
# Do Privacy and Security Matter to Everyone? Quantifying and Clustering User-Centric Considerations About Smart Home Device Adoption

Natã M. Barbosa, Zhuohao Zhang, and Yang Wang,
*University of Illinois at Urbana-Champaign*

https://www.usenix.org/conference/soups2020/presentation/barbosa

This paper is included in the Proceedings of the Sixteenth Symposium on Usable Privacy and Security.

August 10–11, 2020

978-1-939133-16-8

Open access to the Proceedings of the Sixteenth Symposium on Usable Privacy and Security is sponsored by USENIX.

# Do Privacy and Security Matter to Everyone? Quantifying and Clustering User-Centric Considerations About Smart Home Device Adoption

Natã M. Barbosa, Zhuohao Zhang, and Yang Wang

University of Illinois at Urbana-Champaign
*{natamb2,zhuohao4,yvw}@illinois.edu*

## Abstract

While consumer adoption of smart home devices continues to grow, privacy concerns reportedly remain a roadblock to mass adoption. However, it is unclear exactly how the interplay between privacy and other factors affect consumers' purchase decisions, and whether such considerations are held only by certain consumer groups but not others. In order to unpack the decision-making process of smart home device adoption, we conducted a mixed-method analysis using online survey data collected from 631 US participants. Our analysis uncovered motivators and blockers of purchase decisions, along with their relative importance. We found that consumers can be segmented based on their considerations into three clusters: affordability-oriented, privacy-oriented, and reliability-oriented. We present an in-depth quantification of consumer considerations on smart home device adoption along with desired privacy and security features consumers wish to use to protect their privacy in the smart home.

## 1 Introduction

Consumer adoption of smart home devices continues to see steady growth. A recent study in the US by Statista [34] reports 41 million homes with at least one smart home device in 2020. This figure represents a 32.4% household penetration in the US, an increase of 18.7% from the previous year. Such growth continues despite consumer privacy and security concerns reportedly remaining a roadblock to mass adoption [20]. Taken together, these reports pose an interesting conundrum: why do we see increased adoption despite widespread privacy and security concerns? Unpacking this question means understanding how much considerations of privacy and security weigh into purchase decisions alongside other factors, and whether such considerations may be held only by certain consumer groups but not others.

From an economic perspective, consumers consider privacy alongside other factors, and make a decision based on their calculus of whether expected benefits would outweigh expected costs (e.g., associated privacy risks) [1]. Previous work found that privacy is an important factor in the purchase decisions of Internet of Things (IoT) devices for many users, standing only behind features and price [14]. However, it is unclear to what extent considerations of privacy and security stand against the many factors that may motivate or prevent consumers from adopting such devices. For example, would the expected convenience outweigh the privacy concerns? Is the presence of privacy and security features more important than the absence of them? In addition, while many consumers claim to have privacy concerns, the growing number of devices installed each year is evidence that either privacy and security concerns do not stop many from adopting, or such concerns are outweighed by other factors when it comes to actual adoption. Either way, adoption statistics suggest that privacy and security considerations may play a different role to different people, and people could be segmented based on their purchase considerations, since privacy or security may not be a pre-purchase consideration for some people [14].

To address this conundrum, we report on data from a US-based survey with 631 participants where half of all participants reported having a smart home device and the other half did not. Participants were asked what could motivate or prevent them from adopting smart home devices, separately. Our mixed-methods analyses quantify and cluster motivators and blockers in order to provide an in-depth understanding of consumers' decision-making process in smart home device purchases. Participants were also asked about what privacy and security protections they desire for smart home devices.

Our findings show that good privacy or security practice was considered as a motivator only by 11% of the partici-


*USENIX Symposium on Usable Privacy and Security (SOUPS)* 2020.
August 9–11, 2020, Virtual Conference.

pants, while privacy or security concerns/risks being a blocker for half of the participants. Across all participants, the top considerations are ranked as follows, with "+" for motivators and "-" for blockers: +*convenience*, -*privacy*, -*price*, -*security*, +*cost-saving*, -*risk*, -*reliability*, and +*control*. Our clustering analysis reveals three groups of consumers: (1) affordability-oriented; (2) privacy-oriented; and (3) reliability-oriented. We also built a decision tree to predict which consumer cluster a person belongs to based on the person's purchase considerations. The decision tree can help explain what matters in people's purchase decision-making process. We discuss implications of our results on the use, privacy control and regulation of smart home devices.

The contributions of this research include (1) quantified relative importance of different factors as motivators and blockers in people's smart home device adoption considerations; (2) consumer segmentation based on these factors; (3) a list of desired privacy protections for smart homes; and (4) actionable recommendations based on our findings. In summary, our work helps unpack the smart home adoption conundrum and provide guidance on how companies and policymakers could enable consumers to make more informed decisions about smart home device purchases.

## 2 Related Work

### 2.1 Users' IoT Privacy Considerations

A great number of qualitative studies have looked into privacy and security concerns and expectations of IoT and smart home users (e.g., [6–8, 10, 21, 23–25, 38–42]). These studies have revealed privacy and security concerns among both owners and non-owners of smart home devices. More importantly, a recurring theme are the trade-offs between price, functionality, and privacy [8, 10, 40], where some users have reported prioritizing functionality over privacy and *vice-versa*.

A recent interview study by Emami-Naeini *et al.* [14] on user privacy and security considerations on IoT device purchase revealed most device owners did not consider privacy or security prior to purchasing, but did become concerned once the devices were installed in their homes. The attributed reason being lack of access to or information about privacy and security of the devices. Through their interviews, the authors also identified 16 factors that influenced users to purchase smart home devices, and later asked 200 survey respondents to rate the influence of the 16 factors on a 5-point Likert-scale. Their survey results revealed privacy as the third-most influencing factor on participant decisions, standing only after features and price. One of the findings from this study is that privacy and security may not be a consideration for many consumers, and that some would only consider it after being exposed to such considerations. This finding motivated the authors to design and evaluate privacy labels for IoT devices [13, 14] in order to inform and educate consumers about privacy and security aspects before purchasing IoT devices.

