"I don't own the data": End User Perceptions of Smart Home Device Data Practices and Risks

Madiha Tabassum[1], Tomasz Kosiński[2], Heather Richter Lipford[1]
[1]University of North Carolina at Charlotte, [2]Chalmers University of Technology
{mtabassu, Heather.Lipford}@uncc.edu, tomasz.kosinski@chalmers.se

Abstract

Smart homes are more connected than ever before, with a variety of commercial devices available. The use of these devices introduces new security and privacy risks in the home, and needs for helping users to understand and mitigate those risks. However, we still know little about how everyday users understand the data practices of smart home devices, and their concerns and behaviors regarding those practices. To bridge this gap, we conducted a semi-structured interview study with 23 smart home users to explore what people think about smart home device data collection, sharing, and usage practices; how that knowledge affects their perceived risks of security and privacy; and the actions they take to resolve those risks. Our results reveal that while people are uncertain about manufacturers' data practices, users' knowledge of their smart home does not strongly influence their threat models and protection behaviors. Instead, users' perceptions and concerns are largely shaped by their experiences in other computing contexts and with organizations. Based on our findings, we provide several recommendations for policymakers, researchers and designers to contribute to users' risk awareness and security and privacy practices in the smart home.

1 Introduction

Internet-connected utility devices, called smart home devices, are starting to proliferate throughout households thanks to a growing selection of available devices along with decreasing prices. From lights to thermostats to whole sets of sensors and actuators, users can now enjoy home automation and hands-free control. Yet to provide this functionality, smart home devices greatly expand the types and amount of information about ourselves and our environments that can be collected and shared. The security of our homes is also now becoming reliant on the security of our digital home devices. Thus, with this new domain come new risks to users' security and privacy. And new questions as to how to support users in understanding, reasoning about, and mitigating those risks.

Research on mental models of the Internet has demonstrated that users are uncertain about how their data is collected, shared and stored online, and that users' perceptions often depend on their personal experiences and technical education [12]. Smart homes are even more interconnected, with a wider variety of personal data collected from people's homes. Thus, we seek to examine in greater detail users' perceptions in this more complex environment of the smart home. Our work also builds on previous interview studies, primarily of technically skilled smart home early adopters, examining general privacy and security perceptions [33] and concerns regarding specific data collection entities [34] in the smart home. We focus on users' mental models of the data practices of their smart home devices, and their related privacy and security perceptions.

Specifically, we conducted a drawing exercise and semi-structured interview with 23 participants who have experience living with multiple smart home devices. We focused on recruiting both more technical participants who installed their devices, as well as non-technical users who were not involved in the installation process.We investigated the following research questions: (1) What are end users' mental models of the data flows in their smart home? (2) What are end users' perceptions of the data collection, sharing, storage and use by smart home devices and their manufacturers? (3) How do these mental models and perceptions relate to users' privacy and security concerns, considerations and behavior?

We found that the sophistication of participants' threat model and the adoption of protective measures do not depend on their knowledge of how their smart home works. While participants mentioned some threats and protective measures, they often estimated the privacy and security risks from their smart home devices to be too low to trigger any actions.

Our study makes the following contributions:

- Provides a thorough analysis of both technical and non-technical users' perceptions of smart home device manu-

USENIX Symposium on Usable Privacy and Security (SOUPS) 2019.
August 11–13, 2019, Santa Clara, CA, USA.

facturers' data practices and related threats that offer new insights and also confirm, explain, and extend findings from previous studies.

- Among other findings, our results provide new evidence that people are moderately aware of the sensitive information that can be inferred from smart home data, however, are not concerned over the collection and sharing of this data.
- Participants' lower-risk perceptions are shaped by trust, previous experiences within other computing contexts, and their own biases, despite their concerns over the lack of control over their information.

Based on our findings, we provide recommendations for smart home designers, researchers, and policy-makers to provide improved awareness and control of data collection practices and protection strategies, considering the perceptions and capabilities of general smart home users.

2 Related Work

User mental models have been explored in the context of the Internet and software. Usually these have been explored from a task- or tool-specific perspective, such as understanding of the operations of WiFi networks [15], general home computer security [31] or firewalls [14, 28]. For example, Kang et al. explored user mental models of the Internet in general [12], also asking users about their perceptions with regard to data practices on the Internet. They found that participants with a more accurate understanding of the Internet identify significantly more privacy threats than participants with simpler models, but that does not influence their protection behaviors. We believe many of these models will carry over into the smart home. Yet, the smart home is more complex, and is more integrated with people's personal lives, introducing new and unique security and privacy risks. Thus, we aim to examine the mental models of smart home users specifically, focusing on the perceptions of data practices of end users.

A number of researchers have examined end user concerns, expectations and preferences with smart devices in the home. However, early work relied on prototypes or probes within homes to examine users' reactions and perspectives, given the limited availability and adoption of smart home devices at the time. For example, Choe et al. [6] used sensor proxies in 11 households as a cultural probe and found participants had concerns about unintended use of their data and the possibility of data exfiltration. They also found tensions between different members of a household around the use and adoption of such in-house sensing applications. Worthy et al. [32] installed an ambiguous Internet of Things(IoT) device in 5 participants' homes for a week and found that trust in the entities that use the data (in this case, the researchers) is a critical factor in the acceptance of the smart device. Montanari et al. [25] invited 16 participants to interact with two smart home devices during the study session and found that users are primarily concerned with the ownership of their data.

A number of studies have also examined the role of context in users' comfort of sharing IoT-related data. These studies reveal that privacy concerns are indeed contextual, depending on a variety of factors such as the type of data recorded, the location where it is recorded, who the data is shared with, the perceived value of the data and benefits provided by services using that data [5, 7, 10, 15, 18, 19, 21, 23, 26]. Naeini et al. [26] used vignettes to study many of these factors with over 380 different use cases across 1,000 users. Their results indicate that people are most uncomfortable when data is collected in their home and prefer to be notified when such collection occurs. Similarly, a survey study by Lee and Kobsa [19] found that monitoring of users' personal spaces, such as their homes, was not acceptable to participants, as well as monitoring performed by the government or unknown entities.

Other studies have found that people are most concerned with certain types of data, namely videos, photos, and biometric information, particularly when this information is gathered inside the home [4, 9, 19, 20, 26]. In another large vignette study, Apthorpe et al. [5] found that participants' acceptance of data collection and sharing was dependent on both the recipient of the information and the specific conditions under which the information was shared. Their results also suggest that users' privacy norms may change with continued use of specific devices. However, results of a different vignette survey by Horne et al. [11] suggest that those changes are not always towards more acceptance of data-sharing. Each of the above studies examines fine-grained contextual factors through survey methods of potential use cases of smart home devices. Despite these findings showing significant concerns over data collection in the home, many users are installing smart home devices that do collect and share such information. These prior studies have not revealed what adopters of current devices think is actually occurring, and their comfort and concerns with those practices.

With widespread adoption, several studies have recently examined the perceptions of users of consumer devices that they use in their own homes and found less concern by actual, regular users. Lau et al. [16, 17] conducted a combination of a diary and interview study with 17 users and 17 non-users of smart voice assistants. They found that the lack of trust and perceived utility are the main reasons for not adopting the device. They also noticed that adopters of the voice assistant have an incomplete understanding of the privacy risks and rarely use existing privacy controls. Most similar to our study, Zeng et al. [33] conducted an interview study of 15, primarily technical, smart home users and observed limited concern among participants about the potential improper use of their data. They also found that even relatively technical participants have an inaccurate or incomplete understanding of smart home technology, resulting in incomplete threat

models and adoption of insufficient mitigation techniques to resolve potential threats. Zheng et al. [34] interviewed 11 technologically skilled smart home users on their reasons for purchasing smart home devices and the perceptions of privacy risks from these devices. They found that users' concerns over specific external entities (i.e. government, manufacturers, internet service providers and advertisers) are influenced by the convenience they get from the device and those entities. While these two interview studies highlight many general concerns of users, and feelings about data being accessed by specific entities, we believe that a more detailed understanding of users' perceptions of the data practices of their smart home devices is critical to understanding users' behaviors and needs. In addition, these prior studies relied primarily on technically knowledgeable participants who actively set up their smart home and are interested in technology, which may limit generalizability of their results.

