To appear in the Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2022

"I mute my echo when I talk politics": Connecting Smart Home Device Users' Concerns to Privacy Harms Taxonomy

Chola Chhetri, Vivian Motti
College of Engineering and Computing
George Mason University
{cchhetri, vmotti} @gmu.edu

With the proliferation of Internet of Things devices, smart home devices are expected to increase in use. However, experts have raised privacy concerns regarding these devices. As the body of literature on understanding privacy concerns continues to emerge, we realize the need for a privacy concerns taxonomy to standardize and facilitate common understanding of privacy concerns. To address this gap, we conducted 25 interviews of smart home device users and analyzed their privacy concerns qualitatively. This paper contributes analysis of user privacy concerns from the angle of privacy taxonomy theory. It examines whether privacy concerns could be characterized by Solove's taxonomy of privacy, which is a well-recognized privacy taxonomy for informational privacy. We further discuss results and their implications.

INTRODUCTION

With the proliferation of Internet of Things (IoT) devices, smart home devices (SHD) are expected to increase in use and have the potential to transform the home through Internet-based devices and automation systems (S. Li et al., 2015). However, past research shows that privacy presents as one of the challenges to realize the full potential of SHDs (Furszyfer Del Rio et al., 2020).

SHDs have been known to contain privacy vulnerabilities (Chhetri & Motti, 2020) and raise privacy concerns in users and non-users (Haney et al., 2020; Zeng et al., 2017). Researchers and SHD designers cannot develop effective privacy-enhanced solutions without enabling the users (Chong et al., 2019), which requires an understanding of users' privacy concerns.

The study of privacy concerns of users has recently gained attention (Chhetri & Motti, 2019; Zheng et al., 2018). Similarly, studies of other stakeholders, such as non-users (Chhetri & Motti, 2022c) and bystanders (Yao et al., 2019), are also gaining attention. Recent studies are also exploring designs of privacy controls (Chhetri & Motti, 2022a). In this study, we take a theoretical approach to elicit privacy concerns from SHD users and examine whether an existing taxonomy of privacy harms, from the law field, can be used to characterize and understand SHD users' concerns.

In this paper, we aim to address two research questions:
RQ1. What are the privacy concerns of SHD users?
RQ2. How are these SHD privacy concerns connected to the various facets of privacy harms taxonomy?

To answer these questions, we interviewed 25 SHD users and analyzed the transcripts. We performed qualitative analysis of their privacy concerns. We further examined whether participants' privacy concerns could be characterized by Solove's taxonomy of privacy harms, which is a well-recognized privacy taxonomy for informational privacy (Solove, 2005).

The contributions of this paper are twofold: (1) identifying the privacy concerns that users have about SHDs, and (2) generating insight on how SHD users privacy concerns are connected to various facets of *privacy harms* or *violations*. These insights seek to facilitate in developing privacy controls for SHDs.

BACKGROUND

Prior work shows various definitions of privacy. One highly cited work defines privacy as *control over one's personal information* (Westin, 1968). This definition applies to the context of this paper as most privacy concerns regarding SHDs are related to information and its control.

Solove's taxonomy of privacy helps understand privacy problems, in the law field, by identifying four groups of privacy harms: information collection, information processing, information dissemination, and privacy invasion (Solove, 2005). *Information collection* includes surveillance and interrogation. *Information processing* includes aggregation, identification, insecurity, secondary use and exclusion. *Information dissemination* includes breach of confidentiality, disclosure, increased accessibility, blackmail, appropriation, and distortion. Finally, *privacy invasion* includes intrusion and decisional interference (Solove, 2005).

Solove's taxonomy is used in multiple fields of research. We found literature that applied Solove's taxonomy to characterize older adults' threat models about privacy and security (Frik et al., 2019). Our study examines how SHD users' privacy concerns are connected to these various facets of privacy harms.

METHOD

We conducted semi-structured interviews to address the research questions. The semi-structured interview design provided some degree of standardization and consistency in the study protocol. In addition, it provided us the flexibility to probe deeper into participants' responses and request explanations when necessary (Oates, 2005).

The study protocol and all related materials received ethics approval from our host institution's Institutional Review

Board (IRB). Due to the ongoing pandemic, all interviews were conducted online using the videoconferencing application Zoom and an online collaborative whiteboard application Miro. The interview protocol was tested through three pilot interviews which helped us refine the protocol but the responses of the pilot interviews were not included in the data analyses.

