

Assumptions Checked: How Families Learn About and Use the Echo Dot

ERIN BENETEAU, University of Washington

YINI GUAN, University of Washington

OLIVIA K. RICHARDS, University of Michigan

MINGRUI RAY ZHANG, University of Washington

JULIE A. KIENTZ, University of Washington

JASON YIP, University of Washington

ALEXIS HINIKER, University of Washington

Users of voice assistants often report that they fall into patterns of using their device for a limited set of interactions, like checking the weather and setting alarms. However, it's not clear if limited use is, in part, due to lack of learning about the device's functionality. We recruited 10 diverse families to participate in a one-month deployment study of the Echo Dot, enabling us to investigate: 1) which features families are aware of and engage with, and 2) how families explore, discover, and learn to use the Echo Dot. Through audio recordings of families' interactions with the device and pre- and post-deployment interviews, we find that families' breadth of use decreases steadily over time and that families learn about functionality through trial and error, asking the Echo Dot about itself, and through outside influencers such as friends and family. Formal outside learning influencers, such as manufacturer emails, are less influential. Drawing from diffusion of innovation theory, we describe how a home-based voice interface might be positioned as a near-peer to the user, and that by describing its own functionality using just-in-time learning, the home-based voice interface becomes a trustworthy learning influencer from which users can discover new functionalities.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI); User studies, Field studies, Interaction techniques**

KEYWORDS

Voice interfaces; Smart Speakers; Digital Home Assistants; Learning

ACM Reference format:

Author First Name Last Name .2019. Assumptions Checked: How Families Learn About and Use the Echo Dot. *ACM Trans. Graph.* 4, XXXX, 2 (June 2019), XXX pages.

<https://doi.org/124564>

1 INTRODUCTION

Despite over 100 million devices with a pre-installed Alexa voice interface sold [5], prior work has shown that users, particularly those who are not early adopters, do not use a wide range of the available functions of voice interface-based assistants [3,8]. After initial experimentation, users rarely explore the ecosystem of available functions over time [33]. Although this may indicate that users prefer a limited set of features, it is also possible that users fail to discover and learn about personally relevant features they would otherwise choose to use. As a result, there may be untapped opportunities to improve the discoverability of features within voice interfaces.

We sought to understand the learning, discovery, and integration process of adopting a voice interface system in the home environment by the entire family household. We conducted a four-week in-the-wild study in which 10 diverse families adopted an Amazon Echo Dot [37] for

the first time. We examined what participants knew about the device in advance, how they explored the device's functionality during the initial setup, how they used it over time, and how they learned about features during the month-long study period. Our research specifically seeks to understand:

- 1) Which features of the Echo Dot do families use?
- 2) How do families explore and learn to use the device?

Using an inductive analysis approach, we found that design decisions and device capabilities were not always in-line with family expectations of learning and use. We found that families' approaches to discovering their devices' functionality were limited, which quickly manifested into a narrow set of functions used over time. Discovering new functionality as time went on was more likely to occur when learning opportunities came from family, friends, or through trial-and-error, in contrast to learning opportunities sponsored by the device manufacturer. Families also attempted to learn from the Echo Dot about its functionalities through the voice interface itself and were met with mixed success.

Key contributions of this research include a design agenda for improved discoverability of the functionality of digital home assistants utilizing a framework based on social relationships. We provide design suggestions so that families can more easily learn the capabilities of their digital home assistant using a "just-in-time" model with the voice interface itself acting as a "near-peer" and trusted learning influencer. We also highlight the importance of considering the entire family group when designing digital home assistants. We do this by including both children and adults' use of the Echo Dot in our analysis and discussion.

2 RELATED WORK

The field of Human-Computer Interaction (HCI) has had an explosion of research exploring the usage of digital home assistants and voice interaction devices [3,8,11,16–20,26–29,33,34]. Because this technology is relatively recent, many research studies focus on early adopters who purchased the technologies themselves and are already using it in their homes [3,18,19,28,33].

Studies on voice interface assistants indicate that the primary uses of the technology are for simple requests. Frequently cited simple requests include: checking the weather, playing music, setting reminders and setting alarms/timers [3,8,17,19,28,33]. Often, users must adjust their speech and language in order for voice interfaces to respond appropriately and communication breakdowns often result in abandonment and distrust that the device will work [2,8,11,19]. As a result, many users might not understand the full capabilities or features of their device [19]. For example, a lab study by Druga, et. al, found that children quickly lost interest in devices when they couldn't answer questions or understand the children's communication [11]. Similarly, Luger and Sellen [19] found that adults using voice assistants are likely to abandon usage due to unreliable responses and lack of trust that the voice assistant would be able to perform anything but the simplest tasks.

2.1 Digital Home Assistants and Families

Sciuto et al. [33] explored how existing users of the Amazon Echo used their devices through an analysis of user history and by interviewing seven users in their homes. They discovered that after an initial few days of experimentation, usage patterns stabilized and were consistent from week to week. Many families owned more than one device, and the room the device was placed in impacted the nature of the interactions and usage of the device.

Bentley et al. [3] reviewed history logs of 88 homes using the Google Home, with a focus on adult usage. They found that peak usages occur early in the morning and afternoon, with a daily high around 5-6 pm. They state that designers should consider the user's context with respect to important functions for the individual based on the time of day. The authors also analyzed the categories of functions that individuals used over a period of 110 days. Their findings show that functions used did not change much over time, and that users did not engage with a wide variety of functionality, with only 4% of users trying a new functionality in week 14 of use. The authors also suggest that despite similarities between voice interaction with digital home assistants and voice assistants embedded in the phone, early adopters use the functionality of voice interaction differently between the two types of devices.

To explore how families, as a group, use the Amazon Echo on a daily basis, Porcheron et al. [26], used audio capture to record the interactions of five families who were given Amazon Echos to use in their homes for a period of four weeks. The authors found that interactions with the Echo are not "conversational" in nature and they suggest that designers should consider that the design assumptions for human-device requests might be "awkward" for some conversational interactions. Lopatovska et al. (2018) also investigated how 19 people of different ages (4 to 55 years old) used their voice interface assistants in their homes, using a diary study method over four days. The authors found that over time, participants reported decreased usage of their devices, and the authors did not find major differences between how children and adults interacted with their device.

Overall, initial research indicates that digital home assistants are not being used to their full potential in family homes. In addition, prior work has found a gulf of evaluation [23] among users of voice interfaces [10], in which users' expectations of voice interfaces are different from the actual capabilities of the system. The very nature of a voice interface cues the user to expect human-like conversational ability, which the voice interface cannot deliver. Our paper explores if this might be due to learning or if there are other design mismatches between user expectations and the design assumptions built into the device.

2.2 Learning to Use Novel Technologies

The field of interaction design and human-computer interaction has had a long history in understanding how to make interactions and features discoverable to end users. Some of the earliest interaction formats for computing systems, command prompts, were known to have poor discoverability and required participants to learn and memorize complicated command prompts. The advent of WIMP systems (Windows, Icons, Menus, Pointers) were a leap forward in discoverability, as their nature made things visible to end users and allowed people to explore possible interactions more freely. During this era, Jakob Nielsen [38] also developed key heuristics for interactive systems, with two of them relating to discoverability, including "visibility of system status" and "recognition over recall." More emergent technologies focusing on natural user interfaces, including voice, input, and gestures, have seen a retreat back into less discoverable interfaces, which has increased the challenge of designing systems that are usable.

A body of research has attempted to improve discoverability by understanding how users might define their own gestures for certain actions within interfaces with low discoverability. These studies were often done as Wizard-of-Oz studies, and in fact, the first known Wizard-of-Oz prototype was developed at IBM to simulate a voice-dictation system, the Listening Typewriter [14], which is similar to today's voice assistants. These studies of non-discoverable interactions includes early work on command-line email interfaces [13], table-top surface

gestures [36], augmented reality [25], voice interaction systems [4], and children using whole-body interaction systems such as with the Microsoft Kinect [6] or tablets [32]. This research often reveals that people are inconsistent in what they find intuitive, but that patterns can emerge that can be useful in designing discoverable systems.