3 Methodology

We conducted a semi-structured interview study and drawing exercise of smart home residents to elicit their mental models of the data practices of smart home devices, along with their perceived security and privacy risks and concerns.

3.1 Participants

We sought participants who are regular users of smart home devices and thus had mental models of the smart home ecosystem informed by their usage. We recruited participants with at least three devices, similar to Zheng et. al. [34]. We explicitly recruited some participants who did not install the devices themselves (such as family members) to find people who are not as tech-savvy and may have different privacy perceptions. The participants were recruited through advertisement on Craigslist, and IoT-related Reddit communities. Potential participants were asked to fill in a pre-screening survey answering what types of devices they have in their home, whether they set up the devices by themselves as well as demographic information and email address. We recruited participants until we felt we had a sufficiently diverse sample, and then found we reached saturation (i.e., no new codes or new information attained) during analysis, and hence did not seek additional participants.

We recruited a total of 23 participants (see Appendix A.3). Six of them had a background in computer science, either as a student, or as a computing professional or both. 13 participants were male and six were more than 51 years old. All participants were living in the United States, except one in Canada and one in Sweden. 11 participants installed and manage the devices in their home, 3 participants installed some of the devices and 9 were not involved in the installation and configuration process at all. Not surprisingly, participants who installed their devices self-reported a higher level of familiarity (statistically significant) with technology and smart home security and privacy, than users who did not perform the installation. We acknowledge that there can be tech-saavy non-installers; however, we did not find such participants in our study sample.

3.2 Procedure

The researchers contacted selected participants via email to schedule a phone or Whatsapp interview. The interview was semi-structured, with a set of basic questions that were varied depending on the response of the participants. The interviews were recorded via Google voice or an external audio recorder. Interviews lasted on average an hour and participants were given a $10 Amazon gift card for participating. The study was approved by our university Institutional Review Board (IRB).

We started the interview by asking general questions on what smart home devices participants have, and how they use and control those devices. Participants were then instructed to perform a drawing task to elicit their understanding of how their smart home works. Participants were asked to "draw how these devices collect information and how that information flows between the devices and any other involved entities" and to explain their thoughts verbally during the drawing exercise. This has been used as an effective method in capturing mental models in the literature [12, 33]. We utilized remote Google drawing as it was accessible to most of the participants and has been used previously for remote drawing tasks [33]. This could impact the drawings, as the participants utilized shapes and lines rather than free-form strokes. However, participants explained their drawings as they were creating them, similar to an in-person interview. Only 2 participants sent pictures of their drawings via email during the interview because they felt more comfortable drawing on paper. However, after sending the drawing, participants extensively talked about what they drew. We recognize that a drawing exercise in a remote interview is challenging, but we feel the trade-off in finding a more diverse sample was worth it.

We then focused on participants' perceptions of data practices, asking the participants what data they think the smart home devices they own are collecting and where these devices are sending and storing that data. Participants were then prompted to discuss who they think has access to their data and how it is being used, as well as whether the devices are sharing the information, with whom and for what benefit.

Next, we asked participants if they have any concerns regarding those data practices. We then asked them what they do to mitigate their concerns and resolve the threats that they think arise from using their smart home devices. We discussed what controls the participants believe they currently have over the data the devices are collecting, what controls they expect to have and their expectations regarding the security of their data. Finally, we collected participants' demographic infor-

Type of device	Count	Examples	Users' perception of information collection
Intelligent voice assistant	20	Google Home, Amazon Echo	Voice interaction (20); Usage (10); Account info (5)
Smart light	16	Philips hue, LIFX, Sangled	Patterns & usage (11); State of the lights (10);
Smart plug and switch	13	Wemo, Tplink, Insteon, Sonoff	Account info (5); Home location (2)
Smart camera/doorbell	11	Nest Cam, Ring, SkyBell Doorbell	Video (11); Home location (4); Usage (3)
Smart thermostat	11	Nest, Ecobee Thermostat	Temperature (10); Usage (5); Energy use (3); Account info (2)
Hardware hub	8	Samsung SmartThings, Wink hub	Usage (6), Location (3), Other devices in the network (2)
Streaming device	8	Roku, Fire Sticks, Chromecast	Viewing history (4); Account info (3)
Other devices: Smart TV (5), Leak sensor (4), Smart Doorlock (3), Open/close sensor (3), Motion sensor (3), Smoke detector (2), Smart media hub (2)			

Table 1: Summary of the devices owned by participants. Numbers in the parentheses are number of participants

mation at the end of the interview. Interview questions are provided in Appendix A.2.

3.3 Data Analysis

We transcribed the interviews and used an inductive coding process to analyze the data. Two researchers independently coded the interviews of five participants and came up with a list of common themes and patterns. Then the researchers compared and merged the themes and agreed on a shared codebook with 15 structural codes divided into 60 sub-codes. The two coders then independently coded the rest of the interviews. After all the interviews were coded, the researchers met and discussed the codes, resolving any disagreements caused by misunderstanding the codes. We tracked the disagreements and the Cohen's Kappa, a measure of inter-rater reliability, was calculated at 96.37.

The participants' drawings and related verbal explanations were separately analyzed by the primary author, who clustered similar drawings and conceptions into two emerging categories. The clustering was performed based on the complexity of participants' mental model about both the physical architecture of their smart home and corresponding data flows throughout the system. These categories were then discussed among all the authors, and used to examine differences between participants' perceptions throughout the results.

3.4 Limitations

As with similar interview-based studies, we consider sample size to be the biggest limitation of this work. We can only provide limited qualitative results on the posed research questions, yet hope that those revealed patterns can be used in formulating further studies of more representative populations and to inform design. We also believe that the participants, even the non-technical ones, that we interviewed are still the early adopters. They are clearly well educated, and likely of high socio-economic status. They also value the benefit of the devices and decided to have them in their homes. Hence, they have already made the decision that the trade-off is worth the risk; therefore they may not have as many concerns as non-adopters. Thus, these results may not generalize to a broader consumer base who will adopt smart home devices in the future. Still, we hope that many of these patterns would be found in a more general population as we found many of the perceptions did not differ between participants of different levels of expertise. Another limitation is that this was a one time interview, which entails the risk of missing participant concerns that could be discovered in, for instance, a longitudinal study. Finally, almost all of our participants are from the U.S. and may have a different perspective about privacy from other regions. Because we have only two participants from other countries, it was not enough to identify those differences.

