Mental Models and Home Virtual Assistants (HVAs)

Janghee Cho
Michigan State University
East Lansing, MI, USA
janghee.cho15@gmail.com

Abstract
This study examines how users interact with Google Home, which is a type of home virtual assistant (HVA). Users are expected to speak to HVAs in a conversational manner; however, there has been little research looking at users' mental models for what kinds of interactions they think the devices are capable of. To investigate users' mental models, I conducted user study sessions in which I gave novice users several tasks to complete, and asked them to think aloud as they completed those tasks. I elicited two mental models (*push, pull*) from verbal strategies they use to complete the task. My findings help to better understand why users may be reluctant to use HVAs, and provide design guidance for future conversational interfaces.

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3180286

Author Keywords
Mental models; home virtual assistants; information seeking; conversational interface

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

Introduction
Conversation is becoming a key interface of Human-computer interaction (HCI) as a virtual assistant (VA) integrates into personal mobile devices (e.g., Apple's Siri) and home devices (e.g., Amazon's Alexa or Google Home). VA, so called conversational agents, refer to a dialogue system by "combining automatic speech recognition and speech synthesis with natural language processing and dialogue management" [9, p281].

In many cases, VAs' practical value is still in doubt and limited to a few simple tasks (e.g., setting a timer or playing a music list) [3,9]. Part of the challenge in using VAs is attributed to their opaque operation: by only speaking to control the VA without a visually direct cue makes it difficult for users to formulate a question [13]. Moreover, people often use ambiguous expressions, which connote subtle nuances in communication, and they expect the VA to recognize their conversational manner [14]. There is a gap

	Age	Gender
P1	71	F
P2	38	F
P3	20	F
P4	56	F
P5	48	M
P6	56	M
P7	39	M
P8	19	M
P9	34	M
P10	61	F
P11	21	M
P12	35	F

Table 1: Study participants (earlier twelve subjects, $20 compensation for their participations)

between human's natural utterances, human-to-human communications, and the communication between humans and VAs. This gap interferes with the development of the interaction between users and the VAs.

There are some studies of peoples' perceptions and use of VAs [8,12], but little work has looked at users' mental models for what kinds of interactions they think the devices are capable of. Mental models refer to what people have in their heads that they use to interact with the world around them [5]; these models are sets of beliefs and understandings that help users' decisions [10]. Understanding the mental model that people use in the interaction is important in designing a more effective system for users.

In this study, I aim to understand how users decide what questions to ask and commands to give, and how they pose queries to Google Home, which is a type of home virtual assistant (HVA), in order to elicit users' mental models. In this context, my research questions are as follows:

- RQ1. What types of mental models do users have of HVAs?
- RQ2. What strategies are applied for interacting with HVAs?

How to form mental models

Research in HCI has commonly studied users' mental models to understand how people perceive an object, especially those designed with a complex system. Users get information from other people like themselves, from the media, from communications with experts, and from their own experiences [15]. In other words, mental models differ from the institutionalized, professional, legitimated conceptions held by experts [2]. In this sense, "mental models tend to be functional rather than complete or accurate representation of reality" [5, p5]; the user continues to modify the mental model in order to get to solve a problem through interaction with the system [10].

Besides experience and background knowledge, the structure of the human information processing system is also a component of mental models [10]. Encountering new knowledge in the world, users try to understand it by extrapolating from their own knowledge in their heads [11].

Method

Participants

I recruited 20 native speakers of American English in the Midwest United States through a paid research pool organized by our institution. Participants were screened based on three criteria that helped to select novice users. Eligible participants 1) had no experience with any HVA devices, 2) had little or no experience with mobile-driven VA, and 3) were not experts in VAs' technology.