The burgeoning work in HCI regarding modern voice interfaces provides insights into how individuals (primarily adults) use their device: that individuals tend to use a limited set of functions and users are unable to truly “converse” with voice assistants. While there is a well-documented pattern of limited use of the full capabilities of devices, work exploring how individuals learn about voice interface device capabilities is just emerging.

Early research indicates that individuals use guessing and exploration, or a trial-and-error pattern, more than using visual aids when interacting with an unfamiliar voice assistant [7,21,22]. While tutorials can be helpful in providing easy discoverability of some functionality, learning “as-you-go” within the context of actual use appears to be most effective [7]. Learning “as-you-go” was addressed by Corbett and Weber in the form of creating a contextualized help menu for a mobile voice interface assistant, and when compared to tutorials and onboarding tools, the authors found that contextualized help was the most useful form of discoverability for users [7].

Some technology adoption theories address social factors as influencers on the learning and discoverability of technology. The Unified Theory of Acceptance and Use of Technology (UTAUT) [35] incorporates social influence and facilitating conditions into its adoption model. The Senior Technology Acceptance and Adoption Model (STAM) [30] also specifies that social influence and ease of learning are components of technology adoption for the elderly. Diffusion of innovation theory and related concepts includes the social system involved in the dissemination of learning and adoption [10]. Rogers’ diffusion of innovation theory [31] discusses the concept of “near-peers,” who can influence an individual in their adoption of an innovation [1]. Rogers states that individuals often rely on the experiences of others to influence their adoption of innovations and near-peers can model attitudes and ideas about new innovations. Diffusion of innovation work describes how potential adopters search for input from trusted and experienced others in an informal manner and that the framing and timing of a potential user’s openness to learning of innovations is contextually dependent [10].

While these theories and models explore how individuals and groups adopt technology and new innovations, our study has a focus on one specific aspect: learning and discovering the functionality of voice interface technology in the **family** home.

We differentiate our study from earlier work with a focus on families’ expectations, assumptions, and learning about the functionality of their digital home assistant. We focus on families’ assumptions (both children and adults), based on how they learn about and use their Echo Dot over time. We start from the point of initial unboxing of the device, which little research has addressed, but which sets the stage for exploring assumptions of designers and the assumptions of users. Our analysis encompasses both children and adults as members of diverse, multi-generational family homes. As a result, we provide a comprehensive view of how families first engage with, learn about, and use their digital home assistants. We then apply these findings to the concept of near-peers and diffusion of innovation theory to create design guidelines for home-based voice interfaces.

3 METHOD

We recruited 10 families who had at least one child under the age of 18 living in their home. While much of the existing research on digital home assistants is based on early adopters, who also tend to be socio-economically homogenous, we purposely sought out a sample of participants who are less well-represented in the current literature: families who self-identified as having a total household income at or below the median income for the area in which they lived. We advertised for participants through community groups and centers in a centralized metropolitan area on the West Coast of the United States, including recreational centers, community colleges, workplaces, housing centers, and at university campuses. Participant families ranged in composition from families of two to families of five (see Table 1). Two families were bilingual. None of the families had previously owned a digital home assistant or had one in their home.

Families were diverse in their schedules/work hours. Some parents worked evening/night shifts, some parents worked traditional business hours, and some parents were stay-at-home. Some families had children who were part-time at their home and part-time at an ex-partner's home. In addition, some families had one or more grandparents living in the home.

The research team conducted in-person, semi-structured pre-deployment interviews with all family members, in which we asked families about their familiarity with digital home assistants, how they learn about new technologies, and their expectations of how they might use their Amazon Echo Dot. Families were asked to set up the Amazon Echo Dot themselves, and the researchers observed their set up process and provided guidance only if families were unable to successfully set up their Echo Dot after 10 minutes or longer, or if families appeared to "get stuck" and become frustrated.

Families were instructed to place the Echo Dot in a living area where the entire family would be able to access it. Most families chose to place their Echo Dot in the living room or the kitchen area. By week 3 of the deployment, families were e-mailed a document from the first author of the research team, which listed skills and activities geared towards the ages of the children in their home. Links to reviews of the activities were included in the document. Families were told that they were under no obligation to try out any of the activities.

Table 1. Participant Demographics.

Family ID	Total N	Adult N	Adults' age ranges	Children's ages	Ethnicity
A	4	3	41-68+	13	Asian
B	4	3	18-55	12	White/Asian
C	2	1	41-55	9	White
D	4	2	26-55	3, 5	Hispanic
E	5	3	26-67	4, 8	White
F	5	3	18-55	<1, 5	Hispanic
G	3	1	26-40	10, 16	African American/White
H	2	1	41-55	9	White
I	5	2	26-40	2, 4, 6	White
J	2	1	18-25	4	White

After four weeks of deployment, families participated in a post-deployment semi-structured interview in their homes. The post-deployment interview involved asking families to describe their experiences with the Echo Dot, how they learned about the features they used, their expected use versus their actual use, how they might use the Echo Dot going forward, and to provide additional feedback.

Families were compensated with \$100USD in gift cards and their Echo Dot. Our study was approved by our Institutional Review Board and family members provided consent or assent to participate in the study.

3.1 Audio Capture

To understand participants' *in situ* experience with the Echo Dot and the context in which they used it, we used an anchored audio sampling (AAS) approach [15]. AAS is a form of context-aware passive audio sampling that captures both audio that comes after a specific trigger event and, via a continuous backbuffer, audio that precedes the event of interest. Thus, each time the wake word, "Alexa" was spoken, our logging app captured one minute of audio that came before the wake word and three minutes of audio data that followed. This logging app was deployed on a Samsung tablet computer, which was placed near the Echo Dot in each family's home

In addition to this passive audio capture triggered by the wake word, the logging app included a recording interface to support active, participant-triggered data collection. Via a button labeled, "Record a Thought," participants could choose to capture audio at any time, providing a mechanism for them to leave a message for the research team.

To safeguard against passively capturing audio which families did not want the researchers to hear, the logging app included a "delete" button that could be activated with a numeric code (provided to all families). Pressing the delete button permanently deleted any data captured in the prior 10 minutes and logged the fact that data had been deleted.

3.2 Analysis

In total, we captured 14.5 hours of families interacting with the Echo Dot. We also had notes, transcripts and video recordings from pre- and post- deployment interviews. We used an inductive approach to analyze the data [9,24]. A core team of five research team members attended at least one pre- and one post- deployment interview each, and team members discussed trends discovered during interviews in team meetings. For the deployment audio capture and analysis of conversations that families had with the Echo Dot, we utilized memoing during initial reviews of the audio capture.

We then began more formal analysis with two team members reviewing transcripts for specific usage of the Echo Dot, developing open codes while reviewing audio captured interactions between families and the Echo Dot. Team members discussed and reconciled open codes to a formalized codebook detailing families' use of the Echo Dot. The codebook was used as reference for reviewing the audio capture data and coding the data a second time, with a total of three researchers reviewing the final coded data. Audio capture that was not directly related to using the Echo Dot for a specific use were eliminated from the dataset for this analysis. For example, a family discussion about the Echo Dot in which the family did not try to "activate" the Echo Dot for a specific function was eliminated.

Our codebook included a total of 18 codes related to use, which we broadened to a total of three themes of usage functions (Table 2). We also cross-referenced use with the family member

type (adult or child). Two of the researchers used the interactivity of Tableau [39] to explore the audio capture data.

For the purpose of our analysis, we wanted to focus on the types of functionality used over time. Therefore, we analyzed data in terms of “usage instances” in which we noted the types of functions used each hour. For example, if a family member asked the Echo Dot to play music twice in an hour, we noted the use of “music” for that hour segment. If a family member asked the Echo Dot to play music a third time, 90 minutes after the first request, we noted the use of “music” for both one-hour segments. By counting intended use of functions per hour, we are able to visualize how families used the variety of Echo Dot functions over time.