Prior qualitative interview studies have also explored user-centric privacy behaviors, tools and protections (e.g., [14, 39]), revealing strategies such as frequent updates, strong access controls, device re-positioning, using separate routers, data localization, disconnection from the Internet, and the use of a private mode.

### 2.2 Privacy Value in Emerging Technologies

From an economic perspective, if personal and societal privacy are considerations held by consumers, they are pondered as part of a calculus that will inform and mediate decision-making with regard to the adoption of emerging technologies involving privacy risks [1, 4, 32]. Accordingly, numerous studies have looked at consumer considerations of privacy in technological economic transactions (e.g., [2, 9, 11, 19, 29–31]). Notably, Egelman *et al.* [11]'s experiment on smartphone app installs showed that given the appropriate choice architecture highlighting app permissions, users are willing to pay a premium for privacy, potentially leading to more rational decisions. Danezis *et al.*'s study [9] revealed that smartphone users may be willing to allow their location to be monitored for a given price, it being higher when users traveled frequently or communicated with partners using their phone.

Behavioral studies such as these suggest consumers engage in risk assessments heavily controlled by the underlying context, which can be influenced by biases, heuristics, and framing effects [1]. These studies also often point to a potential paradox where people's stated privacy attitudes and preferences deviate from their observed behavior, commonly referred to as the "privacy paradox." [26]. However, contemporary views on privacy decision-making provide possible explanations of and even refute the existence of such a paradox. For example, Adjerid *et al.* [3] argues that the constancy of normative factors (e.g., privacy preferences, settings, regulation) and behavioral effects (e.g., choice framing, defaults) must be challenged in hypothetical and actual choice settings, since "consumers may both overestimate their response to normative factors and underestimate their response to behavioral factors." Solove [32] refutes the existence of a privacy paradox altogether, arguing that behavior is better understood as "choices about risk in specific contexts" and attitudes as "people's broader valuation of privacy, often across many contexts." Solove argues that privacy decision-making should be instead approached as behavior that involves risk in which people's decisions are influenced by many factors. Such an approach is also more suitable to the reality of today, where new products are increasingly made to be Internet-connected and escaping the associated privacy and security risks becomes increasingly difficult. According to Solove, this approach also stands in contrast with the assumptions made in the privacy paradox, which often stems from "leaps in logic" that generalize from specific contexts to broad attitudes across contexts [32].

Arguably, reports indicating privacy is a blocker to mass-adoption of smart home devices (e.g. [20]), yet devices gaining significant adoption in the past few years [34] hint at a potential privacy paradox taking place in the smart home. However, user studies paint a more nuanced picture about smart home device adoption considerations. For instance, some consumers indeed care more about privacy than others and choose not to buy smart home devices [10]. Some consumers do not consider privacy or security before purchasing [14], or deem them less important than other factors [14, 40]. Consumers' expectations and concerns about privacy and security are also shaped by their preceding experiences with computing technologies and their underlying organizations, causing them not to expect privacy by default [35]. Such expectations pose consequences to individual privacy valuations in smart home device purchases, given that consumers may be loss-averse with regard to personal privacy, giving it more value when their current stance "includes" it and lower value otherwise [5].

These studies point to a complex setting where it is unclear where privacy and security considerations may stand in smart home device purchase decisions, suggesting a nuanced assessment of the interplay between functionality, price, convenience, and privacy and security risks. Such a setting also poses an interesting opportunity for the study of privacy given the long-established association of a home and privacy. To the best of our knowledge, no prior works have tried to unpack this interplay and quantify considerations at a large scale. Findings from previous works also point to different, perhaps segmented consumer priorities when considering the purchase of smart home devices, where some may prioritize privacy while others may not, yet this potential segmentation has not been explored.

### 2.3 Distinctions from Prior Work

Interview studies and surveys have found that privacy and security risks are a major consideration with regard to smart home device adoption. These studies have also shown that some consumers may prioritize price over privacy while others will not, and that there is a tension between privacy and functionality, both of which are common considerations. However, to the best of our knowledge, there has not been an attempt to systematically quantify to what extent privacy and security are important alongside other factors such as price, expected convenience, and interoperability. For example, is privacy more of a blocker than price is a motivator? Answering these questions will give developers and researchers a more contextualized understanding of what considerations are being made alongside privacy, giving them knowledge to design effective privacy features and tools.

More importantly, prior interviews, surveys, and smart home adoption statistics suggest privacy and security may not be considerations held by all consumers, and based on our quantification, we conducted a clustering analysis of consumers. Our goal was to understand whether consumers can be segmented with regard to their priorities, and reveal whether there would be one segment of consumers who would be privacy-oriented, and if so, what might the other competing clusters be. This is important to understand given that interview studies with smart home device owners revealed that they did not pay attention to privacy before purchasing, but only became concerned after the purchase. [14]. Uncovering potential segments will provide opportunities for education and awareness to cater to consumers where they stand with respect to privacy and security. For example, if low price is a strong motivator for a segment of consumers, educating them on potential privacy/security risks with Internet-connected products could prevent undesired privacy/security outcomes.

While privacy concerns, expectations, and valuations around the smart home devices have been studied, potential privacy behaviors, tools, and protections remain largely under explored. Such exploration could, for example, identify user-centered privacy features within the context of the smart home. To address this, we present a ranked, comprehensive list of privacy tools, behaviors, and features of smart home devices desired by survey participants.

## 3 Method

In this section, we present details of the survey design and the data analysis procedure.

### 3.1 Survey

The present work encompasses a data analysis from a survey conducted on Amazon Mechanical Turk (AMT) introduced in a previous paper of our authorship [5]. In the present work, we report on a different portion of the survey, not previously reported on. One of the survey's goals was to collect user preferences for different smart home information flows, then create machine learning models to predict such preferences, which was the contribution of the previous paper. Another goal of this survey was to unpack the decision-making process of consumers on smart home device adoption, focusing on where privacy and security concerns and desired protections stand within such considerations. The latter goal had not been addressed before and is the topic of the present work.