Procedure

I conducted a qualitative study that was comprised of two phases. At first, I asked participants to think aloud while conducting the tasks. To control for order effects, I randomized the order of the tasks with the help of a Latin square. Also, I did not pose a time limitation to do each task. After completing all tasks, I had a semi-structured interview with general questions that encompassed overall perceptions but avoided

Definition of task types
Subject search: Require to find any pieces of information that are related to the subject and regarded as useful to the participants
Factual: Search specific information
Instrumental: Search for the instrument to know how to use something
Controversial: Check opinion about a certain topic and make own decision
Predictive: Check the prediction of a certain situation

Table 2: Definition of task types

prejudging the answer [4]. Developing follow-up questions based on the participants' answers, I probed for a set of causal beliefs that impact their behavior during the tasks. Lastly, the participants were asked to complete a survey for demographic information. The study took place from November 2017 to December 2017 at the Behavior Information Technology Lab (BITLab) at Michigan State University.

Home Virtual Assistants (HVAs) & Google Home

In this study, I use the term, HVA, to specify a certain domain for a conversational interface, which is the focus of this study. In the previous research, VAs were not categorized according to different domains or purposes. Given the fact that carrying out multiple tasks to save time is the main motivation for mobile-driven VAs [8], it is necessary to distinguish VAs used in the context of the home (i.e., HVAs) from the mobile-driven context in a portable device. For this reason, I focus on HVAs and exclusively use the Google Home device to draw out adequate verbalizing data from participants. In a recent study from the market research firm [1], Google Home had the highest accuracy rate in answer to the most questions. A sufficient length of users' interaction is necessary to observe the process of how users learn and change their approach during the study. Therefore, I was intentionally choosing a device that allows the participants to talk as much as they can.

Task Design

I designed *subject search* [7] tasks with a scenario-based context so participants could easily pretend to seek a certain information. *Factual and instrumental* [6] tasks were developed based on the ability lists on Google Home's website[1] to demonstrate users' practical actions as the usage would be in real life. To design unanswerable tasks, I introduce two additional types of tasks: a controversial task and a predictive task, not referring to the ability lists on the website but reflecting aspects of everyday life. I intentionally included unanswerable tasks to see how participants change their approach to find the answer.

Task	Type	Answerable	Topic
1	Factual	Y	Travel
2	Factual	Y	Shopping (TV)
3	Instrumental	Y	Passport
4	Controversial	N	Obamacare
5	Predictive	N	Basketball

Table 3: The characteristic of each task (The sixth task is open to participants where I allow them to ask anything they want.)

Analysis

To identify mental models, I iteratively analyzed the data by using an inductive qualitative approach. I identified themes and patterns in participants' utterance data. More specifically, I examined how frequently people talk about a certain phrase (wh-question, direct action, or only a word) or a certain concept.

Preliminary Findings

This paper reports preliminary findings from the interviews and the task analysis for four different types

[1] Available at https://support.google.com/chromecast/answer/7130274?hl=en&ref_topic=7195641

Misunderstanding (task5):
No.10: Hey, Google. When does Detroit play again?
Google Home(GH): They will play the Bears in Detroit on Saturday at 4:30PM.
No.10: Is that basketball or football?

Misunderstanding (task2):
No.9: Okay, Google, where can I find a Sony Bravia TV near me?
GH: ..The first one is Sony Pictures Television International at Movie Drive in Brighton.

The burden of memory (task4):
GH: According to Wikipedia, the Patient Protection and Affordable Care Act, often shortened to the Affordable Care Act or nicknamed ObamaCare, is a United States federal statute enacted by the 111th United States Congress and...
No.7: Okay, that just kinda happened kinda fast and it didn't seem to have a lot of detail.

Figure 1. Verbalized data from the tasks

of tasks (task 2,3,4,5) by examining the earlier twelve participants' data.