In addition, two researchers also conducted a thematic analysis of all interview transcripts. As a final step in our analysis, we were able to cross reference family expectations and assumptions revealed through the interviews with actual functional use based on audio capture. Additional visualizations of our analysis can be found at <https://tinyurl.com/stu5zu2>.

4 FINDINGS

We report on families’ adoption experiences, from pre-adoption expectations, to unboxing the device, to initial set up and use, through a total of four weeks of usage. We use our visualizations and data analysis regarding functional usage over time to provide a broad overview of families’ usage during the deployment period, in essence providing a landscape from which we delve into specific findings on learning and discoverability.

Table 2. Codes of Echo Dot Use

Theme	Function	Example	Frequency of Occurrence
Entertainment	Game	Alexa, open Panda Rescue.	2%
Entertainment	Joke	Alexa, tell me a joke.	4%
Entertainment	Reading	Alexa, tell me a story.	2%
Entertainment	Music	Alexa, play Hawaiian music. Alexa, play C78.5	29%
Assistance	Timer	Alexa, set an alarm for 5 minutes.	6%
Assistance	Alarm	Alexa, set an alarm for 10:45 am.	7%
Assistance	Reminder	Alexa, set a reminder.	6%
Assistance	Calendar	Alexa, put 10:00 am pick up on my schedule.	2%
Assistance	Math questions	Alexa, what’s 12 times 51?	1%
Assistance	Volume	Alexa, increase volume.	6%
Assistance	Shopping list	Alexa, can you add bananas to my shopping list?	1%
Assistance	Communication	Alexa, call Julie.	2%
Information Seeking	Knowledge questions	Alexa, what’s the biggest bee?	12%
Information seeking	Recipe	Alexa, what’s the recipe for beef stroganoff?	1%
Information seeking	Time	Alexa, what time is it?	4%
Information seeking	Weather	Alexa, what’s the weather?	11%
Information seeking	News	Alexa, what’s the news?	2%
Information seeking	Functionality	Alexa, can you sync with Pandora?	2%

4.1 Expectations of the Echo Dot

Before trying the device, families had a wide variety of expectations about how they might use the Amazon Echo Dot. One family thought they might integrate the Echo with a home security system (Family B) while another thought they might use the Echo Dot to help regulate family disputes by monitoring the volume of their voices (Family H). All ten families expressed an interest in using the Echo Dot to access information, such as new recipes, weather, trivia, and facts, with statements such as, “*one of the things that I’m looking forward to is being able to ask [the Echo Dot] lyrics*” (the grandmother in family E).

Seven families anticipated that the Echo Dot might make daily life easier, as described by the mother in Family I: “*It would be nice for the kids to play music so I don’t have to stop what I’m doing and look it up on YouTube and play it for them.*” One parent (Family H) said that she hoped to use the Echo Dot to help with tasks remotely, such as talking to the dog when the family was away from home as well as “*it would be freaking amazing if it could lock the car.*”

The majority of families (seven) expected that they would use the Echo to play music. Four families had heard about the Echo’s ability to control lights. Nearly half of our families expected to use the Echo Dot for reminders or to-do lists. However, we found that families who expected to control the lights, play music, or manage to-do lists, were unaware that an app or additional appliance was required to access these functions. For example, at the exit interview, after one month of using the Echo Dot, the parent from Family G said: “*Oh, one thing I did that I was upset about. I remember... I was like, oh, we can do the smart home thing now with the light bulbs and stuff. I ordered this light bulb, and nothing in it had told me I needed a bridge thing.*”

Based on our interviews, we see that families’ assumptions of the Echo Dot’s functionality and potential use were not always in line with the design of the device.

4.2 Unboxing, Initial Setup, and Video Tutorial

The research team observed each family as they set up their device for the first time. Families were encouraged to engage in the unboxing and set up process as they would naturally. We consistently observed that families assumed the digital home assistant would be able to respond to voice interaction almost immediately. For example, family C is representative of most families’ unboxing and set up experiences:

(Mom): Is that a power button?
(Child): Do you know how to turn it on? Alexa. (pause) It plugs in before you’re going in. Wow.
[pause] Play ‘Sky’ by Alan Walker.
(Mom): It might be charging, I don’t know.
(Child): No? Maybe?
(Mom): Do you need to say Alexa again?
(Child): Yeah. Probably. [pause] Say ‘Alexa.’
(Mom): Alexa. Alexa.
(Mom): Stop.
(Alexa): Hello
(Child): Yay!
(Alexa): Your Amazon Echo Dot is ready for set up. Just follow the instructions in your Alexa app.

Families attempted to talk to Alexa immediately after plugging in the Echo Dot, often collaboratively troubleshooting how they might initiate a conversation with the assistant, ACM Trans. Graph., Vol. 4, No. 2, Article 39. Publication date: October 2017.

indicating a lack of awareness that the setup process included an app. Families assumed that setting up a digital home assistant would be a self-contained experience requiring only the device itself and were often surprised to learn it required an Amazon account and a companion app.

We also observed seven families watch the initial demo video as part of the setup process, in which families learned about selected features of the Echo Dot. For example, our observation and field notes of family E's experience watching the demo video show how the children gather around their mother, watching the video on their mother's phone while other family members are sitting in the living room, looking towards the mother's phone, listening intently:

Alexa video demonstrates how to play music.

(Dad): That's cool [said enthusiastically]

Alexa video demonstrates timer ideas.

(Multiple family members): Wow

Alexa video demonstrates controlling lights.

(Multiple family members): That's cool!

Most families were guided by the formal learning experienced in the demo video, and they completed the setup process by asking Alexa the sample questions listed in the app (e.g., "Alexa, what's the weather?"). However, some families bypassed the questions suggested through the formal learning or quickly branched out to questions of their own (e.g., "Alexa, what's the earth's gravitational pull on the moon?").

4.3 Family Usage Patterns

To understand which Echo Dot features families discovered and engaged with during their month of use, we first examined families' usage patterns using our audio capture data. Consistent with prior work, participants primarily used the Echo Dot for music, jokes, weather reports, alarms and timers. These, and other patterns of use clustered into three larger themes: information, entertainment, and assistance. Table 2 shows the percentage of use for each individual category by all families. Figure 1 shows the frequency of use across the three larger themes, distinguishing between adults and children usage.

Although families often used the Echo Dot for entertainment purposes, they engaged in relatively little exploration and feature discovery in this category. The overwhelming majority of entertainment usage instances from both adults and children were commands to play music,