The survey presented randomly generated vignette scenarios combining different attributes and purposes of use in the template "The manufacturer/developer of your smart home device is accessing or inferring [attribute]. They are using this information for [stated purpose]," asking participants to provide their comfort levels and preferences on whether they would allow or deny a given information flow. Each participant was presented with four of such scenarios. In addition, for each scenario, participants were asked to review and select up to three out of 14 transmission principles that could make them more or less comfortable with the original

scenario. In [5], we also present an analysis and modeling from an economics-related question about the purchase of a voice assistant, in which participants were asked to specify how much they would be willing to receive or pay extra for privacy protections in the purchase of such voice assistant. The survey received responses from 698 participants, with a median completion time of 19 minutes and compensation of $1.50 (USD), plus equivalent bonuses for participants who took longer than average to finish. The survey was approved by Syracuse University's IRB and the survey protocol can be found in Appendix A.1. More details about the scenario and economics-related questions of the survey can be found on the previous paper [5].

Participants of the survey were required to have 95% of all-time approval rate on previously submitted work and be based in the US. Participants also had to pass a manual qualification task which required them to select three out of six devices they believed to be smart home devices after having read briefly about them. A brief explanation with three short paragraphs largely based on the Wikipedia definition for smart home devices preceded the qualification task, along with three pictures, one with an Ecobee smart thermostat, another with an Alexa smart speaker, and a third one with a Nest smart camera. Then users were asked to select three images containing a smart home device, out of images (with alternative text) of a smart thermostat, a voice assistant, a blender, a DSLR camera, a desk lamp, and a smart bulb. Only users who selected the three smart home devices, namely the smart thermostat, the voice assistant, and the smart bulb, were allowed to proceed.

In the present work, we analyze the data from three open-ended questions asked in the survey that have not been previously reported on. Two of the three questions were presented *immediately after* the qualification task and *immediately before* the four random scenarios. The two first questions followed an introductory question posed to elicit participants' thought processes: *What factors do you consider when making decisions about adopting smart home devices? Please answer below.* Then two follow-up questions were asked:

1. *For example, what are factors that could motivate you to purchase smart home products?*

2. *Similarly, what are factors that could keep you from adopting the technology?*

Privacy or security were not mentioned during the qualification step nor in the two questions as not to prime participants. This helped avoid any privacy or security-related bias in participants' thought processes when answering the questions.

The third question was presented *immediately after* participant responses to the four scenarios and *immediately before* the economics-related question reported in the previous paper:

> 3. *What privacy behaviors you would like to be able to adopt in the context of smart home devices? For example, would there be any privacy-protecting tools, configurations, and techniques you would like to use?*

Answers to the three questions were mandatory, and we did not use or report on the data collected from any of these three questions before. We acknowledge that answers to the third question could include biases resulting from the four scenarios presented earlier in the survey, and this is a limitation of the answers to this question. For example, participants who responded to a scenario where the purpose of use was targeted advertisement may have been primed to mention protections against secondary use. We still analyze and report the data given that (1) participants had a comparable experience because scenarios were created randomly; (2) the analysis on this question is a secondary contribution of our work and (3) the answers still provide valuable insights that have resulted from participant's engagement with a survey focused on potential information flows of the smart home. The survey scenarios also indirectly provided a broad grounding around the potential privacy and security risks associated with using a smart home device, enabling them to provide contextualized and meaningful responses, as evidenced by the level of articulation observed in participants' responses to question 3 (see Table 2 in Appendix).

Following the third question within the scope of the present work (i.e., the question about privacy tools, configurations and techniques), participants were asked the economics-related question presented in [5], questions from the Awareness, Control, and Collection dimensions of the Internet User Internet Privacy Concerns (IUIPC) scale [22], and demographics: gender identity, age bracket, hours spent on the Internet weekly, whether they owned a smart home device, how many smart home devices owned, what specific devices were owned (from a list of 16 types), occupation, education level, income bracket, size of household, whether the participant had children, and marital status. We tested whether any of these demographics would be associated with mentioning privacy or security as a motivator or a blocker in questions 1 and 2.

We initially read each response to check the quality of the responses. We manually inspected the answers to each question and removed responses from 67 participants (9.6%) due to their answers not being meaningful and/or being random copy/paste. Our cleaned up data set resulted in responses from 631 participants. This cleaning process generated the data set used in our qualitative and quantitative analyses.

### 3.1.1 Participant Demographics

**Gender Identity and Age** 48.8% identified as female (50.7% male, 0.5% other), 44% as 26-35 years-old, 21% as 36-45 years-old, 16% as 18-25 years old, 9% as 46-55 years old, and 10% over 56.

**Education and Income** 39% of participants reported having a Bachelor's degree, followed by some college but no degree (21%), master's (14%), associate (13%), high school

(9%), professional (3%) and doctoral (1%). 24% of participants reported earning no more than $30k, 34% no more than $60k, 19% no more than $90k, and 22% over $100k.

**Household Size** The average household size was 2.73 (Mdn=3, SD=1.38). 45% reported being married, 45% single, 8% divorced, 2% separated, 1% widowed. 44% of respondents reported having children.

**Occupation** Participants reported a diverse set of occupations, including agriculture, sales, therapist, teacher, attorney, software engineer, student, insurance worker, and accountant. 9% of respondents provided an IT-related occupation.

**Device Ownership** 48% reported owning a smart home device. The most popular type of device owned was voice assistant, followed by security camera, smart lighting, audio/speakers, and thermostat.

**IUIPC Scores** We added up the score for the responses to the questions within each corresponding dimension. The average Awareness score was 19.23 (Mdn=21, SD=2.68, Min=7, Max=21). The average Control score was 18.4 (Mdn=19, SD=2.8, Min=6, Max=21). The average Collection score was 23.68 (Mdn=25, SD=4.56, Min=6, Max=21).