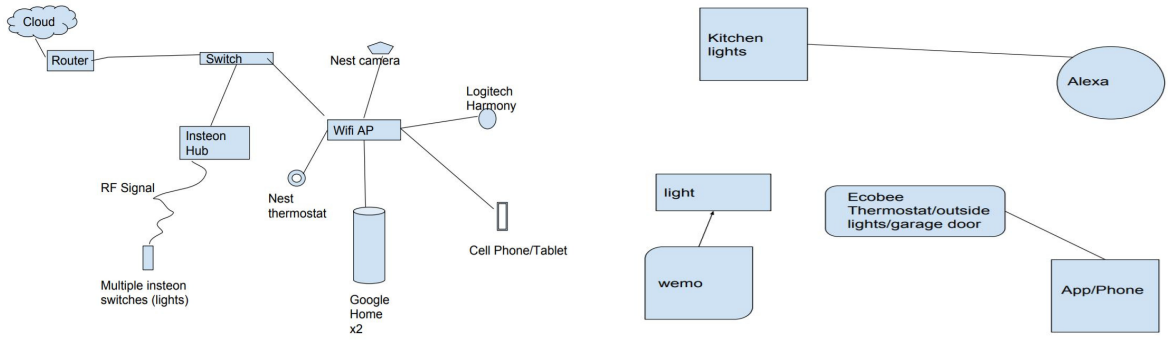
4 Results

Our study goals are to examine users' perceptions and concerns of the data practices of smart home devices. First we describe the devices they have and use, then present the results of our analysis of participants' mental models, their perception of manufacturers' data practices and their related security and privacy concerns and behaviors. Please note that the numbers reported below are not meant to convey quantitative results, but simply reflect the prevalence of particular themes within our experimental sample.

4.1 General Use of Smart Home Devices

Participants own a wide variety of internet-connected devices, including integrated devices (lights, thermostats), home monitoring and safety devices (security cameras, door locks), home appliances (vacuum cleaners, smart refrigerator), and intelligent personal assistants (Google Home, Amazon Echo). We summarize the common devices in Table 1. Participants use these devices in a number of ways. The most frequently mentioned ($n = 11$) use case is household automation (automatically turn on/off the lights, adjust the temperature, etc.), followed by remotely sensing and controlling the home ($n = 10$) (i.e. to turn on/off the lights, check on pets). Another use case ($n = 9$) is increasing the security or safety of the house (by notifications of conspicuous sounds in the house, water leakage, etc.). Other less frequently mentioned use cases are energy saving and help with household chores.

We also asked participants how they interact with their devices. Participants use several different methods, often in combination, depending on the location of the user within or outside of the home, as well as the type of device and



(a) Advanced model (Id9)

(b) Service-oriented model (Id20)

Figure 1: Drawing of participants with different mental models

its compatibility with a controller. Almost all participants ($n = 21$) have a central controller set up, i.e. either a smart voice assistant, hardware hub, an app (e.g. Apple Homekit) or a custom-made controller using the Raspberry Pi. For 13 participants, voice is the primary method of interaction when they are home, utilizing either Amazon Echo or Google Home. Some participants also mentioned setting up triggers based on other sensor data or timers to make devices fully automated (i.e. using IFTTT services).

4.2 Mental Models of Smart Home

Our analysis shows that participants with different technical backgrounds and experiences with the devices have different mental models of how their smart home works. We asked users to describe how data flows in their smart home, and participants chose different ways to express this. We grouped based on similarity of participants' understanding of how devices are connected and how information flows in the smart home and this resulted in our categorization. Two models emerged: advanced (9 participants) and service-oriented (12 participants), based on participants' drawings and verbal explanations of their smart home. We did not include Id6 and Id21 in our categorization. The recording of Id6's drawing explanation was distorted, and Id21's spouse was helping her with the drawing during the interview.

Participants with the advanced model consider their smart home as a complex, multi-layer system. These participants have a reasonable understanding of the logical topology of the smart home, connection mechanisms (Ethernet, WiFi, ZigBee/Z-Wave) and the role of some network components (routers and hub) in communication (Figure 1a). All the participants with this model also discussed how data flows back and forth between the devices and servers in the cloud when interaction happens. For example, Id19 said,

"*When its (Echo) not being used, it is just waiting for one of four trigger words and when that triggers, then it opens up the connection back home(Amazon) and start parsing out the commands for different devices and passes it along to the smart things which takes over from there.*"

Participants with the advanced model discussed how information flows through the infrastructure as well as to the companies' servers and comes back to the device. These participants personally installed all of their smart home devices. Also, a number of them did some customization in their smart home, i.e., used IFTTT to automate the devices, installed a personal server or built a central assistant using Raspberry Pi. It can be one of the reasons behind their more comprehensive understanding of the network topology. Additionally, these participants are also more informed of the complexity of the flows as well as the fact that devices are sending information to companies' servers as soon as they interact with them.

Participants with the service-oriented model (n=12) have a reasonable understanding about which devices communicate with each other inside the house, but do not have deeper technical knowledge of how that communication happens other than via the WiFi. Their mental models of the smart home mostly consist of the interaction between the smart devices (i.e. lights) and the controller (e.g. Google Home) they use to control the device, but no awareness of the role of other networked components in the device interaction (Figure 1b). There were a few participants in this group who brought up that information is going to the cloud initially when drawing their smart home; the other participants didn't. However, when asked directly during later interview questions they all indicated that information the devices are collecting is not stored locally, it is leaving their home to the cloud or some server. However, the participants with the service-oriented model expressed no or very shallow awareness of the role of the cloud in the device interaction.

4.3 End Users Perception of Data Practices

In our analysis, we found that participants' mental models of how their smart home devices work do not often relate to their perceptions of smart home device data collection, usage, and sharing practices. Rather, their understanding is primarily based on interactions they have had with the devices

or what they see in the corresponding applications. Some participants ($n = 10$) beliefs of data practices were informed by their perception of particular companies and experiences with those companies in other (non-smart home) contexts, which sometimes leads to inaccurate conclusions on what really happens. For example, Id11 thinks that companies will not sell any data because it would upset their consumers and the companies would lose reputation. Only two explicitly reported privacy policies as a source of their knowledge about data practices. Below we discuss the findings on end users' perceptions of data practices in detail.

4.3.1 Data Collection:

Not surprisingly, participants' perception of data collection was informed by the type of the device and their experience with that device. For example, all participants who have a smart doorbell or camera were aware of the device collecting video recordings, but none demonstrated any awareness of the corresponding applications tracking their location whenever they use it. In other words, participants were well aware of the primary data that the device is collecting but may overlook secondary data that does not directly correlate with the type of device or basic utilities received from the device. In Table 1, we summarize user perceptions of what information is collected for different devices. Only the devices owned by more than five people are listed.

For most of the devices, participants believe that their usage and interaction patterns are being recorded and they were not that concerned about the data collected by the smart home devices. We also looked more specifically at audio and video data since previous studies [26, 34] found that these data are more sensitive to people. When put in practice, video is still considered as the most sensitive; however, participants for the most part were able to find practices that allowed them to be comfortable with the collection of video. For example, using the camera in a live streaming mode without recording the video, starting video recording only when the house is empty, or using an outdoor camera or video doorbell, so nothing inside the house gets recorded. For instance, Id7 mentioned:

"*Only information I would potentially ever be concerned with, like the way I use my device, is the images on the camera. But again, the camera is turned off when I am home, and on when I am not home.*"

However, in one extreme case, a participant (Id22) removed an indoor camera from her apartment. She reported being unable to use the camera outside because of concern of residents of her apartment complex. She was not aware of other alternative configurations for her camera such as using the camera only for live streaming or removing the recordings from the cloud, which she may have been more comfortable with.

Participants did not show much concern about the collection of their audio data. They know that the voice assistant is recording after the trigger word, and they were comfortable with the audio being recorded in that way. One non-installer participant (Id20) was uncomfortable with the voice assistant as she suspected that Amazon Echo may be listening to her even if she is not calling it using the trigger word. Her Echo had recently showed an Amazon package delivery notification with yellow lights, and she misinterpreted it as the device listening to her conversation. Even though her husband later clarified the misunderstanding, the participant was still very uncomfortable and did not want the device in her house at the time the interview was conducted. Another technical participant decided not to buy any commercially available voice assistants because of a worry over companies harvesting the audio.