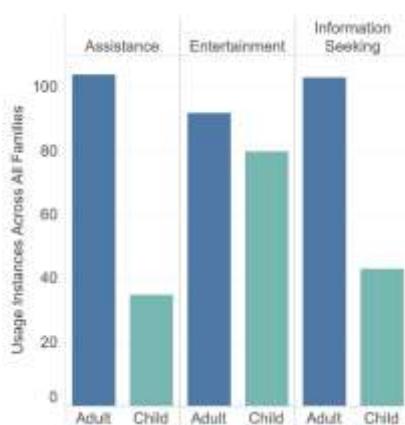


Figure 1. Themes of Use by Family Role

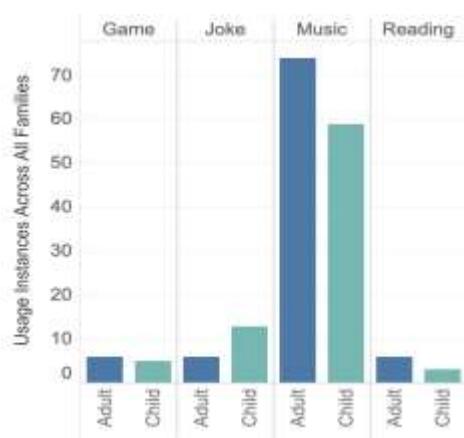


Figure 2. Entertainment Categories Used by Adults and Children

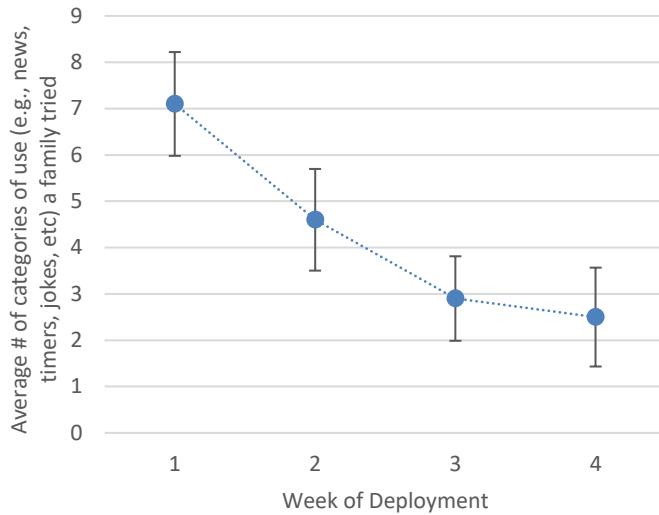


Figure 3. Number of Categories of Use Families Tried Each Week

with only minimal engagement with other entertainment features like jokes, games and reading (Figure 2).

As the deployment progressed, families tried fewer and fewer features, narrowing to a small set of categories by the fourth week (Figure 3). At post-deployment interviews, families described engaging with the Echo Dot less than they originally expected, such as the parent in Family F: “*And I honestly thought that we would do more...But I don't know if it's because we've never had it before or if we, maybe also because we don't know what skills it has. And so as you learn, you use those right. You use what you learn.*” Similarly, the parent of Family H said: “*I got my habit of listing, reminders and music. And then I just stopped being interested in it.*” Such statements both confirm what we saw empirically (i.e.- that families engaged with a limited set of features) and also suggest that this narrow set of uses may reflect a struggle to discover and learn of new features.

All families attempted to use the Echo Dot as a source for information, with varying degrees of success. Families reported disappointment with how frequently the Echo Dot was unable to answer their questions when they turned to the Echo Dot as an authority and information source, as explained by the child in Family C: “*I'm kind of tired of hearing, 'I just can't help you with that...' What if it could be like, 'I think you should try again, I might've not got that correctly.'*” Family F also explains their frustration: “*It's frustrating when you ask a machine a question and the machine responds 'I don't know, I don't have that information.' Why don't you know? That's your job to know.*”

With these quotes, we see how family members experienced frustration when their assumptions of the Echo Dot’s capabilities did not match the devices’ functionality. As described by the child of Family C, the Echo Dot’s responses could be modified to be perceived as more helpful.

4.4 Learning about the Echo Dot’s Functionality

During pre-deployment interviews, we asked families what they knew about the Echo Dot's functionality and how they had learned that information. One family intentionally visited a physical store to learn about and try out the Echo. However, the majority of families indicated that they learned about the Amazon Echo and other digital home assistants through incidental exposure, such as through commercials, friends, or family members who owned an Amazon Echo or similar device. Families also mentioned learning about specific features through their family and friends, explaining, for example, that, "*Grandma's Alexa, it tells jokes*"(child, Family J) and, "*They're [friend's children] able to call or message her [friend] when she's at work...and so, even though they don't have their own cell phones, they can still communicate while she's still at work*"(parent, Family I).

Families reported that, more generally, when they learn about the functionality of other new technologies they acquire (such as tablets, TVs and phones), they use strategies such as Internet searches (such as Google and YouTube), reading manuals, and engaging in trial-and-error. However, we found that families' actual approaches for learning about the Echo Dot's functions were significantly more limited during the study deployment period.

At follow-up interviews, we asked families to discuss how they learned to use the Echo Dot over the course of the month-long deployment. Families confirmed their earlier statements during initial interviews, explaining that they had tried functions that they'd learned about from family and friends, as well as through "*messing around*" and "*experimenting*." Six families reported learning of functionality through family members, friends, or in collaboration with other family members. Family D reported looking at the Alexa app to learn about functionality. In addition to family self-reporting during exit interviews, audio capture reveals that half of the families (Families B, C, D, F, and H) attempted to learn of the Echo Dot's features directly from interactions with the Echo Dot itself by asking Alexa.

Collectively, participants engaged in two major forms of learning during the study:

- 1) Learning about the Echo Dot from the Echo Dot
- 2) Learning about the Echo Dot from Outside Influencers.

In this section, we describe these two learning sources in-depth, and provide examples of the differences between effective and ineffective learning experiences.

4.4.1 Learning About the Echo Dot from the Echo Dot

As families interacted with the Echo Dot, they learned about its capabilities and limitations, which at times increased learning and discoverability and at other times, inhibited learning and discoverability. Families learned from the Echo Dot in three distinctive ways: 1) Device-initiated learning, 2) User-initiated learning, and 3) Trial-and-Error. Using audio capture and interview data, we describe each of these patterns of learning about the Echo Dot from the Echo Dot.

4.4.1.1 Device-Initiated Learning: Building on Context

As families used the Echo Dot, the device itself attempted to stretch the user's understanding of the device's functionality. The Echo Dot provided these context specific suggestions in response to users' commands or queries. An example of a successful device- initiated learning moment occurs with Family A:

(Parent): "Alexa, what package just arrived?"

(Alexa): "To avoid ruining any surprise gifts, you can choose whether I say names of products being delivered. Should I say them from now on?"

(Parent): "Yes."

(Alexa): “Okay. I’ll say the product names. If you’d like to change this, go to settings, notifications in the Alexa app.”

Similarly, we see additional examples with Family C and Family F. For example, when Family F asked the Echo Dot for a recipe, at which Alexa responded: “*Do you want me to tell you the ingredients, send this to your phone, or hear the next?*” the adult replied “*Send it to my phone.*” Family C had a similar experience when asking Alexa for a listing of events in their local area and Alexa responded “*Hmm, maybe the skill [city] event calendar would help. Want to try it?*” to which the adult replied “yes.”

At other times, these device-initiated learnings were unsuccessful, even when they were within the context of the user’s-initiated communication topic. Audio capture evidence demonstrates how families ignore the device’s learning suggestions and resume engaging with a previously discovered functionality, such as playing music, as demonstrated with Family C:

(Child): Alexa, is Dan TDM livestreaming?

(Alexa): Sorry, I’m not sure.

(Child): Alexa, is Crickcraft livestreaming on YouTube?

(Alexa): I couldn’t find any video skills that could play that. Go to the music, video, and books section to see those skills and devices.

(Child): Alexa, stop.

(Child): Alexa, play “Means” by Navarro.

Finally, device-initiated learning took on a marketing tone at times. Amazon Music services featured as an example of this type of device-initiated learning in both audio capture and at exit interviews. For example, audio capture from Family H demonstrates how the flow of the user-device request is disrupted somewhat by the Echo Dot attempting to provide information regarding its functionality in the form of Amazon Music services:

(Child): Alexa, please play Havana by Camila Cabello.

(Alexa): Havana isn’t included in Prime but is available in Amazon Music Unlimited.

Would you like to learn more?

(Child): No.

(Parent): Not right now.

(Alexa): Okay, you can sign up anytime by saying, sign up for Amazon Music Unlimited. Here’s a sample of Havana by Camila Cabello. [plays said song]

Device-initiated learning interactions with a marketing tone led to confusion for some families, as described by the child in Family B: “*At first I was really confused on how to listen to the samples of it [music], so I think that it should just be simpler. Not the whole, ‘Oh, you have to subscribe to this Amazon Music thing in order to ...’ It was really weird that it let you play a sample. I’m not complaining, but it’s just that it was so strange, the way that they asked... I think they should say something along the lines of, ‘You can get a subscription or we can just play a sample of a song.’ And I think there should be some sort of thing that you can say so it just stops asking you every single time for the subscription.*” As a result, some families remained unaware of the full music capabilities available with the Echo Dot, despite music being the dominant entertainment function used by families during the month-long deployment.