## 3.2 Data Analysis

### 3.2.1 Privacy/Security and IT Annotation

With a focus on privacy and security, we annotated each row with whether the participant referred to privacy or security as a motivator or a blocker. As a first step in our analysis, we grouped privacy and security responses together because prior works found that IoT users had limited knowledge of privacy and security and often could not distinguish between them [14]. This grouping also enabled us to start with a high-level analysis involving descriptive and test statistics. Examples of when we flagged privacy or security are if participants mentioned "privacy concerns" or said "hacking, " "tracking and monitoring," or "stalking me to market to me." We later used this annotation to generate descriptive statistics about the overall number of responses mentioning privacy or security, in addition to conducting statistical significance tests for relationships with demographics. Although we grouped privacy and security responses for a high-level analysis, we considered them separately during the coding and segmentation analyses. We also report the statistical test results when considering privacy and security separately.

Additionally, we annotated whether each participant's occupation was related to IT. We did this because we wanted to be aware in our analysis when participants could have heightened technical expertise. Some occupation examples where participants were marked include "IT Help Desk Analyst," "Software QA," "Programmer," and "Computer Technician." Only 9% of participants reported an occupation related to IT.

### 3.2.2 Coding

We conducted inductive coding on the open-ended participant responses, coded by two researchers. We first read each answer in order to get acquainted with the responses and underlying, recurring themes. Then, we drew a random sample of 15% of responses and coded them individually. After coding the sample individually, the two researchers met in person to review their individual codes, discuss, and converge into a code book that would be used for the remaining of the responses. A code named "other" was created for which answers not belonging to any of the codes in the final code book were assigned. The final code book contained 23 categories for motivators, 20 categories for blockers, and 19 categories for privacy tools and behaviors. Using this code book, the two researchers coded the remaining 85% of the responses. In our coding procedure, each answer was allowed to have more than one category. We calculated inter-coder agreement between the two coders using Cohen's Kappa: 87% for motivators, 91% for blockers, and 88% for privacy tools. These values indicate excellent agreement between the two coders [15].

### 3.2.3 Quantitative Analysis

We merged the coded data sets resulting from the coding procedure based only on the agreements between the two coders. For example, if both coders assigned the same given category to a response, then the category was assigned in the final data set, indicated with a value of 1, otherwise this value was 0. Once our final data set was generated, our quantitative analysis was divided into three parts.

The first part consisted of testing relationships of demographics with whether participants reported privacy or security as a motivator or blocker. We used Chi-square association tests for categorical variables such as gender identity, or owning a smart home device, and logistic regression for numerical and ordinal variables, such as the age bracket, education level, and IUIPC awareness, control, and collection dimensions.

The second part consisted of analyzing the frequency of each factor either as a motivator or a blocker, and quantifying the relative importance of the factors side-by-side. The latter task involved creating a wide data set with each column representing a factor mentioned either as a motivator or a blocker from the coded and merged data set. If the factor was mentioned in the motivator question, the column was given the value of 1. If the factor was mentioned in the blocker question, the column was given the value of -1, and 0 otherwise (i.e., not being mentioned in either). This allowed us to compare motivators and blockers side by side, surfacing whether each factor is more of a motivator or a blocker, as determined by their calculated average values. For instance, a device being privacy-invasive might be a blocker whereas not being privacy-invasive might not be a motivator.

The third part consisted of conducting a clustering analysis with the considerations and creating a decision tree model to

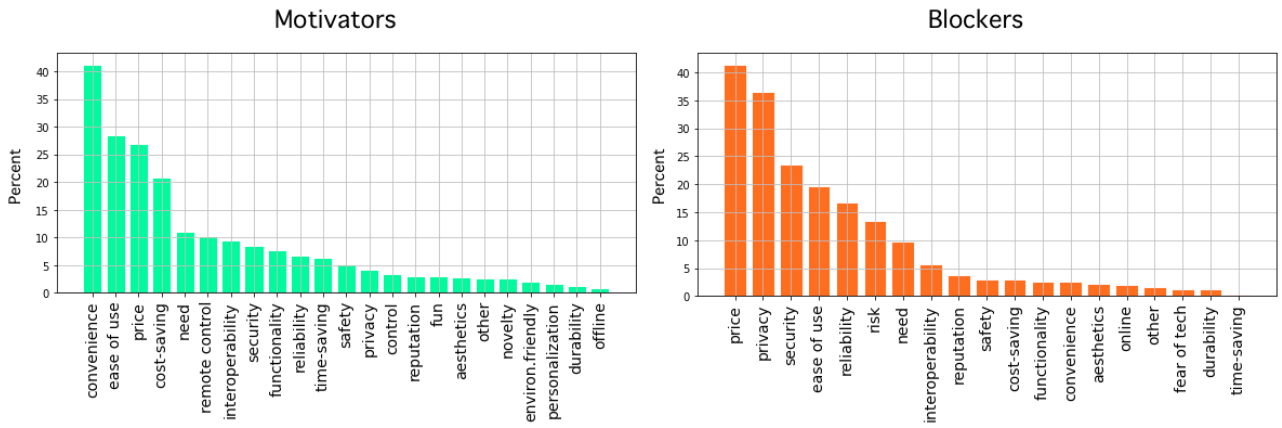


Figure 1: Frequency (%) of motivators (left) and blockers (right) across all survey respondents. Top motivators are convenience, ease of use, price, cost-saving, and need. Top blockers are price, privacy, security, ease of use, and reliability.

predict the assigned cluster based on participant considerations. With the wide data set encoded with -1,0,1 columns, we generated a dendrogram to visualize hierarchical clusters. Then, we assigned each response to a cluster with k-means clustering with $k = 3$. This segmented the participants based on their purchase considerations. We report the ranking of motivators and blockers for each of the three participant clusters and cross-checking of the clusters with specific demographics, such as having technical background, owning a device, or the reported gender identity. The last step in this part consisted of creating a classifier to predict the consumer clusters. We created a decision tree classifier and evaluated it with 10-fold cross validation. We report our results, along with the resulting decision tree of the trained model.