Despite their awareness, many participants ($n = 15$) believe smart home devices are collecting more information than they should. However, some participants ($n = 9$) said the data collection was mostly positive. These participants explicitly mentioned that most of the data these devices collect is needed in order to either provide them the services they expect or to make the devices more convenient to use.

We also asked participants what can be inferred about them from the data these devices are collecting. In contrast to a previous study [34], we found that participants are somewhat aware of the sensitive information that can be inferred from the seemingly innocuous data collected by the smart home devices. For instance Id11 mentioned,

"*They can probably tell that I don't have the lights everywhere at my home. That I am out of the house during the day time. They can probably tell when I am sleeping because the lights are not turned on that time.*"

The types of inferred information that were mentioned are: habits and preferences (i.e. buying habits, music preferences, etc.; 14 participants), daily schedule (i.e. when home or not, when using which devices, etc.; 11 participants), tentative location of the house (8 participants), other occupants in the house (i.e. have pets or kids; 3 participants), political views (2 participants), sleeping patterns (2 participants) and other devices in the house (2 participants). Three explicitly mentioned that these companies can infer a lot of things that consumers can't even imagine.

4.3.2 Data Storage:

When prompted, all participants reported being aware that at least some of the smart home device data is being stored externally, with twenty specifically mentioning the cloud or a server operated or owned by the manufacturer of the device. However, three other service-oriented participants expressed a vague idea such as 'somewhere in some kind of database.' For example, Id15 said:

"*I don't know really where it goes or what happened to it but I imagine that it does get stored somewhere, some kind of database and somebody is able to analyze and see different trends through it. But I have no idea.*"

Eleven participants explicitly mentioned there is either no or very limited local storage of the data, that everything is stored in the cloud. Participants frequently mentioned they have no control over the data once they shared it; however, some ($n = 5$) hypothesized that it might be possible to remove their data by contacting the device manufacturers. Interestingly, 4 participants suspected that even if they remove the data, it will still be in the cloud. A number of participants ($n = 8$) also mentioned companies are doing the bare minimum to protect their consumers' data in the server. Most participants were not sure about companies' data retention practices except for the retention period of the video. Some of them also made interesting inferences, for example, five participants believe that Google and Amazon store data forever or for a very long time because these companies have enough resources to store such data, while smaller companies do not.

Interestingly, all the participants who installed the camera or video doorbell themselves ($n = 7$) know about the video deletion option or after how many days the video will be automatically removed from the server. On the other hand, participants who have not installed ($n = 3$) the camera or the video doorbell are not sure about the storage policy of the video or the option of deleting the video. Video is the one exception where some participants are very aware of the data storage practices and available controls, but only those who installed it, and as a result, they found practices that they were comfortable with and configured their device accordingly. But, participants who are not the installer did not get that understanding, which in one case led to a lot of discomfort and removal of the device.

We did not find as many difference between installers and non-installers regarding their knowledge of data storage policies and controls provided by the devices that collect audio. Out of 20 participants who had a smart voice assistant, 15 are familiar with the device usage log where they can review their voice interactions with the assistant. However, some of them either are not familiar with the data deletion option ($n = 5$) or skeptical that Amazon or Google may keep the data even after they delete it from the log using the available interface ($n = 4$). However, all the participants who did not know about the device usage log were also not involved in device installation. For one participant, this lack of awareness also lead to more discomfort about using the device, as stated by Id20:

"*I have asked my husband to disconnect the Alexa(Echo) multiple times. Just because I'm not comfortable with it. But if it did collect data, I would have no idea how to find it and to remove it so I would just disconnect it.*"

4.3.3 Data Use:

Participants discussed three primary uses of the data their smart devices collect. The most frequently mentioned use case is targeted advertising or marketing to sell products to consumers ($n = 19$). For instance, Id19 said:

"*They have put a lot of money in this product, and then they are selling it. So, they must be using it for something other than me telling my house to turn on my bedroom light. They are building advertising model of me. They want to know who I am and how I work so they can try to sell me something.*"

Participants were aware that their habits, preferences, and daily schedules can be inferred from the data smart devices are collecting and can be used for targeted advertising. However, targeted advertising seems to have became so integral to participants' lives that they accepted it as a price of living in the age of the Internet.

Many ($n = 17$) mentioned that the companies are using the data to improve the current product, for instance by fixing malfunctions/errors (4 participants), improving the user experience or tailoring the device to customers needs (4 participants) or improving the services provided by the device (2 participants). As Id7 stated:

"*(Companies use the information) in order to better the products I guess. I guess if there are errors like you know if I ask Google Home to do something, and the lights don't respond, they're surely collecting that kind of information*"

A number of participants ($n = 9$) also believe that the information companies are gathering can help them to recognize users' needs and come up with new products.

4.3.4 Data Sharing:

Participants identified a number of entities that they believe have access to the data their smart home devices are collecting: the manufacturer of the device/the data analysts working with the company ($n = 23$); third parties/advertisers interested in the data ($n = 9$); parent companies, subsidiaries or affiliates of the device manufacturers ($n = 7$); hackers ($n = 7$); legal organizations such as government security agencies ($n = 4$); the manufacturer of the device/app that is used to control the device ($n = 3$) and other people who have accounts with the device ($n = 2$).

We then asked participants if they think companies share any information with third parties. Twenty-two participants agreed that they do. Nine further believe that companies are sharing only their demographics or preferences but not any personal information; however, 4 participants mentioned they believe companies are sharing everything. Participants also made interesting inferences about how the sharing happens, such as that the big companies (Google, Amazon, Apple) do not share data at all while only the small companies share their consumers' data (6 participants). For example, Id8 said:

"*I think Amazon would be like the top consumer of this information; I think they're collecting this for themselves. I don't think they would share it. I think a smaller company... if the Ring wasn't purchased by Amazon, I think Ring might share that information with Amazon...I have a feeling that's why Amazon bought them.*"

Most of the participants ($n = 18$) said they agreed to this

sharing by signing the terms of service or privacy policy or saying 'yes' to everything during the installation process. But similar to previous research [13, 22], participants reported not reading privacy policies and pointed out the usability issues of such agreements. Three service-oriented participants believe they consented just by using the product. Some participants ($n = 9$) stated that once the data is sent to the cloud, it is out of their hands and control. Id12 stated:

"I'm sure they do... absolutely they do it (share data)... they are allowed to do that...they can do whatever they want with it, that data is considered as their property. They can keep everything for their own or they share."

Many participants reported that the only way they can opt out from this sharing is to stop using the product ($n = 15$), while a few mentioned modifying the applications' settings for partial opt-out ($n = 4$) or by contacting the company (n=2).

To summarize participants' perceptions of data practices: they base their understanding of what data is collected on their experiences and interaction with the devices. For the most part, they expect that their data resides in the cloud and that it can be and is shared by companies, with little ability to control that. However, participants expressed a great deal of uncertainty when they discussed the ways companies are collecting, using and sharing their data. The only exception is the video data where all the participants who installed the device were aware of where the video is stored and video retention time. Several participants ($n = 5$) explicitly expressed their concern about companies not being transparent enough about their data practices. Many participants mentioned that they want more transparency from the device companies ($n = 14$). For instance, Id9 said:

"*If these companies are sharing my data with third parties, I'd like to know who they are sharing with, maybe like if I go to the Insteon website they say, hey we share your data here. So a website that keeps track of all this stuff would be good.*

Participants also want companies to take enough measures to ensure their data is protected ($n = 9$). A few participants ($n = 4$) also believe there are not enough regulations in place and that policymakers should enact and enforce more strict laws to protect consumer data. Finally, ten participants expressed the desire to have explicit control over data collection and sharing and to be able to remove their data from the cloud.