4.4.1.2 User Initiated Learning: Asking the Echo Dot about Itself

In addition to learning about the Echo Dot’s features through context-dependent interactions with the Echo Dot, families also proactively sought to learn about new features by asking the Echo Dot directly, such as when the adult in Family C asked, “*Alexa, what can you do?*”, and Alexa responded with “*There’s a lot of things I can do. You can say, ‘Set a 10 minute pasta timer,*” ACM Trans. Graph., Vol. 4, No. 2, Article 39. Publication date: October 2017.

'Wake me up to music,' or 'Wake me up to 80s music.' Do you want to hear some more?". However, families who attempted to learn about the Echo Dot's functionality from the Echo Dot were met with mixed success. Family F described their experience: "So I asked about the skills that she had available. The one thing I didn't like was that she started all the way from the top. So every, 'we have this skill, would you like to try it', right? So every time. 'So we have this skill would you like to try it?' No, next. 'Would you like this skill?' Oh yeah, more and more. And so, 'do you want to install it or do you want to open it?' No. Start all the way over. I didn't like that."

In this case, the Echo Dot was able to describe its functionality, as requested, but in an inefficient and laborious way which did not meet the user's needs. Ultimately, this resulted in abandonment of attempting to learn about the Echo Dot by asking the Echo Dot about its skills.

These unsuccessful learning attempts at times resulted in missed opportunities for users to discover new functionality of the Echo Dot, such as with family F:

(Adult): (singing) Alexa, do you support Pandora?

(Alexa): For help with that question, go to the help and feedback section of the Alexa app.

(Adult): Alexa, next song (singing). Alexa, turn the volume up.

Here we see the family attempting to learn about the Echo Dot from the Echo Dot, seeking to learn about its potential functionality, however, the suggestion of moving from a voice interface modality to another interface (the app) inhibits learning within the context desired by the user.

4.4.1.3 Trial-and-Error: Guessing and Experimenting

Prior work refers to the act of "guessing" as a user strategy when interacting with voice user interfaces [7,21,22]. Our audio capture and interview data support this as one of the learning approaches used by families in the home environment. For example, the child in Family C referred to their learning approach as: "*I experimented.*" Trial-and-error approaches rely heavily on discovering key terms that can be understood by the device. Families expressed difficulty in finding the right words to use to ask the device about itself. For example, when the child in Family C wanted to play a version of a song where its explicit lyrics were redacted, his mother told him to ask the Echo Dot to find that version of the song. The child explained that, "*He [the singer] has one [a non-explicit version of the song] but...I don't know how to ask.*" The parent of Family H described her struggle as well, explaining: "*I feel like there's certain language, you know, there's certain words that cue her to do certain things. Like, I would stumble on them accidentally. So like I would say things over and over again, and I would try different ways of asking her, and there's one time that I was able to add multiple things, but I don't remember what I said.*"

When the correct language is used, families were able to learn about and discover new functionality through trial-and-error. For example, the mother of Family E explained: "*I did figure out on my own about the whole shopping list... and so I went to do it, she's like, 'Oh I don't have an app set up,' so we went into the Alexa app to find out which apps can we use for the shopping list... So that's how we set that one up and we love that function.*"

Family E also discovered customization settings both for the calendar and music functions utilizing trial-and-error. The parent from Family E describes their experience with discovering music functionality: "*So I was like, 'Alexa, open Spotify.' She opened Spotify, and then we were able to play it. So then I had to go into the Alexa app and look to see which was the default music player. Once we changed it from Amazon to Spotify everything was fine.*"

Similarly to Family E, Family D learned about the Echo Dot's music settings through trial-and-error: "*It took me a little while to like, realize that I didn't just have to use Amazon music...It's*

very geared towards like, the Amazon life, which is fine. But, it wasn't until after a while that I figured out I can actually connect her to other like Spotify and Pandora."

However, rather than persisting in exploring the Echo Dot's capabilities through trial-and-error, some families reverted back to familiar technologies in which functionality and settings were known. For example, Family A expressed a preference for using various functions on their phones because they felt that their phones were more capable than the Echo Dot. The Echo Dot's attempts at providing in-context learning of its Amazon Music feature further added to Family B's lack of interest in continuing to explore and discover features of the device. During their exit interview, Family B explained that they reverted to using their phones as primary resources during the research study, and said: "*The phone can do everything Alexa can do, you [researcher] didn't say anything, but Alexa can probably do things that the phone can't do, right? How do we learn what those things are without Alexa looking like they're trying to push marketing stuff on us?"*

Thus, we saw that families were interested in exploring new features and saw the Echo Dot as a potential source of information about itself. When the device answered family member's questions successfully, they learned about, tried, and even adopted new features. However, the device's inability to efficiently explain its capabilities in a relevant and meaningful way to users ultimately undermined users' ability to ask the device to guide them to discover new and relevant features.

4.4.2 Learning About the Echo Dot from Outside Influencers

While the most prevalent methods of learning occurred from learning about the Echo Dot from the Echo Dot, families also learned about the Echo Dot's functionality from outside influencers, ranging from other family members and friends to corporate learning and marketing materials.

4.4.2.1 Learning from Family and Friends

During exit interviews, six families reported learning of functionality through family members, friends, or in collaboration with other family members during the study deployment. The parent of Family C said "*So I always kind of watch and see what [child] does, because he just jumps [in and tries things].*" In the audio capture data we also find that family members build on functionality introduced by other family members, such as with Family I: (Adult) "*Alexa, make volume 4,*" immediately followed by the child saying, "*Alexa, make volume 50.*" In addition, family members directly explained to other family members about the Echo Dot's functionality during the deployment, as audio captured from Family E:

(Father): Hey if you need anything while I'm here, just tell Alexa.

(Father): It's under shopping, so say [quietly] Alexa, add white bread to the shopping list.

(Grandma): Alexa, please add white bread to the shopping list.

(Alexa): I added white bread to your shopping list.

(Grandma): Thank you!

(Father): We basically have where she creates the list you tell her and then we –

(Grandma): Gain access whenever we want.

(Father): Yes, you can access it whenever we want on our phones in the store.

Family members learned about the Echo Dot's functionality within the home environment from other visitors to their home, such as friends. The parent in Family I explains: "*When one of my friends was over, she's the one that has a bunch of them [Echos] and uses them extensively...and so she has Audible on hers, and we don't have Audible...Yeah, I didn't think about*

Audible and her [Alexa] being able to tell stories." In this case, the family member learned of new functionality and possible applications from a friend, who also is an experienced user and earlier adopter of the technology.

In addition, family members learned about the Echo Dot's functionality from other family members outside of their home environment. For example, the child in Family J reported learning about using an Echo at another family member's home, and the parent in Family C reported learning a new functionality (listening to the news) from how another family member used the device in their home.

4.4.2.2 Learning from External, Formal Sources

As discussed in the initial interviews, families also learned of some basic functionality through external sources, such as commercials. After the study deployment, families did not describe commercials as a major learning source, however, during the study deployment families did learn of new functionality through other external sources. The least frequent source of external learning came from internet searches. One family member of one family member (Family B) reported using Google to find "*funny things to ask your Alexa*," another family member (Family F) reported using YouTube to learn about privacy, and one family member (Family D) actively searched the Alexa app to learn new skills. All instances were reported as one-time activities, rather than ongoing resources for continued learning.

As part of the study, our research team created a resource of Alexa skills customized to reflect the age ranges of the children participating in the study. The first author, the primary research contact for the participant families, emailed the skills sheets to the point of contact for each participant family. Families were told that they were not expected to use any of the skills on the skill sheets but had the option to try them if they wanted.