# 4 Results

## 4.1 Privacy and Security Considerations

Across all participants, only 11% mentioned privacy and security among the factors which would *motivate* them to adopt smart home devices (separately, privacy=4.12%, security=8.56%). For example, participants mentioned *"if it's non-intrusive"* and *"the security of the system and how protected it is from outside tampering,"* as motivators related to privacy and security. In comparison, 50% mentioned privacy or security as something that would *prevent* them from adopting (separately, privacy=36.61%, security=23.61%). For instance, participants responded with *"the security of the item, could it be hacked? Could it have a camera that could turn on and be hacked? Would my personal information be safe?"* and *"mostly, companies obtaining information on my personal life. Just because consumers buy from a company doesn't mean the company can own them."* Among participants who reported owning a smart home device, 44% mentioned privacy or security concerns as a blocker (separately, privacy=16%, security=10%), while this number was 56% for participants who reported not having a smart home device (separately, privacy=21%, security=14%). A Chi-square association test examining the relationship between having or not having a smart home device and mentioning or not mentioning privacy or security as a blocker produces a statistically significant result: $\chi^2$ (1, N = 631) = 8.901, $p < 0.001$, suggesting that people who have privacy and security concerns are less likely to be associated with having a smart home device. When testing for privacy and security separately, this relationship is also significant: privacy $\chi^2$ (1, N = 631) = 4.041, $p < 0.05$, and security $\chi^2$ (1, N = 631) = 3.8942, $p < 0.05$. A Chi-square association test showed no difference between having one *versus* multiple devices.

Further, we investigated the relationship between privacy or security considerations and the demographics collected in our study. Namely, we tested gender identity, age bracket, whether participants had an IT-related occupation, education, income bracket, household size, whether participants had children, and marital status. Given that there were no pre-planned hypotheses or theoretical model for testing these demographics, we applied Bonferroni correction to control family-wise Type I errors, thus taking .00625 as our significance level considering 8 tests. None of the tests yielded statistically significant results at the corrected p-value. The results were the same when testing for mentioning privacy and security separately.

Finally, we tested the relationship between the IUIPC constructs and stating a privacy or security consideration with a logistic regression model using the three IUIPC dimensions as predictors. For both the motivator and blocker question, the IUIPC Collection dimension was a significant predictor ($p < .05$, exp(estimate) = 1.06 for blocker, 1.11 for motivator), indicating that people who were more concerned about data collection in general (based on the IUIPC) were more likely to be associated with mentioning privacy or security considerations in our study. In a model comparison via the


Figure 2: Ranked importance of considerations. The X axis represents the absolute average value of each consideration. If the non-absolute average value is positive, color is green, or red otherwise. While convenience and cost-saving are the top motivators, privacy, price, security are the top blockers.

Likelihood Ratio Test, both the motivator and blocker models containing the IUIPC predictors resulted in a statistically significant difference against the null model ($p < .001$).

These results suggest that privacy and security considerations may be preventing actual adoption, may not be associated with particular demographics, and may have a relationship with how participants felt about online data collection.

## 4.2 Relative Importance of Motivators and Blockers

For all respondents, we calculated the frequency of each motivator and blocker. The top five motivators were convenience, mentioned by 41.2% of participants as a motivator, ease of use (28.37%), price (26.94%), cost-saving (20.76%), and need (10.94%). The top five blockers were price, mentioned by 41.36% of participants, followed by privacy (36.61%), security (23.61%), ease of use (19.65%), and reliability (16.8%). Figure 1 shows the percentage of participants who mentioned the motivators and blockers, and Table 1 (Appendix) shows all factors with examples. 234 participants (37%) mentioned at least one factor both as a motivator and a blocker, with the most frequent being price, with 18.2%, then 11.6% for ease of use, 5.1% for security, 3.8% for need, 3% for reliability, 2.5% for privacy, 2.5% for interoperability, then six more factors mentioned as both motivators and factors by fewer than 2% of participants, with the remaining 15 factors being mutually exclusive, meaning they were either mentioned only as a motivator or as a blocker.

When combining motivators and blockers via their average values across all participants, it is possible to determine whether a factor was mostly a motivator or a blocker. Figure 2 shows the distribution of the factors ordered by their absolute average value. The top motivator is convenience, followed by privacy, price, and security as top blockers. This suggests that most consumers might consider the three top blockers after convenience, then whether the device will save money in the long-term, then the risk of owning the device, etc, according to the ranking in Figure 2.

## 4.3 Clustering Consumers

One of our research goals was to examine whether participants could be clustered with regard to their purchasing considerations. To do this, we used the wide data set with each possible consideration as a column, resulting in 29 columns of value 1 if it was a motivator, -1 if it was a blocker, and 0 otherwise. We identified the optimal number of clusters via a dendrogram generated from Agglomerative Clustering, a bottom-up hierarchical clustering approach. The dendrogram analysis (Figure 3) revealed three major clusters, as indicated by the number of vertical lines crossed by the horizontal black line placed at the end of the longest vertical line. Agglomerative clustering starts by assigning each data point to its own cluster, then moves up, grouping instances based on the smallest distance, such as the Euclidean distance, eventually making all data points belong to a single cluster. The optimal number of clusters is chosen by crossing a horizontal line over the longest vertical line and verifying how many vertical lines

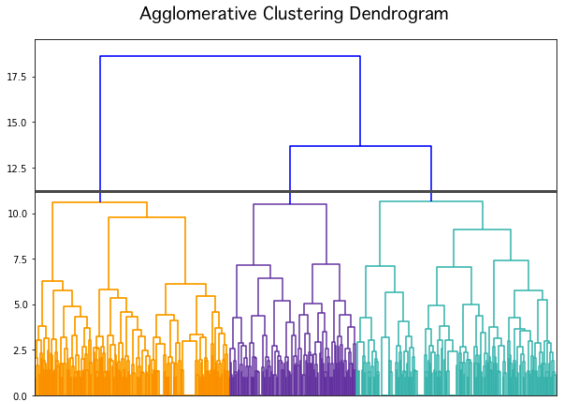


Figure 3: Hierarchical clustering dendrogram. The graph reveals three clusters, as indicated by the number of horizontal lines crossing the longest vertical line.