4.4 Security and Privacy Threats and Consequences

We now turn to participants' perceptions of the risks and behaviors for protecting their information. Participants identified several threats and discussed how these affect their security and privacy. However, we again could not find many differences between participants with different technical knowledge levels and mental models. Instead, many of the concerns participants mentioned came from their experiences with the Internet, computers and mobile phones instead of threats specific to smart home devices.

4.4.1 Threats:

The most concrete and frequent threat mentioned by participants ($n = 17$) is a data breach in the cloud and their personal information being compromised. Two participants also suggested hackers could gain access to aggregated profile data from the cloud. Id2 stated his concern as:

"*I mean especially the states of data breaches lately. That is concerning because they're not viewing in a way that hey, these are actual consumers out there, these are real people. Then they may not have the best security practices, and that data can get out somewhere.*"

Some participants ($n = 11$) also pointed out that their smart home devices or the WiFi can be hacked and remotely controlled by adversaries for various reasons, i.e. to spy on them, break into their house, etc. For example, Id19 said:

"*someone could access my lights, someone could turn my heat up ... umm ... if I had a smart lock, someone could have access that to get in my house but I don't have a smart lock. Just like I wouldn't use banking through any of these devices because the consequences are too severe in case there was a breach... the same with a lock, I wouldn't use one of those.*"

Six participants also identified improper use and sharing of their data with third party companies as a potential threat. Unlike data breaches and device hacking, participants were more vague about this threat, i.e., third party companies may use my data for some nefarious reasons or their server may not be secure, etc. Id12 said:

"*The person you shared that data with can share the data with somebody else. Like if you shared data with the company that follows all the rules and if they share with a company that doesn't follow any rule that is out there. I don't think these companies have any methodologies in place to ensure that whether their partner will maintain the data safety or not.*"

4.4.2 Consequences of the threats:

These threats were then associated with specific negative outcomes. Similar to the concerns expressed in previous papers on smart homes [33, 35], participants most frequently mentioned the violation of their physical security and safety ($n = 10$). They implied that smart home devices know when they are home or not, and what other devices they have in their home, and that this information can be used to rob them or physically harm them. Id3 mentioned:

"*I guess if it was a criminal group like a gang or something they could use that data to know when I'm home or not home. If they want to rob, what is the best time to rob, where to go in my house, what my house looks like, that kind of information.*"

Participants also mentioned the possibilities of identity or financial theft ($n = 4$). Three advanced participants expressed

their discomfort about the abilities of companies to manipulate their decisions, judgment or perception of things in some way. Id23 said: "*I think they can show me what I like; I think they can alter the world I am living into the world that is preferential to me, as a consumer.*"

Other risks that participants identified are profiling (n=2), criminals/companies using data to uniquely identify people (n=2), spear phishing (n=1) and social engineering (n=1).

Interestingly, some participants ($n = 6$) shared a general discomfort around the feeling of surveillance, of people knowing too much information about them and being able to use that for nefarious reasons specially around the devices that collect audio and video. For instance Id20 mentioned:

"*Makes me feel uncomfortable that I am in my own home and I can't just say whatever I want without somebody listening you know?*"

Participants with the advanced model identified more examples of threats, and 8 of the 9 were concerned with data breaches. However we found no additional differences between participants based on their mental models. In line with the previous work [33], we found that despite participants identification of these threats, only a few expressed significant concerns or worry about them. However, participants did take some actions to protect the security and privacy of their smart home as we will further discuss below.

4.5 Protective Measures

Participants reported a diverse range of protective measures that they perform or are aware of to reduce their security and privacy risks. Both traditional security best practices and use of protection tools/services were discussed by participants.

4.5.1 Behavioral/non-technical mitigations

Many participants ($n = 12$) mentioned self-censoring their way of using smart home devices. It took various forms, such as turning the device off, changing behavior in front of the device, or avoiding the use of certain device functionality ($n = 6$), as well as limiting the amount of information disclosed to the device ($n = 8$) by not providing more information than absolutely necessary while signing up for an account, or by using someone else's account. For instance, Id22 mentioned changing her behavior in front of the camera:

"*It knew when I woke up and walked to the kitchen... it is in the living room... so it kind of sees that I come around the corner to the kitchen...I kind of try to stay by the wall because I didn't want my robe or pajamas or whatever I was wearing to be on camera.*"

Some participants ($n = 8$) also expressed concerns about their financial information and mentioned frequently monitoring their bank accounts and using credit monitoring services.

4.5.2 Technical mitigations:

Participants discussed using various traditional technical security practices ($n = 9$), such as changing and using strong passwords and using two-factor authentication. Two also reported using certain devices offline to limit access to their data. Two participants with the advanced model also discussed using a separate network for smart home devices. Id8 stated:

"*I have a closed WiFi network for my IoT devices. I do password changes and what not, also my WiFi isn't broadcasted.*"

4.5.3 Tool-based mitigations:

Participants also discussed using some tools or services to protect their privacy around smart home devices ($n = 7$). Two participants hosted local servers and customized the devices to work with that. Others mentioned using different network security devices, installing firewalls or a VPN to protect their network from outside attacks. Id3 stated:

"*I do have a firewall set up on my network that apparently helps with if people try to get the data from me... I can't do anything about the data stored on the cloud. Hopefully the firewall cuts down on any devices that might be compromised or part of a botnet or something like that.*"

A number of participants ($n = 5$) expressed their awareness of such tools or services but were not using those at the time the interview was conducted.

The tool-based mitigations were primarily discussed by the more technically knowledgeable users; nine of the twelve who mentioned tool-based mitigations had the advanced mental model. Furthermore, only the participants with advanced mental models demonstrated familiarity with customizable tools/services for preventing their data from being sent out to the Internet ($n = 5$). On the contrary, most of the participants with the service-oriented model attempt to mitigate their concerns by following traditional security practices (e.g. changing passwords) derived from other computing contexts or changing their behaviors around the devices.

In summary, participants have demonstrated an understanding of some risks from the smart home, but they are not very concerned about many of them. Only a few technical participants did use tools specifically to protect their smart home. Others kept on following the best practices they know from other contexts either because they don't know about what actions to take in the smart home context or the cost of finding and taking those actions is way bigger than their concern. Participants discussed a number of reasons for their lack of concern and unwillingness to take protective measures, as discussed in the next section.

4.6 Reasons for lack of concern and protective actions

While participants could all discuss perceived threats to their security and privacy, most did not express strong concerns.

Several themes emerged when we asked participants why they are not concerned about their security and privacy in the smart home.

Acceptance of trade-off: Most of the participants (n=15) mentioned that they have to give up some of their data and accept the risks for the convenience and services provided by these smart home devices. Four participants also mentioned feeling powerless over this trade-off. For instance, Id12 said:

"*Once I bought all these devices that was it. These functions come with these risks no matter what and I can't do anything about that. There are no third option. If you want the device you have to accept those risks, otherwise don't use it at all.*"

Though participants accepted the trade-off between their privacy and the convenience, 13 of them stated a desire for more transparency from the device manufacturers.