General Thoughts and Feelings
The participants generally showed ambivalent opinions toward Google Home. Google Home was perceived as a useful device once they learned how to ask questions. However, they thought having a conversation with Google Home was challenging ("*I think that it might be difficult to ask questions properly to get the answer you want, but once you learn how to ask questions, it can give you information that you never knew." – P5).* Some even mentioned it is frustrating and annoying when Google Home misunderstood their questions and gave bizarre responses that are beside the point (Figure.1). They often did not know how to ask questions in a way that Google Home could find the right answer because of inadequate responses ("*Well, I wonder if there's a specific format to use to ask questions and if that's how it's programmed than the user needs to know how to ask the questions if that's what's wrong." - P1).* Also, the participants were struggling to follow up Google Home's long responses with unfamiliar or unrelated topics in order to develop the next question (i.e., the burden of memory). Compared to the usage of Google search on desktop or mobile, they explained obtaining the information by only using a conversational interface was too difficult to obtain further information ("*It's obviously not like Google where they give you 20 million choices and then you can scroll down and pick which ones sound the best." – P2).*

Users' Strategies and Their Models
Based on participants' way to talk, I identified a number of different strategies that applied to their interaction with Google Home. Then, I divided strategies into two high-level categories – *push models* and *pull models* – which seemed to indicate differing approaches to participants' information seeking behavior. The *push models* that are elicited in participants' verbalizing strategies emerged from the formation of the question that participants clarify what they want to find, offering a certain piece of information to Google Home. On the other hands, the *pull models* demonstrate a pattern that participants attempt to extract information from Google Home by using relatively simple forms of questions.

Push Models
One of the strategies in *push models* is to explain personal information or needs to Google Home. The participants applied this strategy, failing to receive an answer by using a wh-question. This strategy is derived from the experience in a normal conversation with a human, which means that they believe articulating a specific personal information or a desire makes the miscommunication clear *("That's what I would talk to my friends about. Right? That's how you give information about yourself." – P4).*

The participants also tended to give more information based on their knowledge when Google Home did not answer by a general wh-question. They believed using a better-known synonym would be of help to get the answer from Google Home ("*I know it's called the Affordable Care Act, I didn't know if they would know what Obamacare was." – P10).* Based on their memory, they also specified their questions by adding a representative service or place. They thought it would help Google Home understand more clearly what they

Push models' examples:

"I already have health care, so I don't need Obamacare?" (P10)

"Detroit Pistons are a basketball team. Do you know anything about them?" (P6)

"So, maybe if I ask about Sports Illustrated predictions for the basketball playoff season, that might help if Google Home could access that. Hey, Google, can you tell me who Sports Illustrated predicts will make it to the NBA playoffs in 2018?" (P12)

Pull models' examples:

"Can you give me your opinion on anything?" (P11)

"What do you know about ObamaCare?" (P9)

"Tell me the highest rated TVs." (P8)

Figure 2. Each model's examples from the tasks session

were looking for *("Because I know that you can buy a television at Best Buy. That's where I would go." – P5).*

Some participants thought they could develop Google Home's capability by providing a certain idea related to the task. They believed Google Home harnesses the power of Artificial Intelligence so that it is actually able to learn by being coached by the user on a specific topic *("Like a learning capability or some sort of AI in it that it can. Assuming that it has the capability, then yeah, I think it would learn over time what you're asking for..." – P9).*

The participants sometimes changed their way of speaking more slowly and clearly. They believed Google Home's failure was due to their voice so that altering their enunciation or repeating the same sentence would help the device to understand their question *("…I didn't know if I wasn't getting the answers I wanted 'cause I was talking too fast." – P12).*

Pull Models

The participants usually asked questions to see whether Google Home is able to draw out an answer for a particular type of information. For instance, they posed the question to pull out any information Google Home had *("I didn't know if they would know what Obamacare was." – P.10*). They applied this strategy to only unanswerable tasks after failing several times. They were doubtful about Google Home's ability or knowledge on a basic topic *("it wasn't answering my question. And so I wanted to see what it knew about the NBA." – P5).* By using this strategy, participants took a step backward and tried to think broadly about how to formulate the right question.