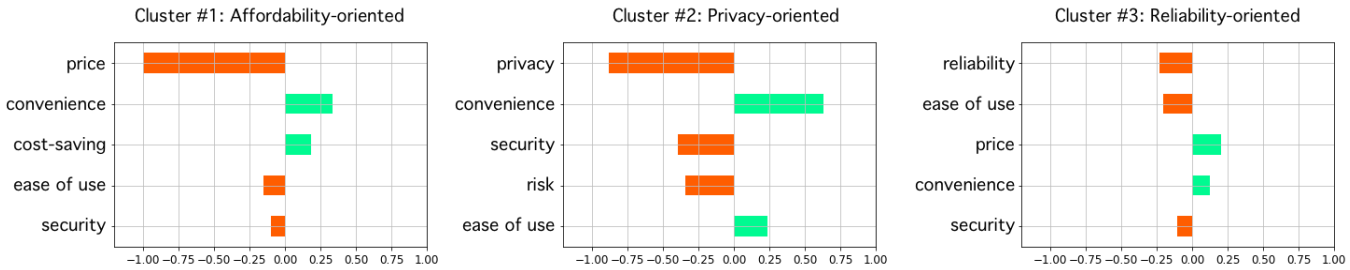


Figure 4: Top five considerations in each of the clusters. Negative and red means the factor is a blocker, positive and green a motivator. Clusters were largely defined by price, privacy, and reliability being blockers.

cross this horizontal line. Clusters are indicated by colors.

Knowing that the optimal number of clusters was three, we clustered participants using k-means with $k = 3$. The clustering resulted in 28% of participants being assigned to the cluster named reliability-oriented, 34% to the cluster named privacy-oriented, and 38% being assigned to the affordability-oriented cluster. We named the clusters based on the sorted absolute average value of factors within each cluster in order to represent the importance of the motivator or blocker. Figure 4 shows the top five considerations for each cluster. The fact that the clusters are largely defined by price, privacy, and reliability blockers suggests that consumers may be segmented with regard to reservations that they may have on these factors. While the average value for the top consideration in the reliability cluster is smaller than the other clusters, a triangulation analysis of quantitative and qualitative data further reinforces the segmentation: reliability is the main and differentiating factor of this cluster (e.g., ease of use is a common factor across all clusters). Also, reliability is not pronounced in the other two clusters, with averages around -0.1, and open-ended responses from this cluster show a recurring theme of reliability, e,g. “fear of malfunction.”

Participants assigned to cluster #1 would see price as a major blocker. For example, if the device is not affordable or too expensive. Participants in cluster #2 would see privacy risks as a major blocker. Finally, participants in cluster #3 might not purchase a device if it is not reliable. For instance, consumers in this cluster would care about how dependable or high-quality the device is, and what happens when Internet connection is lost. These clusters revealed that consumers may approach their decision-making process with different priorities, and that consumers who value price and reliability may not particularly consider privacy as a major factor.

Knowing that the identified clusters were related to privacy, affordability, and reliability being mentioned as blockers, we conducted additional statistical tests in order to understand whether any of the demographics would be associated with participants mentioning price or reliability as blockers. Using the corrected p-value of .00625 (.05/8 demographics), none of the tests came out significant. In other words, we did not find significant relationships between participants mentioning price or reliability and their reported demographics. We also conducted individual multinomial logistic regression analyses where the dependent variable was the cluster and demographics the independent variables, using separate models for each demographic. The results were the same: no statistically significant relationships found between the assigned consumer cluster and people’s demographics.

The percentage of participants in the privacy-oriented cluster who reported not owning a smart home device was 56%, whereas this percentage was 49% for the other two clusters. A Likelihood Ratio Test of a multinomial logistic regression model with the cluster as the dependent variable and whether participants reported owning a device did not produce a statistically significant result.

## 4.4 Cluster Classification and Decision Tree

We created a decision tree classifier to predict the cluster of each participant and elucidate/reconstruct the decision-making process of participants in each cluster. The goal of this classifier is to be able to segment consumers based on the considerations they might have. For example, one could use our classifier by asking users to select among the factors we identified in our study which ones they consider as motivators or blockers. Separating motivators from blockers in this analysis is important given how consumers may have different considerations in their purchase decisions. For example, while a product not being environmentally-friendly may not be a blocker, being environmentally-friendly may become a motivator. In other words, separating motivators and blockers can uncover more nuanced decisions. Then, based on the selections, a cluster can be assigned to a consumer which will help understand the consumer’s priorities. Does the consumer prioritize price, privacy, or reliability more? The interpretation of this decision tree classifier can uncover how considerations of motivators and blockers can segment consumers.

We initially evaluated a classifier using all motivators and blockers, without specifying a maximum tree depth, with 10-fold cross-validation. This classifier achieved F-1 scores

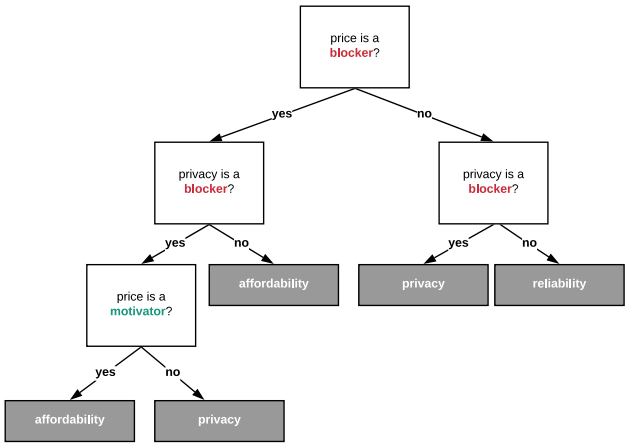


Figure 5: Simplest Decision Tree Classifier to assign clusters based on top considerations. The decision tree suggests that even if privacy is a blocker, consumers may focus on affordability if price becomes a motivator.