Trust of the manufacturers: Another common reason was participants' trust in the device manufacturers. Eleven participants stated that they trust that companies will not misuse their data because it would damage the company's reputation or will not be financially profitable. Id7 said,

"*I don't think they (companies) are selling it to Russian, I don't think they are trying to steal my identity. I don't think there's anything other than just trying to improve the product, trying to use the information for marketing and advertising.*"

Optimism bias: A number of participants ($n = 9$) expressed a low likelihood of being affected under the assumption that they are not an attractive target for hackers. For instance, Id10 mentioned: "*I also went to college and have student debt. So, I don't feel like an attractive target for someone to try to steal my identity or really do anything.*

Marginal risk: Participants tend to judge the risk from smart home devices by comparing it with how exposed they already are. Several participants ($n = 9$) were not concerned because they believe a wide array of information about them has already been collected or available otherwise and the smart device won't increase the risk. For instance, Id13 said:

"*I've been using the Internet since like I was in middle school... so I don't really have an expectation of privacy.*"

Ten participants believe the data that smart devices are collecting are not that useful or sensitive and would not be harmful to them in the future. Five participants also explicitly mentioned not being concerned because smart devices do not have any critical information about them, i.e., financial details, SSN, etc. Id16 mentioned:

"*I would be worried about just the things like my credit card information or maybe like social security... that hasn't been shared with any other companies... as for like my habit I don't really think that's (concerning) because the companies will only be able to tailor the things we want.*"

Three of these participants also felt that they have already taken enough action to keep their smart home safe.

Trust of regulators: Four participants believe that there are appropriate regulations or overseeing bodies in place which will protect their data from potential misuse by companies. Id19 said: "*If they(company) violate it(rules) it's either going to be corrected or will be most likely to be shut down by a government agency or something.*"

High cost of protective actions: A few participants (n=3) with the advanced mental model also discussed the inconvenience of implementing useful protective measures. For example, Id9 explained the inconvenience of locally hosting the services:

"*You know if I wanted some services that did not connect to the Internet then I kind of have to purchase that myself and run everything that way to prevent, you know, things on my network from going out to the Internet.*"

5 Discussion

We will now report the key insights learned from our study and discuss implications and recommendations for designers, policy makers and researchers.

Knowledge of smart home does not influence threat model or trigger actions: Even though participants had different levels of understanding about how their smart home works, their perception of device manufacturers' data practices was quite similar and not much different from the findings of the earlier work on Internet perceptions [12]. Furthermore, our participants' knowledge about their smart home and manufacturers' data practices did not affect their awareness of possible threats in the smart home. Rather, participants with advanced and simple mental models both frequently mentioned threats and protective actions that are known from the context of the Internet, but also applicable in the smart home. However, participants with the more advanced mental model did show more awareness of the protective measures unique to the smart home, such as preventing data from going outside of the home. Yet, despite awareness of the threats and protective measures, most of the participants choose not to put those into practice. Instead, participants' decisions of protective actions were more influenced by their own biases and concerns related to general Internet usage.

Difference in knowledge (or a lack thereof) between different participant groups: The two groups that emerged in our analysis, i.e., participants with the advanced and service-oriented model, seem to differ primarily in their technical detail and understanding of their smart home. While the participants with advanced model were all installers, there were installers with the service-oriented model as

well. However, we did not find many differences between participants with these two mental models and installers vs. non-installers in terms of their perceptions of data practices. The only difference in knowledge is that the installers of smart cameras and doorbells are more aware of companies' video data storage practices. One reason for installers having this awareness can be the fact that the users need to buy an additional subscription to store the video in the cloud for many of the devices (i.e., nest aware subscription for nest camera, ring protect plan for ring doorbell). This added step exposed the installers to the company's policy regarding video data storage.

Users' lack of exposure to companies' data related policies, in general, may be the reason for the similar perceptions of different groups of participants. This asserts the need for including such information about data practices as a part of the application that is used to control the device and designing nudges and cues for users (installers and non-installers) to get exposed to that information.

Trust paradox: Participants know about much of the data collection occurring with their smart home devices. Many of them are also aware of companies' lack of security in the cloud and data sharing with third-party organizations. Some of them also believe that there is not enough legal protections for consumers. Yet, participants justified their lack of concerns and protective actions with trust that companies will not misuse their data as it will tear down their reputation and regulators will close the company. This paradox can be explained by the notion of learned helplessness seen in many participants, where they ignored possible negative consequences because they feel they have no control. Participants described how once data is collected from their devices, it's beyond their control. And sometimes coped by censoring themselves in some way to keep data from being captured by a device and entered into an application in the first place. Participants thus primarily rely on the organization to keep their data secure and expect governments and policymakers to regulate what is occurring, rather than taking many actions by themselves.

Estimated risk is too low to take action: One of the main reasons for inaction is that participant's estimated risk from the smart home devices is quite low. They are aware of the fact that their daily schedule and habits can be inferred from the data smart home devices are collecting and that companies may use that for targeted advertising. However, companies have been using data such as buying habits for targeted advertising for a long time; it was nothing new to the participants and not viewed as an added risk. Even the risk of a break-in was also not able to raise participants' concerns as they believed they would not be a potential target. A number of participants also didn't think that the use of smart home devices may increase their risk of identity theft as they think there is already enough information out there on the Internet if someone wants to target them specifically. Even the participants who have been a victim of identity theft were quite comfortable with their smart devices as they believe they put enough protection on their financial accounts. None of the participants showed awareness about news of potential smart home device or data misuse, and may not realize the breadth of risk imposed by their devices. Rather, all the participants accepted the trade-off between the benefit of smart home devices with their lower perceived risk as mentioned by Id19, "*I wouldn't let something that I personally see so small affect something that I am enjoying using so much. Something that I personally think more serious, like access to my bank and things like that.. I would lock it down and stop using it immediately.*"

Lack of awareness about data practices and controls impede usage: Despite participants' perceptions and expectations of a large amount of data collection and sharing, we also note that participants are still very uncertain about the device manufacturers' data practices, echoing prior work on users' perception of the Internet and cloud storage more generally [3, 8, 12]. Many participants were also uncertain or unaware of the controls they have on their devices. For a few participants, these uncertainties led to not using certain device functionalities or using the device only at specific times or specific places and may also influence their freedom of expression. In two extreme cases of non-installer participants, Id20 and Id22, it led to the desire of removing the device from their house. However, from their interviews, it appeared the awareness of the available controls may have influenced their privacy behaviors, as mentioned by Id22, "*If I had an easy way to do it... if I had to push a button to remove it(camera recordings) then I would surely remove it.*" In other words, more familiarity with controls may have led those participants to be more comfortable using the device. This underscores the importance of future research to examine ways to nudge users, especially those who are not involved in the set-up and configuration of their smart home, to discover and utilize the available controls.

5.1 Implications and Recommendations

Enhance transparency and control: People want more transparency and control over the data collected and shared by smart home device manufacturers. Participants should have the ability to remove the data and set sharing preferences of their data where possible, for instance, sharing only aggregated data, sharing only usage data, etc. Companies can provide more transparency and controls to users by designing a dedicated web-page or privacy setting in the mobile application where users can view the data points collected by the devices. Another suggestion is to provide privacy and data-related information in addition to the set-up information

in the box, which as Peppet [27] reported, many of the IoT device manufacturers do not. Multiple participants appreciated Google for the transparency and added control in their devices, whereas some were more skeptical about buying devices from lesser-known companies. New smart home start-ups can improve their reputation by providing more transparency and control over users' data.

Researchers have also proposed and developed dedicated devices and tools to give users more security and privacy controls [2, 29, 30]. For instance, Karmann et al. developed 'Alias,' a device that paralyzes the voice assistant by preventing it from listening and only activates the assistant for a custom wake word from the user [2]. Mennicken et al. proposed a calendar-based interface, Casalendar, that visualizes triggered actions and the sensor data collected in a smart home to facilitate users' understanding [24]. We advocate for more such research on novel security and privacy tools and controls beyond the features currently available within a device. While few of our participants were actively looking for additional tools, we believe that easy to use off-the-shelf tools, if commercially available, may increase the comfort of privacy-sensitive people and provide more options for privacy preserving use and adoption.