When discussing external, formal learning sources, families were more receptive to the researcher's email than to emails from Amazon for learning. "*I didn't even know that it could play games until [researcher's name] sent the list*" (parent, Family E). "*I really appreciated the list that I received because I was able to sit down and look at it and compare and really pick and choose, like the guessing game*" (parent, Family F). Family G reported: "*[Researcher's name] had sent me some different ideas, games and stuff, so we tried that. It was like a fill your own adventure book type of thing. It actually understood her [young child], which I thought was very cool.*"

Family I also reported that their children were in the room when they read the researcher's list of "*things we could do*" and as a result, the children learned about the Echo Dot's functionality from the researcher's email as well as the adults.

In contrast, families' reactions to Amazon's emails were viewed more often as spam and impersonal: "*I don't think it got to my primary inbox*" (parent, Family E), "*I probably just skimmed them*" (parent, Family J), "*I just thought it was some spam email*" (parent, Family B). One family, Family F, did indicate that they read at least one of the emails from Amazon: "*I did [see the emails] and I don't remember what it was but I came home and I was asking about the skills so that's one of the things that got me to start asking about the skills that Alexa is capable of doing.*" In this case, the email from Amazon facilitated the family's attempts to learn about the Echo Dot from the Echo Dot by asking it questions about its skills (described earlier in section 4.4.1.2).

4.4.3 Limited Learning Impacts Usage

As a result of limited learning, some families went through the entire four-week deployment without knowing about some features that they later said they wished were available. For

example, both families C and F did not realize that the Echo Dot could connect to Spotify until they were told about it by the research team. At the post-deployment interview Family F said: “*I didn’t even know you could connect Spotify to it...I have Spotify on my tablet...I suppose it really doesn’t say what it [Echo Dot] can do, what it can connect to.*” The parent of Family C contrasted the music services of Alexa with Spotify during the post-deployment interview, at the time, unaware that the two services could be integrated: “*So he [child] uses Spotify on my phone, and he finds a lot more on Spotify than he did with Alexa.*”

During our initial interview visit, in which families unboxed and set up their Echo Dot for the first time, we observed the majority of families watched the informational video in the Alexa app about using the Echo Dot. They then used that information to form assumptions about how they could interact with the Echo Dot. Interestingly, in post-deployment interviews, when families were asked if they reviewed the video again or remembered the video, the majority of families had forgotten about it, such as Family G: “*I forgot it existed.*” We see how this video was a perfect “just-in-time” learning experience, in which families received structured, relevant learning at the time that they needed to apply that learning. We did not observe other, similar learning events occur consistently throughout the deployment period.

Our families are not early adopters, and our findings show that the strategies that these families employ to learn about their new technology did not result in discovering the full functionality of their digital home assistants. We also see that a mismatch exists between how families attempted to learn about the Echo Dot, and the learning experiences designed by the manufacturer, resulting in limited learning from the manufacturer, an outside influencer. As Family F put it during their exit interview from the study: “*I wish that the makers would not assume that their consumers are as high-tech or half the level that they are at.*”

5 DISCUSSION

We found that families decreased their exploration and breadth of use of their digital home assistants over time, consistent with prior work studying individual users [3]. Our interest in this study is to explore the heart of the problem: how families are learning about and discovering the functionality of their digital home assistants. In our discussion, we address this problem in the context of our findings. Building on concepts from diffusion of innovation theory [10,31], we provide a framework in which we suggest how designers of home-based voice interfaces can incorporate just-in-time learning into digital home assistants. We also provide specific design suggestions for developing home-based voice interfaces as trusted learning partners within the family home.

5.1 Discovery of Voice Interfaces Should Be Voice-Based, Just-in-Time, Trustworthy Experiences

We found that from initial unboxing through to the end of the month-long deployment, families expected their interactions with the device to be voice-based. Our results build upon prior work, which indicates that users of voice interfaces do not rely on visual inputs for learning about and discovering functionality of their devices [21]. These findings, paired with previous work, strongly support the idea that designs for increasing discoverability of functions should be voice-based.

Previous work has indicated that individuals more often employ a “guessing” tactic to try to overcome obstacles and learn device functionality [7,12,21,22]. We also found that the majority of the families in our study employed “trial-and-error” as one approach to interacting with their

digital home assistants. While this approach is somewhat haphazard and inefficient, designers can build on this common method of attempted discovery to employ “just-in-time” learning opportunities through a voice-interface system.

The concept of just-in-time learning has been discussed widely within the instructional design communities for workplace learning [40–42], and these concepts can be applied to voice interface home technology learning. Technology designers can partner with instructional designers to promote just-in-time learning opportunities within the voice interface ecosystem. For example, when a similar key word is used in succession, but the device is unable to understand the command/request, a learning prompt can be deployed, such as *“I seem to be having trouble understanding your request. Would you like to hear some of the things that I can do related to [keyword]?”* or *“I’m having trouble understanding your request. Say ‘help’ and I can give you some phrases that I can understand.”*

One of the key concepts related to designing just-in-time learning opportunities for home-based voice interfaces is that the device itself becomes a trusted learning partner. As we saw from our interviews, attempts at just-in-time learning that pushed Amazon Music services did not instill trust or increased discoverability with our participant families. We encourage designers to think of the home-based voice interface as a “near-peer,” in which communication takes place between entities which are similar, yet not the same [31,43]. The modality of a voice interface creates a platform in which communication between the user and device are similar, yet decidedly not the same. Voice interaction creates the potential for a digital home assistant to be viewed as a near-peer by both children and adults: a trusted and experienced learning influencer with which the potential adopter (the user) can talk to and learn about the device’s functionality [10,31].

5.2 A Framework for Learning about Voice Interface Devices in the Home

As described in diffusion of innovation theory, the social system and the influence of trusted opinion leaders are key factors in the adoption of an innovation [10]. Our findings demonstrate that families learned of new functionality during the study deployment period primarily through family, friends, and from the research team. All of these outside influencers had established social ties to the learner. Therefore, we build on Rogers’ concept of “near peers” [31] and have developed a framework to describe families’ receptiveness to learning based on their social system. We describe learning influencers in terms of social distance, incorporating our

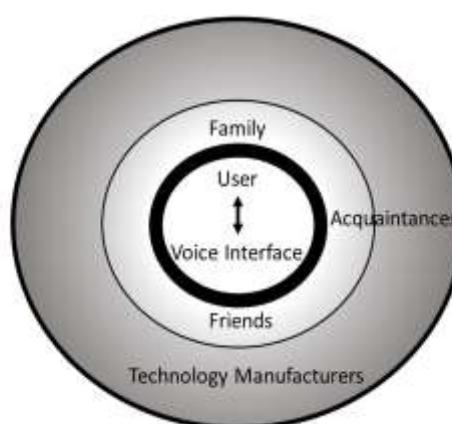


Figure 6. Technology Learning Circles of Influence: Learning about Technology from the Technology and Learning from Outside Influencers. ACM Trans. Graph., Vol. 4, No. 2, Article 39. Publication date: October 2017.

findings of “learning about the Echo Dot from the Echo Dot” (Section 4.4.1) and “learning from outside influencers” (Section 4.4.2). We visualize a series of circles of influence on learning, which represent social distance from the user (Figure 4). We place the home-based voice interface in the inner circle, closest to the user in physical and social distance, in which the user can learn about the technology from the technology itself. By placing home-based voice interface technology itself as a near-peer we extend the concept of trusted opinion leaders and learning influencers [10] from human learning partners to a technological learning partner.

The dark line encompassing the user and the technology represents the concept of learning about the technology from itself. This inner circle draws from our findings of families attempting to learn about the Echo Dot from the Echo Dot, and formalizes the concept.