Participants

We recruited 25 participants via Twitter and our institution's event listserv. There were 13 respondents who identified as women and 11 men. Fifty-two percent of the participants were aged 18-25 and 40% were aged 26-35. Forty-eight percent of the participants reported having a bachelor's degree and 40% reported having a master's degree. Thirty-six percent of the participants reported to be Asian, 28% reported to be White, 12% reported to be African-American, and another 12% reported to be Hispanic. We summarize the participant demographics in Table 1.

All participants owned at least one smart home device. The maximum number of devices owned by a participant was 8 and the median was 2 devices. The majority of participants reported using smart speakers (72%), such as Amazon Echo and Google Home.

Table 1. Summary of participant demographics.

Characteristics	Frequency	
	n=25	%
Gender		
Female	13	52
Male	11	44
Other	1	4
Age		
18-25	13	52
26-35	10	40
36-45	2	8
Highest educational level		
Master's	10	40
Bachelor's	12	48
High School	2	8
Other	1	4
Ethnicity		
Asian	9	36
White	7	28
African-American	3	12
Hispanic	3	12
Other	3	12

Procedure

The study advertisement included an online form with details about the study. Interested participants could provide an email address to be contacted for the interview and answer a question about whether they owned or used SHDs.

The participants who expressed interest in the study, reported using SHDs and consented to participate in the study were contacted with available meeting times. Participants who further confirmed a meeting time were provided with links to an online meeting and an online collaborative application.

During the interview, participants reviewed the consent form and completed the demographic information form. Then, they answered questions about their SHD usage experience and privacy concerns about SHDs. Lastly, they explained their specific concerns about privacy per room (kitchen, living room, bedroom, bathroom, and overall) and the device(s), sensor(s), information collected and shared as well as services that could use the data collected. The interview was concluded by discussing design recommendations for users of SHDs to gain control over their privacy. In this paper, we present only the analyses related to privacy concerns relevant to the research questions discussed in the Introduction section. We present details of privacy controls needs of SHD users in another paper (Chhetri & Motti, 2022b).

The interviews lasted an average of 48 mins (minimum 30 minutes, maximum 60 minutes) and were audio recorded (with consent) for transcription purposes. Each participant was compensated with a USD 20 gift card.

Data Analysis

The first author transcribed the interviews. Another researcher verified the transcripts. Both researchers then analyzed the interview transcripts (Braun & Clarke, 2006).

Following a widely used method of qualitative analysis (Braun & Clarke, 2006), we immersed ourselves with the interview transcripts before coding the privacy concerns inductively. Then, we examined the relationships or patterns among codes and followed a deductive approach to cluster similar codes into higher-level themes from Solove's taxonomy (Solove, 2005). We used affinity diagramming (Harboe & Huang, 2015) to cluster similar codes. Privacy concerns codes and categories obtained from the analysis are detailed in Table 2.

Two coders began the analysis by familiarizing themselves with the dataset (interview transcripts) and coded five transcripts independently. They then met to discuss the generated codes and to identify and reconcile differences. All disagreements were resolved by consensus. This process was repeated in batches of three transcripts until the last transcript was coded.

We reached saturation of codes after the 22nd transcript but we coded and analyzed all 25 transcripts. Between the two coders, we achieved an inter-rater reliability measure of 0.89 using Cohen's kappa, which is considered strong (McHugh, 2012).

RESULTS

In this section, we describe the results of our qualitative analysis. We will describe the privacy concerns of our participants regarding SHDs.

The privacy concerns varied among participants. While some users reported concerns about how their information will be used in the future, others noted the lack of transparency. Participants were concerned about information they considered private: information about finances, health or religious beliefs, medication intake, secret recipes, TV shows, and political affiliation or belief.

Privacy concerns raised by our participants can be understood as the threat models or risks from smart home devices. With inspiration from prior literature (Frik et al., 2019), we grouped the privacy concerns codes from our analysis into categories based on Solove's taxonomy of privacy harms (Solove, 2005), which includes information collection, information processing, information dissemination, and privacy invasion, as described in the Background section.

Our analysis generated codes that were connected to all aspects of Solove's taxonomy. We also found two codes that did not match any aspect of the Solove's taxonomy. These two codes were: Lack of consumer knowledge and Lack of policy. The authors, therefore, through deliberation and discussion, organized the two codes into a new category: Policy and Awareness. We describe the privacy concerns in this section and include a comprehensive breakdown of the codes and their mapping to Solove's taxonomy in Table 2.

Privacy Concerns Connected to Information Collection

One major concern was gathering of audio and video recordings by SHDs, which was raised by 19 participants. Some SHDs, especially those with microphones and cameras, are always listening. This made some participants uncomfortable in having conversations, especially those containing sensitive information, personal data, and private or intimate conversations. For example, according to P02:

"Well, the speaker's basically acting upon a word that I call it, so that means it's listening to me and my environment all the time. So, in order to activate, it's continuously collecting audio data from me. So, that was kind of concerning." (P02).