Best practices for companies and users: As smart home devices become more widespread, smart home attacks will also become more common. Yet, participants who have simpler mental models of their smart home are often aware of and adopted only common traditional best practices (i.e. changing passwords) that may not always help against the security and privacy risks unique to the smart home. Current measures that can help (i.e. locally hosted services) are too technical for the vast majority of potential users. Yet, it is also unclear what best practices are - what are the best methods for average consumers to protect themselves, their data, and their homes? Thus, we concur with Zeng et al. [33] that security researchers, policy makers, and manufacturers need to develop an additional set of best practices for smart home users. However, we want to emphasize that such best practices should be developed by keeping the mental models of users and their technical capabilities in mind. Our findings also revealed that participants rely on companies and policy makers to protect their data. With the widespread use of multiple smart home devices, it will be burdensome for users to manage and take responsibility for all of the data collected and shared by smart home devices. Our study also reinforces the need for the enforcement of a set of privacy best practices for smart home device manufacturers [34]. Policymakers should consider how to administer these rules and penalize companies that do not comply with regulations.

Develop mechanisms to increase user awareness about visual indicators and controls: Researchers need to explore how additional awareness mechanisms can be incorporated directly into smart home devices and applications. For instance, exploring ways to nudge users toward available controls or designing observable cues that provide added awareness of data collection and sharing. For example, Amazon Echo shows blue light patterns when it starts listening. However, designers need to be careful while designing visual indicators, as we found that use of similar indicators (i.e., showing yellow light patterns as a delivery notification by Echo) can be confusing to users. In addition to developing visual indicators, designers should also explore ways to inform users, especially non-installers, of those indicators as a primary part of interaction with the device. For instance, on the first interaction with new users, the voice assistant can speak out loud about the controls they have over their data.

Educate people about future risk: Most of the recent news on IoT misuse is about the use of devices for Distributed Daniel of Service attacks. People do not feel personally targeted when they learn about such generalized attacks. Furthermore, even though participants were aware of the sensitive information that can be inferred from their smart home data, they were unaware of how that data can be used other than for advertising. Centralized online resources are needed where people will be able to learn about the data practices and possible risks from different smart home devices, so that existing users can asses their risk, and potential buyers can decide whether and which device to buy. Mozilla already provides one such online guide [1], however none of our participants mentioned it. Strategies should be taken to educate users about possible risks and available public resources to find information about their devices.

6 Conclusions

In this qualitative interview study of smart home users, we found that participants generally understand that a wide range of information is being collected about their interactions with smart home devices, and shared with a variety of entities to provide useful functionality as well as for marketing and advertising. Much of this information is stored in the cloud, where it is out of the control of users. Yet users are also highly uncertain about these data practices, and desire greater awareness and control over what is occurring. Participants also identified several threats common across computing contexts - such as breaches and financial theft, as well as home safety and security. Yet, despite this awareness of potential threats, they did not view these as serious risks and practiced few mitigation strategies beyond trying to provide devices with no more information than necessary. These findings provide new information about how users perceive what is occurring in the smart home and suggest the need for greater awareness and user friendly control mechanisms as well as cues and visual indicators to inform and contribute to users' security and privacy practices in their homes.

Acknowledgments

We thank our user study participants and pilot participants for their time and input. Tomasz Kosiński was partially supported by the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program (WASP), funded by the Knut and Alice Wallenberg Foundation.

References

[1] Mozilla - *privacy not included. https://foundation.mozilla.org/en/privacynotincluded/. Accessed: 2019-02-13.

[2] Project alias. http://bjoernkarmann.dk/project_alias. Accessed: 2019-02-13.

[3] Mark S. Ackerman, Lorrie Faith Cranor, and Joseph Reagle. Privacy in e-commerce: Examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM Conference on Electronic Commerce*, EC '99, pages 1–8, New York, NY, USA, 1999. ACM.

[4] Noura Aleisa and Karen Renaud. Yes, I know this IoT device might invade my privacy, but I love it anyway! a study of Saudi Arabian perceptions. In *2nd International Conference on Internet of Things: Big Data and Security (IoTBDS 2017)*, pages 198–205, 2017.

[5] Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. Discovering smart home internet of things privacy norms using contextual integrity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(2):59:1–59:23, July 2018.

[6] Eun Kyoung Choe, Sunny Consolvo, Jaeyeon Jung, Beverly Harrison, Shwetak N. Patel, and Julie A. Kientz. Investigating receptiveness to sensing and inference in the home using sensor proxies. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 61–70, New York, NY, USA, 2012. ACM.

[7] Eun Kyoung Choe, Sunny Consolvo, Jaeyeon Jung, Beverly L. Harrison, and Julie A. Kientz. Living in a glass house: a survey of private moments in the home. In *UbiComp*, 2011.

[8] Jason W. Clark, Peter Snyder, Damon McCoy, and Chris Kanich. "i saw images i didn't even know i had": Understanding user perceptions of cloud storage privacy. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1641–1644, New York, NY, USA, 2015. ACM.

[9] Anupam Das, Martin Degeling, Xiaoyou Wang, Junjue Wang, Norman Sadeh, and Mahadev Satyanarayanan. Assisting users in a world full of cameras: A privacy-aware infrastructure for computer vision applications. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1387–1396. IEEE, July 2017.

[10] Marco Ghiglieri, Melanie Volkamer, and Karen Renaud. Exploring consumers' attitudes of smart TV related privacy risks. In Theo Tryfonas, editor, *Proceedings of the 5th International Conference on Human Aspects of Information Security, Privacy, and Trust (HAS)*, Lecture Notes in Computer Science, pages 656–674, Cham, 2017. Springer.

[11] Christine Horne, Brice Darras, Elyse Bean, Anurag Srivastava, and Scott Frickel. Privacy, technology, and norms: The case of smart meters. *Social Science Research*, 51:64 – 76, 2015.

[12] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler. "my data just goes everywhere:" user mental models of the internet and implications for privacy and security. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 39–52, Ottawa, 2015. USENIX Association.

[13] Z. Kaupas and J. Ceponis. End-user license agreement-threat to information security: a real life experiment. In *Proceedings of the IVUS International Conference on Information Technology*, pages 55–60, 2017.

[14] Predrag Klasnja, Sunny Consolvo, Tanzeem Choudhury, Richard Beckwith, and Jeffrey Hightower. Exploring privacy concerns about personal sensing. In *Proceedings of the Seventh International Conference on Pervasive Computing*, 2009.

[15] Predrag Klasnja, Sunny Consolvo, Jaeyeon Jung, Benjamin M. Greenstein, Louis LeGrand, Pauline Powledge, and David Wetherall. "when i am on wi-fi, i am fearless": Privacy concerns & practices in eeryday wi-fi use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1993–2002, New York, NY, USA, 2009. ACM.

[16] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. "alexa, stop recording": Mismatches between smart speaker privacy controls and user needs. https://www.usenix.org/sites/default/files/soups2018posters-lau.pdf. Accessed: 2018-09-10.

[17] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. Alexa, are you listening?: Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):102:1–102:31, November 2018.

[18] Scott Lederer, Jennifer Mankoff, and Anind K. Dey. Who wants to know what when? privacy preference determinants in ubiquitous computing. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, pages 724–725, New York, NY, USA, 2003. ACM.

[19] H. Lee and A. Kobsa. Understanding user privacy in internet of things environments. In *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, pages 407–412, Dec 2016.