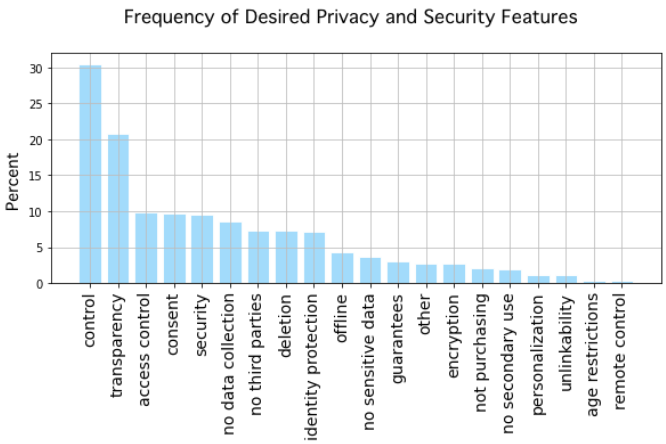


Figure 6: Frequency (%) of desired privacy and security features. The top features are control, transparency, access control, consent, security, no data collection, no third parties, deletion, and identity protection.

from 97% to 98% for all clusters, with a resulting tree depth of 7. The resulting tree included considerations about price, privacy, time-saving, convenience, interoperability, need, ease of use, remote control, and safety. While this classifier helped elucidate the decision-making process with regard to purchase considerations, it was overly complex to interpret.

In order to arrive at a more practical solution – one with good performance and interpretability – we empirically tested different numbers of factors and tree depths using 10-fold cross-validation. The best classifier used only the privacy and price motivators and blockers and a tree depth of 3. F-1 scores for each cluster were 98% for privacy-oriented (Precision=99%, Recall=98%), 100% for reliability-oriented (Precision=100%, Recall=99%), and 99% for affordability-oriented (Precision=98%, Recall=99%). The error rate from 10-fold cross-validation was 1.6%. We then trained this classifier with all of our data and generated the decision tree shown in Figure 5. Based on the process outlined in the decision tree, participants would be affordability-oriented if price is a blocker and privacy is not a blocker or if price is a blocker and affordability is a motivator. Participants would be classified as privacy-oriented if price and privacy are blockers and price is not a motivator, or if price is not a blocker but privacy is. Finally, participants would be classified as reliability-oriented if neither price or privacy are blockers. The fact that the decision tree classifier was able to be trained with only two features and achieve good performance shows that it is likely the decision-making of consumers might rely heavily on price and privacy assessments. For example, the decision tree shows that even if privacy is a blocker, consumers may still be influenced by price if it becomes a motivator.

## 4.5 Desired Privacy Features

Survey respondents provided 19 unique privacy features they wish were available. The top desired privacy features in the survey were: control, mentioned by 30.43% of participants, followed by transparency (20.75%), access control (9.83%), consent (9.67%), security (9.51%), no data collection (8.56%), no third parties (7.29%), deletion (7.29%), identity protection (7.13%), offline operation (4.12%), no sensitive data (3.65%) and guarantees (3.01%). Figure 6 shows frequencies of all features as percentages of participants who mentioned them, and Table 2 (Appendix) shows all codes, along with examples.

**Control** Three types of control were mentioned by participants: physical control such as shutting off the devices, controlling what data are collected, and data use opt-out. For example, P99 noted *"Being able to choose exactly which data is being collected and how it is used. Have complete control."*

**Transparency** Transparency features focused on having the manufacturer/developer show users what data are being collected, whom they are shared with, for what purposes, and whether their system was vulnerable or breached. Participants also mentioned wanting to have such information periodically such as weekly or monthly, and receiving notifications and/or seeing physical visual indicators about data activities. For instance, P183 responded *"I would like full reports on where my data is going from smart home devices sent daily or weekly."*

**Access control** Participants whose answers hinted at access control features want to be able to have strong passwords, two-factor authentication, and biometrics, as well as to prevent access to their system by anyone else besides themselves. For example P151 said *"to turn on only when activated and had a voice recognition devices so if someone asked for my information it would not display it."*

**Consent** Consent means that participants wish to be able to know before certain data collection or sharing occurs and be able to allow or deny it. For instance, P175 was concerned about tracking of search habits and noted *"Only allowing tracking of my search history with my consent."*

**Strong security** Participants who wanted strong security emphasized that they wanted to be able to secure the data in storage and in transit, and make sure that their network was secure. As an example, P220 suggested using firewalls and other software, saying *"I would like to have any device that is used online to have a firewall and virus protection program installed with it."*

**No data collection** Participants who explicitly mentioned not wanting the device to collect or share any of their data were representative of this category of features. For instance, P227 noted *"Well I don't know at the moment. As long as I am not being tracked and none of my data is being collected, then I would be fine with whatever security or privacy protection tools available out there to keep me safe. Especially in my own home."*

**No third parties** Participants whose comments fell into this category were explicit about not wanting their data to be shared with any third parties, such as marketing firms or the government. As an example, P270 expressed *"If I were to use one of these devices, I don't want them linked to my identity and I don't want the information shared with third parties. I can see where a utility company might need some of this information in order to bill me properly. But beyond that, I would want a user-friendly interface that allows me to shut off access to anyone else."*

**Deletion** Participants who emphasized deletion either wanted data to be deleted automatically after a certain period, or have the ability to "go in" and delete any data, either via a user interface or a physical reset button. For example, P228 suggested a feature where she could access and delete information at any point in time, saying *"I'd like to be able to delete things regularly. Permanent deletion."*

**Identity protection** Participants who wish this feature were explicit about not wanting to have any of the data associated with their identity, and that they wanted specific protections from it. For instance, P217 suggested *"anything to protect my identity, so only the device and I know it."*

**Offline mode** Offline mode means that participants would want either the device to be offline at all times, or for it not to be online at all times, or for them to be able to control when devices go online or offline. For example, P295 noted *"All smart home devices need to be able to operate offline, without requiring a web app or account with a company. There is no need to gather data and send it to the Internet to operate these devices, they just want to. You should be able to set them up on a local network and control them yourself without them being tied to a brand or company. This helps when Google buys your thermostat company then bricks them."*

**No sensitive data** Participants in this category would not want any sensitive data to be collected or stored. For example, P85 noted *"Anything that would block personal data from being shared. If I have to enter anything personal to use the equipment, I would want to be able to lock it and that it be never stored. And that I have the say so of what data about me I consider private and personal."*

**Guarantees of privacy and security** Participants who mentioned this feature wanted to be given guarantees either by the manufacturer or the applicable laws that their privacy and security would be protected and that there would be penalties otherwise. For instance, P117 expressed *"Auto-deletion of pertinent data and guarantee, with legal repercussions, that data will not be shared."*

While some of these features have been uncovered in user-centric studies before, such as control, transparency, strong access control, an offline mode [14, 39], our results show the frequency in which such privacy features were mentioned, which can help developers and regulators navigate what is most important for their users with regard to users' privacy and security considerations.