Participants also raised concerns about collecting personal data without consent. For example, P05 mentioned: "[My] privacy concern is the device is listening, collecting that personal data without my consent when I'm unaware of it."

Participants expressed the perception that SHDs make data collection easier for companies that are already collecting data from other sources, such as online browsing. Participant sentiment included that companies can "hear and see everything we do" (P10), and the desire for limiting the collection of personal data. Participants also feared that this large-scale data collection and sharing could lead to future data abuse. For example, P20 feared a social credit system based on data collected:

"I don't want all of our information being out there constantly being recorded, or watched, [...] our data also being shared. Like, I think if there were to be some kind of, like, a healthcare system watching us or something. I don't want some kind of social credit system where people can use our data against us. So, I just don't want everything to be stored by individual companies." (P20).

Participants with concerns in this category also worried that SHDs could collect location information and use it for tracking purposes. SHDs with location awareness would raise safety issues, especially if the information fell in 'wrong hands'. Two participants were also concerned that corporations and governments could spy on them through SHDs. Four participants mentioned that their concerns were triggered after they began to see advertisements related to their conversations, even though they had not directed those conversations to their smart assistants. One participant also noticed that their device was aware of their sleep habits, since they started receiving advertisements for sleeping pills.

Table 2. Privacy concerns codes generated from the qualitative analysis and their mapping to Solove's taxonomy. Policy and Awareness category was not part of Solove's taxonomy. It was generated from our analysis.

Taxonomy	Codes	Example Quotes
Information Collection	Always listening Hear/see everything Easy to collect Video feed Data storage Location awareness Location tracking Government spying	"We're always really concerned that our [*device*] is listening to us. And we've definitely had some instances where we talk about something and we haven't searched on it on our phone or anything. And then the next day, it shows up as an ad or a recommended news article." (P6)
Information Processing	Hacking Steal information Unauthorized access Easy access Misuse Leakage Voice fingerprint Profiling Protection of credentials	"I don't want my data to be shared with anybody else. Of course, I know that my data will help to design a better implementation what I were looking for, so, that is okay. But, *without my consent?* That would be my biggest concern." (P20)
Information Dissemination	Selling information	"… So, my concern is the fact that they're selling my information and using it to advertise back to me." (P22)
Privacy Invasion	Intrusion Unsolicited ads Targeted ads	"I am concerned about talking politics with a smart device, so I mute my echo when I talk politics" (P01)
Policy and Awareness	Lack of consumer knowledge Lack of Policy	"People are unaware what can be done with their information." (P16) "… policy is behind …" (P05)

Privacy Concerns Connected to Information Processing

All 25 participants raised at least one concern which was coded under the category of Information Processing. The major concerns raised by participants in this category were intentional or accidental misuse of data, unauthorized access to data, leakage, stealing or hacking of data, unsolicited advertisements, and profiling.

Participants were also concerned that unauthorized individuals could hack into SHDs and remotely control the SHD device, part of the house, or the entire house. For instance, one participant mentioned:

"If someone hacks into the smart home device, they can control the entire house." (P25).

Participants also worried about the potential for companies and governments to misuse the data they have about users. Another concern of participants included if data were to get into wrong hands, such as criminals, they could cause harm, affecting the safety of the smart home resident. For example, P19 worried about leakage of audio recordings: "I would be worried of, if like, the recordings that they have of us gone out somehow, or if someone found of a way to listen in."

Participants wondered if Internet-enabled SHDs provided the same level of protection on a computer, and there were concerns related to accidental data leaks to third parties and users being profiled with machine learning algorithms. Newer methods of marketing and surveillance also concerned participants. For instance, one participant noted:

"Just by analyzing your voice, people can access data about you that you did not give permission to release and it can be used as an added way of surveillance. It can be used like an additional way of marketing. [...] Marketers are really excited about algorithms that use your voice to tell your age, race, contraceptives, and the like because they could better target specific groups [...] which is great on the marketing end. I think that can cause some privacy issues there." (P14)

Privacy Concerns Connected to Information Dissemination

Another concern was the sales of personal information collected by companies manufacturing SHDs. Four participants raised this concern. As P11 stated, it is "pretty likely that most big corporations are selling your data to other [...] and to advertisers. That's why you get so many targeted ads based on what you ask even." Another participant who did not worry about data collection raised a concern about companies selling personal information:

"I don't care that they have my address or my birthday or anything like that because if someone wants it, they're going to find it. It's more so what they do with the information. So, my concern is the fact that they're selling my information and then using it to advertise back to me." (P17)

Privacy Concerns Connected to Privacy Invasion

Participant concerns in this category were related to the possibility of intrusion into private space, someone being able to reveal secrets, or being able to interfere with decisions. We coded six participants' concerns into this category. For example, participants were concerned about revealing their political affiliation or belief and worried that information could be used against them. As P01 mentioned, "I am concerned about talking politics with a smart device, so I mute my echo when I talk politics".