Using a simple imperative form of the question, instead of wh-question, is also another strategy to derive information from Google Home. While some participants used this strategy reached a deadlock, the other participants used it at the beginning of the task. They explained they had learned this strategy from a HVA's commercial *("I just feel like I've heard that said to Google Home before, maybe like commercials or whatever. That's the way that they want you to use it..."– P11).*

Discussion and Conclusions

My preliminary findings indicate that the participants generally expect Google Home can provide an answer if they input a specific piece of information. Although they believe speaking in a simpler sentence is much easier to interact with Google Home, they tend to speak in a human conversational manner. However, I found participants learned how to communicate with Google Home, rephrasing other sentences by using different strategies. In this sense, an error handling strategy in HVAs is seen a critical factor to avoid users' reluctance to use them. Participants were disappointed in the way how Google Home gave an answer to explain its inabilities or limitations (e.g., "I don't know how to help with that"). They wished Google Home would clarify what users just said to know how to formulate their questions and what was wrong with the previous question when Google Home could not answer. Moreover, having a minimum memory to understand conversational context is necessary to avoid bizarre responses. Luger and Sellen (2016) found that users' unmet expectations in their interaction with virtual assistants hinders their use of the virtual assistant. Such misunderstandings could worsen the users' effort to keep having a conversation than non-understandings.

In this study, I show what mental models guides users' actions and how people develop their chain of causation influenced by interaction with Google Home. Analyzing additional data collected from eight participants, I will verify and extend key mental models to envision what future interactions with conversational agents might be like.

References

1. ERIC ENGE. 2017. Digital Personal Assistants: Which Is Smartest? | Stone Temple. Retrieved December 29, 2017 from https://www.stonetemple.com/digital-personal-assistants-test.
2. Motahhare Eslami, Karrie Karahalios, Christian Sandvig, et al. 2016. First I "like" it, then I hide it: Folk Theories of Social Feeds. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 2371–2382.
3. Milica Gašić, Dilek Hakkani-Tür, and Asli Celikyilmaz. 2017. Spoken language understanding and interaction: Machine learning for human-like conversational systems. *Computer Speech and Language* 0: 5–7.
4. M. Granger Morgan, Baruch Fischhoff, Ann Bostrom, and Cynthia J. Atman. 2002. Mental Models Interviews. *Risk communication. A Mental Models Approach*: 63–83.
5. Natalie a. Jones, Helen Ross, Timothy Lynam, Pascal Perez, and Anne Leitch. 2011. Mental Model an Interdisciplinary Synthesis of Theory and Methods. *Ecology and Society* 16, 1: 46–46.
6. Jeonghyun Kim. 2006. Task as a predictable indicator for information seeking behavior on the Web. *ProQuest Dissertations and Theses*. Retrieved from http://search.proquest.com/docview/305277561?accountid=14643%5Cnhttp://mlbsfx.sibi.usp.br:3410/sfxlcl41?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+Global&atit.
7. Kyung-Sun Kim and Bryce Allen. 2002. Cognitive and task influences on Web searching behavior. *Journal of the American Society for Information Science and Technology* 53, 2: 109–119.
8. Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*: 5286–5297.
9. Roger K. Moore. 2017. Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. *Lecture Notes in Electrical Engineering*, 281–291.
10. Donald A. Norman. 1983. Some Observations on Mental Models. In *Mental Models*. 7–14.
11. Donald A. Norman. 2013. *The Design of Everyday Things. Revised and expanded edition*. .
12. Martin Porcheron, Joel E Fischer, and Sarah Sharples. 2017. "Do Animals Have Accents?": Talking with Agents in Multi-Party Conversation. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 207–219.
13. Stuart Reeves. 2017. Some Conversational Challenges of Talking with Machines. *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17 Companion*, 431–436.
14. Xin Rong, Adam Fourney, Robin N. Brewer, Meredith Ringel Morris, and Paul N. Bennett. 2017. Managing Uncertainty in Time Expressions for Virtual Assistants. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 568–579.
15. Rick Wash and Emilee Rader. 2011. Influencing mental models of security. *Proceedings of the 2011 workshop on New security paradigms workshop - NSPW '11*, 57.