[20] Linda Lee, Joong Hwa Lee, Serge Egelman, and David Wagner. Information disclosure concerns in the age of wearable computing. In *Proceedings of the NDSS Workshop on Usable Security (USEC '16)*. Internet Society, 2016.

[21] Nathan Malkin, Julia Bernd, Maritza Johnson, and Serge Egelman. "what can't data be used for?" privacy expectations about smart tvs in the us. In *Proceedings of the 3rd European Workshop on Usable Security (EuroUSEC)*.

[22] Thomas Maronick. Do consumers read terms of service agreements when installing software?a two-study empirical analysis. *International Journal of Business and Social Research*, 4(6), 2014.

[23] Faith McCreary, Alexandra Zafiroglu, and Heather Patterson. The contextual complexity of privacy in smart homes and smart buildings. In *HCI in Business, Government, and Organizations: Information Systems*, pages 67–78, Cham, 2016. Springer International Publishing.

[24] Sarah Mennicken, David Kim, and Elaine May Huang. Integrating the smart home into the digital calendar. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5958–5969, New York, NY, USA, 2016. ACM.

[25] Alessandro Montanari, Afra Mashhadi, Akhil Mathur, and Fahim Kawsar. Understanding the privacy design space for personal connected objects. In *Proceedings of the 30th British Human Computer Interaction Conference (British HCI 2016)*, 07 2016.

[26] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. Privacy expectations and preferences in an iot world. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 399–412, Santa Clara, CA, 2017. USENIX Association.

[27] Scott R. Peppet. Regulating the internet of things: First steps toward managing discrimination, privacy, security, and consent. *Texas Law Review*, 93:85–179, 11 2014.

[28] Fahimeh Raja, Kirstie Hawkey, and Konstantin Beznosov. Revealing hidden context: Improving mental models of personal firewall users. In *Proceedings of the Symposium On Usable Privacy and Security (SOUPS)*, 01 2009.

[29] A. K. Simpson, F. Roesner, and T. Kohno. Securing vulnerable home iot devices with an in-hub security manager. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 551–556, March 2017.

[30] V. Sivaraman, H. H. Gharakheili, A. Vishwanath, R. Boreli, and O. Mehani. Network-level security and privacy control for smart-home iot devices. In *2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 163–167, Oct 2015.

[31] Rick Wash. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 11:1–11:16, New York, NY, USA, 2010. ACM.

[32] Peter Worthy, Ben Matthews, and Stephen Viller. Trust me: Doubts and concerns living with the Internet of Things. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*, pages 427–434, New York, 2016. ACM.

[33] Eric Zeng, Shrirang Mare, and Franziska Roesner. End user security and privacy concerns with smart homes. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 65–80, Santa Clara, CA, 2017. USENIX Association.

[34] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. User perceptions of smart home iot privacy. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW):200:1–200:20, November 2018.

[35] Verena Zimmermann, Merve Bennighof, Miriam Edel, Oliver Hofmann, Judith Jung, and Melina von Wick. 'home, smart home' - exploring end users' mental models of smart homes. In Raimund Dachselt and Gerhard Weber, editors, *Mensch and Computer 2018 - Workshopband*, Bonn, 2018. Gesellschaft für Informatik e.V.

A Appendix

A.1 Recruitment Survey

- How many smart home devices do you own?
- Please select all the smart home devices you own? (Choices: A list of devices with option to include devices that are not listed)

- Do you have a degree on any Computer Science related major? (Choices: Yes, No)
- Who installed and automated the smart home devices in your house? (Choices: I installed all the devices, I installed some of the devices, Someone else installed all the devices)
- Your name .....
- Your email .....
- Your Age (Choices: Less than 20 yrs, 21-30 yrs, 31-40 yrs, 41-50 yrs, 51-60 yrs, More than 60 yrs)

A.2 Interview Questions

General questions:

- What smart devices do you have in your house?
- How did you use these devices?
- How do you control these devices?

Drawing exercise:

- Can you draw how these devices collect information and how that information flows between the devices and any other involved entities?

Data collection: (for each device)

- What information is collected by the device?
- Do you think that data should be collected?
- Do you think it needs to be collected? If so, for what purpose?

Data storage

- Where do you think the device transmits this information?
- Where do you think the data are stored? What data are stored? For how long?
- Is it possible to check what data are stored? If yes, how?
- Do you have any control over the stored data?
- Is it possible to remove this data? Have you ever considered removing data?
- Can you remove your device usage log?

Data sharing:

- Who can access and use the data that have been stored?
- How do the device manufacturer/others use the data?
- Does your device manufacturer share these data with any other companies and organizations? If yes with whom?
- Why do you think they share the data? What are the benefits? To them and to you?
- Do you think you opt-in to this sharing? When and how do you opt into sharing?
- Do you think you can opt out? Do you consider opting-out?
- Are the devices sharing data between themselves? What data, how and for what purposes?

Data inference:

- Does that concern you about the way the device manufacturers use your data? What are some of the concerns?
- How can a third party use your data? Does that concern you? What are some of the concerns you have regarding this?
- What can be inferred about you from this data by the entities or organizations that have the data?
- What do you think some of the threats are to your data or yourself?

Mitigation techniques:

- Have you done anything to resolve these threats and to protect your data?
- What you think you should be doing?
- What controls do you have on your data? How hard it is to use these controls?
- What controls do you want to have or would like to be able to do regarding your data privacy?
- What do you think companies are doing to protect your data privacy? What do you expect them to do?

Closing question:

- Is there anything else or any concern you want to share with me about your smart home or expected me to ask?

Demographics:

- What is your ethnicity?
- what is your primary occupation?
- What is the highest level of education you have completed?

- What was your major?
- Did you have any degree on a computer science related topic?

Self-reported technical skill [33]:
On a scale of 1(very weak) - 5(very strong)

- How would you rate your knowledge of technology in general?
- How would you rate your knowledge of computer security and privacy?
- How would you rate your knowledge of smart home technology?

A.3 Summary of Participants' Demographics

ID	Gender	Age	Education	Profession	Installed the devices?
ID1	M	21-30	MS: Computer Engineering	Grad student	Yes
ID2	M	21-30	BS: Computer Science	Programming consultant	Yes
ID3	M	21-30	Juries Doctorate	Attorney	Yes
ID4	M	31-40	Doctorate: Medicine	Product manager	Yes
ID5	F	21-30	BS: Biology	Banking	No
ID6	M	61-70	BA: Urban Planning	Retired computing professional	Yes
ID7	M	51-60	Associate Degree: Arts and Science	Computing professional	Yes
ID8	M	41-50	Diploma: Media Arts	Network engineer	Yes
ID9	M	31-40	BS: computer science	IT sales	Yes
ID10	F	31-40	MS: Kinestheology	Unemployed	No
ID11	F	21-30	MS: Kinestheology	Clinical researcher	Yes
ID12	M	31-40	Post Graduate: Chemistry and Physics	Business entrepreneur	Yes
ID13	F	31-40	MS: educational counseling	Education administration	Yes
ID14	M	51-60	BA: Criminal Justice	Banking	No
ID15	F	31-40	BA: Russian	Human Resource	No
ID16	F	21-30	Bachelors: Biology and Psychology	Insurance verification specialist	No
ID17	M	31-40	Masters: Sociology and Applied Research	Higher education administrator	Yes
ID18	F	21-30	Bachelors: Elementary Education	Fifth grade teacher	No
ID19	M	31-40	High School	Customer Service	Yes
ID20	F	61-70	Bachelors: Accounting	Accountant	No
ID21	F	61-70	College	Retired	No
ID22	F	51-60	BA: Practical Civilization	Administrator: call center	No
ID23	M	21-30	BS: Biomedical Sciences	Graduate student	Yes