Additional Concerns: Policy and Awareness

Two participants raised concerns regarding the lack of policies or regulatory framework to control SHD privacy and the lack of awareness among SHD users on data collection and usage practices of companies. For example, a participant (P24) who raised concerns about consumer knowledge of SHD risks thought that most people do not know that current technologies can allow voice fingerprinting: "People are unaware that personal details can be constructed from their voice data."

DISCUSSION

Our findings show that privacy concerns related to SHDs are connected to the various facets of privacy harms based on Solove's taxonomy. We found that privacy concerns connected directly to the information collection, processing, dissemination, and privacy invasion aspects of Solove's taxonomy. Prior literature also shows concern from experts and users regarding invasion of privacy (Abdi et al., 2019; Benlian et al., 2020; Zheng et al., 2018). However, this paper was able to contribute to the literature by systematizing the knowledge to a formal taxonomy through empirical findings.

Additionally, our findings show that the lack of consumer awareness and inadequate government regulation on data protection generates privacy concerns among SHD users. Although these aspects of users' concerns from our findings were not directly part of Solove's taxonomy, the taxonomy was created to systematize knowledge on privacy harms so that it helps all stakeholders including lawmakers (Solove, 2008). Thus, we demonstrate how the taxonomy can be used to characterize SHD users privacy concerns. Our results also emphasize the need for awareness, training and support for users regarding privacy controls in SHDs.

The implications are that SHD manufacturers need to provide privacy controls to address users' concerns. Users have also emphasized the need for regulations and industry certifications to ensure users' privacy is protected. Privacy controls based on context and users' preferences could serve users better than a one-size-fits-all approach. In addition, privacy controls that meet users' needs should also be usable.

Fine-grained privacy controls and settings can provide a feeling of control to users; however, it is likely that they may not be utilized as it can get overwhelming (Zheng et al., 2018). Optimum privacy solutions should be achieved through multiple stakeholders: users, developers, government and industry (Haney et al., 2020). Our empirical findings also echo these privacy-related recommendations made by security and privacy experts.

Privacy related certification programs that verify the privacy features of SHDs can help build users confidence. Frameworks such as privacy-by-design (Perera et al., 2016) can help software developers guide towards implementing data privacy principles, such as minimization, anonymization, encryption, and control.

LIMITATIONS AND FUTURE WORK

The results of this paper must be considered within the context of some limitations. The study participants were based in the United States, young and well-educated. Prior research shows that privacy perceptions can vary based on culture (Y. Li et al., 2017). Thus, the findings may not be generalizable to

other cultures. However, Solove's taxonomy is widely used in multiple fields. So, we believe our findings still provide good theoretical grounding. Future studies should consider eliciting privacy concerns from other cultures in light with the taxonomy. Additionally, we think future research work that benefits users include privacy controls needs assessment and guidelines for privacy controls development in SHDs.

CONCLUSION

In this paper, we identified the privacy concerns of smart home device users through 25 qualitative interviews. We further presented a privacy concerns taxonomy to understand those privacy concerns in a comprehensive and consistent way. Our findings show how privacy concerns related to SHDs are connected to the various facets of privacy harms based on Solove's taxonomy.

This paper contributes to the literature by systematizing the knowledge about SHD users' privacy concerns to a formal taxonomy through empirical findings. This work should lay a foundation towards research and development efforts that are focused on privacy controls in the SHDs, particularly in addressing the users' concerns.

ACKNOWLEDGEMENT

We thank our participants and anonymous reviewers. We are grateful to Yoseif Berhe for his help with coding. This material is based upon work supported by 4-VA, a collaborative partnership for advancing the Commonwealth of Virginia, and Commonwealth Cyber Initiative (CCI). Any opinion, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of our research sponsors.

REFERENCES

Abdi, N., Ramokapane, K. M., & Such, J. M. (2019). More than smart speakers: security and privacy perceptions of smart home personal assistants. *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, 451–466.