Outside of the inner circle are outside learning influencers, which include all of the other learning influencers described in our findings. The receptiveness of the learner to these outside influencers is represented by the darkening shading within these learning circles. The user’s family is next closest in terms of degrees of separation in the home environment and are the most immediate “near-peers” that the user will interact with in addition to the home-based voice interface technology. Friends are the next closest, particularly those friends who visit the home and model interactions with and attitudes about the technology. Acquaintances bridge the span between the user’s more trusted “near-peers” to learning sources that are further separated from the user. Ironically, the entities that are furthest from having influence on the learning of the user are the entities that have a formal role to play in creating learning opportunities regarding the technology: technology manufacturers. Nonetheless, our findings revealed that families were less receptive to emails from Amazon than learning less formally from their family, friends, and from the device itself.

5.3 Applying the Model to Voice Interfaces in the Home

We propose that designers create just-in-time learning experiences embedded within voice interfaces, using characteristics one would expect from a near-peer. A key concept in applying this model to the design of voice interfaces for the home is to create the voice interface as a trustworthy learning influencer and opinion leader [10]. Our findings show that families learned from each other and from friends during the study, therefore, we encourage designers to think of the qualities a friend would have as a learning partner. For example, to establish trustworthiness, a voice interface should be transparent and specific in what it can and cannot do.

Half of the families in our study assumed that the Echo Dot would be able to provide information about its capabilities and functions through voice interaction. Sometimes their queries were successful and sometimes they were not. In order to establish the voice-interface as a trusted learning partner, the technology needs to be more successful and more transparent about its capabilities when queried about its functionality.

To demonstrate the application of our model, we contrast an excerpt from our audio capture regarding music services with what the interaction might have been, using our framework (Tables 3 and 4).

In both of the examples listed in Tables 3 and 4, the interaction based on our conceptual framework establishes the voice interface as a trustworthy learning influencer. The information provided is within the context of the user’s request, and the device does not attempt to provide more information after the user indicates a desire to end the learning experience.

Table 3. Contrasting Actual Interaction with Possible Interaction Using Framework, Option 1

Excerpt from Audio Capture, Family H		Possible Version of Interaction Using Voice Interface Near-Peer Framework	
Child	Alexa, please play Havana by Camila Cabello.	Child	Alexa, please play Havana by Camila Cabello.
Alexa	Havana isn't included in Prime but is available in Amazon Music Unlimited. Would you like to learn more?	Alexa	I can't find that song within [default music program] but it might be available via Pandora, Amazon Music Services or Spotify. Would you like to setup one of these music services now?
Child	No.	Child	No.
Parent	Not right now.	Alexa	Ok.
Alexa	Okay, you can sign up anytime by saying, sign up for Amazon Music Unlimited. Here's a sample of Havana by Camila Cabello. [plays said song]		
Excerpt from Audio Capture, Family H		Possible Version of Interaction Using Voice Interface Near-Peer Framework	

Table 4. Contrasting Actual Interaction with Possible Interaction Using Framework, Option 2

Child	Alexa, please play Havana by Camila Cabello.	Child	Alexa, please play Havana by Camila Cabello.
Alexa	Havana isn't included in Prime but is available in Amazon Music Unlimited. Would you like to learn more?	Alexa	Havana isn't included in Prime, but may be included in other music services. Would you like to hear a list of the different music services I can access?
Child	No.	Child	Yes
Parent	Not right now.	Alexa	I can play music from a variety of music services, including Amazon Music, Spotify, and Pandora. Would you like to learn more about these services?
Alexa	Okay, you can sign up anytime by saying, sign up for Amazon Music Unlimited. Here's a sample of Havana by Camila Cabello. [plays said song]	Child	No
		Alexa	Ok

A key component of our framework is for designers to create near-peers that are transparent with their just-in-time learning information and do not overtly push one type of functionality.

To become an effective learning partner and near-peer, the technology should be trusted and considered valuable and relevant in relation to the human user's information needs. Designers will need to consider transparency with privacy, security, and marketing practices as components of designing a trusted near-peer.

5.4 Resolving the Assumptions Mismatch through Learning

Despite indications that families are not using voice interaction technologies to their fullest potential, voice interface technology continues to grow in the marketplace. As the family home becomes more crowded with voice interface systems, designers will better serve families if they enable them to discover the features of these systems in a way that is personally relevant and valuable.

To resolve the mismatch between user's assumptions and the technologies' capabilities, we suggest the following principles for designing home-based voice interfaces:

- 1) Create just-in-time learning opportunities in which the device offers to provide relevant information based on a new and/or unique user query or a series of repeated failed queries.
- 2) Create short, interruptible learning experiences, in which the device can also be customized through settings to adjust how frequently it offers learning opportunities.
- 3) Create learning opportunities in which the device is transparent about its abilities.
- 4) Create learning opportunities that perpetuate trust with the user, including user-driven settings which disable/enable the device to share new features from the manufacturer and/or third parties.

- 5) Create learning opportunities based on user interests (including child-friendly learning opportunities) through user-driven settings.

If designers of home-based voice interfaces conceptualize their technologies as potential “near-peers” who are trusted and transparent learning influencers, family members could rely on asking the technologies about their capabilities as a primary form of learning. Family members could use this same learning strategy across a variety of home-based voice interfaces. As a result, family members would not need to struggle to understand how to learn about a device and its functionality, rather, they could focus on exploring and discovering the variety of capabilities of the technologies in their homes.

Limitations

This study is limited to 10 families in one region of the United States. While family participants include adults of a range of ages (from 18 to 64+), we did not conduct an analysis of the differences between adults within the home and we limited our analysis within families to the differences between children and adults. We also did not conduct quantitative analyses, such as distinct counts for each individual function’s instance of use, particularly since this is well covered by earlier research and due to our smaller sample size [3]. We also recognize that we propose a framework for learning that has not yet been tested but is grounded in the data analysis from our study. Future research on learning the functionality of digital home assistants is needed to continue development of this framework, including families from a variety of socio-economic and cultural backgrounds.

6 CONCLUSION

We conducted a four-week deployment study with 10 diverse families using the Amazon Echo Dot and found that families’ overall usage and exploration of new features decreased over time. Families indicated that they learned of new features during the study through trial-and-error and from people they knew. Families were more receptive to informational emails from the research team about features of the Echo Dot compared to emails from Amazon about the Echo Dot. Families attempted to learn about the Echo Dot from the Echo Dot itself, assuming that the voice interface was a natural conduit for exploring and learning about the device’s functionality. However, asking the Echo Dot about its functionality was met with mixed success. As a result, we suggest that designers create home-based voice interfaces using the concept of a “near-peer,” in which the voice interface is a trustworthy learning partner which aids in increased discoverability of its own functionality.

ACKNOWLEDGMENTS

We sincerely thank the participant families. We also thank the team who helped with this research study, including: Bella Chiu, Leanne Liu, Sharon Heung, Kate Yen, Yuxing Wu and Ashley Boone. We also gratefully acknowledge Mozilla’s funding and support of this research study.

REFERENCES

- [1] Yvonne Barnard, Mike D. Bradley, Frances Hodgson, and Ashley D. Lloyd. 2013. Learning to use new technologies by older adults: Perceived difficulties, experimentation behaviour and usability. *Comput. Hum. Behav.* 29, 4 (July 2013), 1715–1724. DOI:<https://doi.org/10.1016/j.chb.2013.02.006>