Benlian, A., Klumpe, J., & Hinz, O. (2020). Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. *Information Systems Journal*, *30*(6), 1010–1042.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101. https://doi.org/10.1191/1478088706qp063oa

Chhetri, C., & Motti, V. (2022a). Designing and Evaluating a Prototype for Data-related Privacy Controls in a Smart Home. In N. Clark & S. Furmann (Eds.), *International Conference on Human Aspects of Information Security, Privacy, and Trust* (p. 10). Springer, Cham.

Chhetri, C., & Motti, V. (2020). Identifying Vulnerabilities in Security and Privacy of Smart Home Devices. In K.-K. R. Choo, T. Morris, G. L. Peterson, & E. Imsand (Eds.), *National Cyber Summit {(NCS)} Research Track 2020, Huntsville, AL, USA, June 2-4, 2020* (Vol. 1271, pp. 211–231). Springer. https://doi.org/10.1007/978-3-030-58703-1_13

Chhetri, C., & Motti, V. G. (2019). Eliciting Privacy Concerns for Smart Home Devices from a User Centered Perspective. In N. G. Taylor, C. Christian-Lamb, M. H. Martin, & B. Nardi (Eds.), *Information in Contemporary Society, Proceedings of iConference, LNCS 11420* (pp. 1–11). Springer Nature. https://doi.org/10.1007/978-3-030-15742-5_8

Chhetri, C., & Motti, V. G. (2022b). User-Centric Privacy Controls for Smart Homes. *Proceedings of ACM on Human-Computer Interaction*, *6*(CSCW2), 34.

Chhetri, C., & Motti, V. G. (2022c). Privacy Concerns about Smart Home Devices : A Comparative Analysis between Non-Users and Users. *International Conference on Applied Human Factors and Ergonomics*, 10.

Chong, I., Xiong, A., & Proctor, R. W. (2019). Human factors in the privacy and security of the internet of things. *Ergonomics in Design*, *27*(3), 5–10.

Frik, A., Nurgalieva, L., Bernd, J., Lee, J., Schaub, F., & Egelman, S. (2019, August). Privacy and Security Threat Models and Mitigation Strategies of Older Adults. *Fifteenth Symposium on Usable Privacy and Security ({SOUPS} 2019)*. https://www.usenix.org/conference/soups2019/presentation/frik

Furszyfer Del Rio, D. D., Sovacool, B. K., Bergman, N., & Makuch, K. E. (2020). Critically reviewing smart home technology applications and business models in Europe. *Energy Policy*, *144*, 111631. https://doi.org/https://doi.org/10.1016/j.enpol.2020.111631

Haney, J. M., Furman, S. M., & Acar, Y. (2020). Smart home security and privacy mitigations: Consumer perceptions, practices, and challenges. In A. Moallem (Ed.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12210 LNCS* (pp. 393–411). Springer International Publishing. https://doi.org/10.1007/978-3-030-50309-3_26

Harboe, G., & Huang, E. M. (2015). Real-world affinity diagramming practices: Bridging the paper-digital gap. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 95–104.

Li, S., Da Xu, L., & Zhao, S. (2015). The internet of things: a survey. *Information Systems Frontiers*, *17*(2), 243–259.

Li, Y., Kobsa, A., Knijnenburg, B. P., Nguyen, M.-H. C., & others. (2017). Cross-Cultural Privacy Prediction. *Proc. Priv. Enhancing Technol.*, *2017*(2), 113–132.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Oates, B. J. (2005). *Researching information systems and computing*. Sage.

Perera, C., McCormick, C., Bandara, A. K., Price, B. A., & Nuseibeh, B. (2016). Privacy-by-design framework for assessing internet of things applications and platforms. *Proceedings of the 6th International Conference on the Internet of Things*, 83–92.

Solove, D. J. (2005). A taxonomy of privacy. *U. Pa. L. Rev.*, *154*, 477.

Solove, D. J. (2008). Understanding privacy. *UNDERSTANDING PRIVACY*.

Westin, A. F. (1968). Privacy and freedom. *Washington and Lee Law Review*, *25*(1), 166.

Yao, Y., Basdeo, J. R., Mcdonough, O. R., & Wang, Y. (2019). Privacy Perceptions and Designs of Bystanders in Smart Homes. *Proc. ACM Hum.-Comput. Interact.*, *3*(CSCW). https://doi.org/10.1145/3359161

Zeng, E., Mare, S., & Roesner, F. (2017). End User Security & Privacy Concerns with Smart Homes. *Symposium on Usable Privacy and Security (SOUPS)*.

Zheng, S., Apthorpe, N., Chetty, M., & Feamster, N. (2018). User Perceptions of Smart Home IoT Privacy. *Proceedings of ACM Human-Computer Interaction*, *2*(CSCW), 200. https://doi.org/10.1145/3274469