- [2] Erin Beneteau, Olivia K. Richards, Mingrui Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication Breakdowns Between Families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 243:1–243:13. DOI:<https://doi.org/10.1145/3290605.3300473>
- [3] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3 (September 2018), 1–24. DOI:<https://doi.org/10.1145/3264901>
- [4] Niels Ole Bernsen, Laila Dybkjær, and Hans Dybkjær. 1998. *Designing Interactive Speech Systems: From First Ideas to User Testing*. Springer London, London.
- [5] Dieter Bohn. 2019. Exclusive: Amazon says 100 million Alexa devices have been sold. *The Verge*. Retrieved May 15, 2019 from <https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp>
- [6] Sabrina Connell, Pei-Yi Kuo, Liu Liu, and Anne Marie Piper. 2013. A Wizard-of-Oz Elicitation Study Examining Child-defined Gestures with a Whole-body Interface. In *Proceedings of the 12th International Conference on Interaction Design and Children* (IDC '13), 277–280. DOI:<https://doi.org/10.1145/2485760.2485823>
- [7] Eric Corbett and Astrid Weber. 2016. What Can I Say?: Addressing User Experience Challenges of a Mobile Voice User Interface for Accessibility. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '16), 72–82. DOI:<https://doi.org/10.1145/2935334.2935386>
- [8] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. “What Can I Help You with?": Infrequent Users' Experiences of Intelligent Personal Assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '17), 43:1–43:12. DOI:<https://doi.org/10.1145/3098279.3098539>
- [9] John W. Creswell and Cheryl N. Poth. 2018. *Qualitative inquiry & research design: choosing among five approaches* (Fourth edition, ed.). SAGE, Thousand Oaks, CA.
- [10] James W. Dearing. 2009. Applying Diffusion of Innovation Theory to Intervention Development. *Res. Soc. Work Pract.* 19, 5 (September 2009), 503–518. DOI:<https://doi.org/10.1177/1049731509335569>
- [11] 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* (IDC '17), 595–600. DOI:<https://doi.org/10.1145/3078072.3084330>
- [12] Anushay Furqan, Chelsea Myers, and Jichen Zhu. 2017. Learnability Through Adaptive Discovery Tools in Voice User Interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '17), 1617–1623. DOI:<https://doi.org/10.1145/3027063.3053166>
- [13] Michael Good, John Whiteside, Dennis Wixon, and Sandra Jones. 1984. Building a user-derived interface. *Commun. ACM* 27, 10 (1984), 1032–1043. DOI:<https://doi.org/10.1145/358274.358284>
- [14] John Gould, John Conti, and Todd Hovanyecz. 1983. Composing letters with a simulated listening typewriter. *Commun. ACM* 26, 4 (1983), 295–308. DOI:<https://doi.org/10.1145/2163.358100>
- [15] Alexis Hiniker, Jon E. Froehlich, Mingrui Zhang, and Erin Beneteau. 2019. Anchored Audio Sampling: A Seamless Method for Exploring Children's Thoughts During Deployment Studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, 1–13. DOI:<https://doi.org/10.1145/3290605.3300238>
- [16] Hyunhoon Jung, Changhoon Oh, Gilhwon Hwang, Cindy Yoonjung Oh, Joonhwan Lee, and Bongwon Suh. 2019. Tell Me More: Understanding User Interaction of Smart Speaker News Powered by Conversational Search. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), LBW2717:1–LBW2717:6. DOI:<https://doi.org/10.1145/3290607.3312979>
- [17] Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. 2018. Talk to me: Exploring user interactions with the Amazon Alexa. *J. Librariansh. Inf. Sci.* (March 2018), 0961000618759414. DOI:<https://doi.org/10.1177/0961000618759414>
- [18] Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a Mindless Companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (CHIIR '18), 265–268. DOI:<https://doi.org/10.1145/3176349.3176868>
- [19] Ewa Luger and Abigail Sellen. 2016. “Like Having a Really Bad PA”: The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 5286–5297. DOI:<https://doi.org/10.1145/2858036.2858288>
- [20] Emily McReynolds, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. 2017. Toys That Listen: A Study of Parents, Children, and Internet-Connected Toys. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17), 5197–5207. DOI:<https://doi.org/10.1145/3025453.3025735>
- [21] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 6:1–6:7. DOI:<https://doi.org/10.1145/3173574.3173580>
- [22] Chelsea M. Myers, Anushay Furqan, and Jichen Zhu. 2019. The Impact of User Characteristics and Preferences on Performance with an Unfamiliar Voice User Interface. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 47:1–47:9. DOI:<https://doi.org/10.1145/3290605.3300277>
- [23] Donald Norman. 1990. *The Design of Everyday Things*. Doubleday Business, New York.
- [24] Michael Quinn Patton. 2015. *Qualitative research & evaluation methods: integrating theory and practice* (Fourth edition, ed.). SAGE Publications, Inc, Thousand Oaks, California.

- [25] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn. 2013. User-defined gestures for augmented reality. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 282–299. DOI:https://doi.org/10.1007/978-3-642-40480-1_18
- [26] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, 640:1–640:12. DOI:<https://doi.org/10.1145/3173574.3174214>
- [27] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. “Accessibility Came by Accident”: Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*, 459:1–459:13. DOI:<https://doi.org/10.1145/3173574.3174033>
- [28] 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. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*, 2853–2859. DOI:<https://doi.org/10.1145/3027063.3053246>
- [29] Aung Pyae and Tapani N. Joellsson. 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 (MobileHCI '18)*, 127–131. DOI:<https://doi.org/10.1145/3236112.3236130>
- [30] Karen Renaud and Judy van Biljon. Predicting Technology Acceptance and Adoption by the Elderly: A Qualitative study. 10.
- [31] Everett M. Rogers. 2003. *Diffusion of Innovations, 5th Edition* (5th edition ed.). Free Press, New York.
- [32] Karen Rust, Meethu Malu, Lisa Anthony, and Leah Findlater. 2014. Understanding children's mental models for touchscreen tabletop interaction. In *Proceedings of the 2014 conference on Interaction design and children - IDC '14*, 201–204. DOI:<https://doi.org/10.1145/2593968.2610452>
- [33] Alex Sciuto, Armita 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 Designing Interactive Systems Conference (DIS '18)*, 857–868. DOI:<https://doi.org/10.1145/3196709.3196772>
- [34] Fumihide Tanaka, Aaron Cicourel, and Javier R. Movellan. 2007. Socialization between toddlers and robots at an early childhood education center. *Proc. Natl. Acad. Sci. U. S. A.* 104, 46 (November 2007), 17954–17958. DOI:<https://doi.org/10.1073/pnas.0707769104>
- [35] Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Q.* 27, 3 (2003), 425–478. DOI:<https://doi.org/10.2307/30036540>
- [36] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI '09*, 1083. DOI:<https://doi.org/10.1145/1518701.1518866>
- [37] Amazon Echo (2nd generation) – Alexa Speaker. Retrieved September 10, 2018 from https://www.amazon.com/dp/B07456BG8N?tag=googhydr-20&hvadid=251386627342&hvpos=1t1&hvnetw=g&hvrand=407967149010440752&hvpone=&hvptwo=&hvqmt=e&hvdev=c&hvdvcndl=&hvlocint=&hvloctphy=9033307&hvtargid=kwd-295921611050&ref=pd_sl_2g7cb1h5ze_e#tech
- [38] 10 Heuristics for User Interface Design: Article by Jakob Nielsen. *Nielsen Norman Group*. Retrieved August 23, 2019 from <https://www.nngroup.com/articles/ten-usability-heuristics/>
- [39] Tableau: Business Intelligence and Analytics Software. *Tableau Software*. Retrieved February 20, 2019 from <https://www.tableau.com/>
- [40] 3 Approaches for Adding Informal Learning. Retrieved February 24, 2019 from <https://www.td.org/insights/3-approaches-for-adding-informal-learning>
- [41] 5 tips to keep Alexa from peering into your life. Retrieved August 12, 2018 from <https://uw-media.usatoday.com/video/embed/651786002?sitelabel=reimagine&placement=uw-smallarticleatophtml5&keywords=amazon+com+inc%2Calexa+speech+recognition%2Cvoice+command-devices%2Cvoice+activated+digital+assistants%2Cgoogle+home%2Csiri%2Ccortana&simpleTarget=&simpleExclusion=disasters&pageType=story&cst=tech%2Ftalkingtech&ssstech%2Ftalkingtech&series=>
- [42] Elephant-Sized Impact. Retrieved February 24, 2019 from <https://www.td.org/magazines/td-magazine/elephant-sized-impact>
- [43] NEAR-PEER | definition in the Cambridge English Dictionary. Retrieved August 23, 2019 from <https://dictionary.cambridge.org/us/dictionary/english/near-peer>