Finally, we verified whether the most desired privacy features would be different for people in each consumer segment according to our clustering analysis, and the top two features wished for are the same for the three segments: control and transparency, accounting for more than 30% of the responses in all three segments.

# 5 Discussion

## 5.1 Summary of Findings

Our findings suggest convenience, ease of use, price, cost-saving, need, remote control, and interoperability are the top motivators for consumers to adopt smart home devices. The top blockers are price, privacy, security, ease of use, reliability, risk, and lack of need. Our analyses showed that participants who see privacy or security as a blocker in purchasing decisions were less likely to own smart home devices at the time of the survey, and that considerations of privacy and security were not associated with demographic traits such as gender identity, age, education or income. Our clustering analysis uncovered three consumer segments with regard to their purchase considerations: affordability-oriented, privacy-oriented, and reliability-oriented. The most desired privacy protections for smart home devices from the survey are: control, transparency, strong access controls, consent, security, no data collection, no third parties, deletion, identity protection, offline operation, no sensitive data, and guarantees.

## 5.2 Paradox or (Bounded) Rationality?

Our results indicate that consumers heavily weigh privacy and security as blockers, and that these may be preventing them from adopting smart home devices. This finding suggests

that a paradox may not be the explanation for the mixed adoption signals – perhaps privacy-oriented consumers are really not buying smart home devices. On the other hand, our decision tree analysis showed that even if privacy is a blocker, consumers may still value affordability over privacy if the price becomes a motivator. This becomes a potential problem that could lead to undesired outcomes when such devices are sometimes given for free as part of promotions from big tech companies, such as Spotify giving the Google Home Mini to premium subscribers [33], or Google giving the Google Home Mini to Pixel 2 phone buyers [27] and even randomly [17]. In such cases, bounded rationality may lead consumers to overlook privacy considerations, which may become a concern only after a device has been installed [14].

The fact that our decision tree was effective with only the privacy and price feature suggests what consumers may be ultimately considering is whether a low price point for these devices is worth the expected, associated privacy risks. Such price–privacy interplay has been noted in prior work which suggests users would pay a premium for privacy [5, 12, 36], but that even in the context of the smart home, they would pay less than they would be willing to take in exchange for it [5]. Considering our findings in light of prior work, we posit that: (1) privacy-oriented consumers may expect privacy by default or else many may not adopt such technologies, (2) making privacy and security a motivator attached a higher premium may not work well, and (3) if privacy concerns are only a consumer afterthought due to bounded rationality or if a device is extremely affordable, then the market puts consumers in a "privacy-not-included" scenario and thus the value of privacy may be inadvertently or unintentionally diminished.

Paradoxically, one could argue that the knowledge revealed through our analyses can give opportunistic developers more tactical information to manipulate consumers in ways that motivators are highlighted such that privacy considerations remain obscure. This could reinforce bounded rationality and lead to decisions against the consumer's best interest [28]. The lack of visibility of privacy and security-related information on smart home devices is an existing problem, and legislators on this topic have proposed adding concise and accessible labels (e.g., [16, 37]), but little guidance has been provided on how they should be presented. Recent research efforts have focused on how to implement such labels (e.g. [13, 18]) in order to educate and equip consumers and prevent decisions driven by bounded rationality. We endorse such efforts, and based on our segmentation analysis, we further posit that more transparency is needed in order to equip consumers to make informed, rational decisions that suit their specific needs and expectations with regard to the interplay between price, privacy, and reliability. Accordingly, we present several practical recommendations for device developers and policymakers based on our findings which could lead to more informed consumer decisions and meaningful device comparisons.

## 5.3 Recommendations

### 5.3.1 Device Developers

Our quantification enables developers to market products based on whether something is more of a motivator or a blocker. For example, marketing strategies may focus on showing empathy about users perceiving adoption coming with a risk of privacy and then offering certain guarantees that their data are not to be shared or used for secondary purposes. Other examples may include highlighting a device's reliability when the Internet goes down, or estimating long-term savings, as this is important for affordability-oriented consumers.

Consumer segmentation based on considerations has further implications for targeting of the products. For example, the reliability-oriented segment may care more about what happens when the Internet goes down, whereas the privacy-oriented segment may care more about whether their data will be shared with third parties or they will be monitored without consent. For this reason, device developers could clarify how devices consider the segment-differentiating factors (i.e., price, privacy, reliability) in order to help consumers find the right device for them among the options available.

The properties that segmented the participants may also be inherently at conflict. For example, if a smart camera is to have onboard object recognition capabilities so that it can work offline both for privacy and reliability reasons (when the Internet goes down), it may end up costing more. Surfacing and communicating such trade-offs to consumers may be a promising strategy and prevent instances of bounded rationality. For instance, our proposed decision tree model could be used by retailers and developers to ask consumers four questions on whether privacy or price are motivators or blockers to them, then determine what their consumer segment is in order to better inform their purchasing.

Some of the privacy features uncovered in our analysis have been identified via qualitative studies before. For example, Yao *et al.*'s co-design study [39] uncovered control, transparency, offline and private modes, and Emami-Naeini *et al.*'s study [14] uncovered strong access controls. We contribute further with a comprehensive list of features presented in our paper, which can guide developers in prioritizing and implementing tools and features that may enable them to appeal to consumers in the privacy-oriented segment. For example, a developer can draw from this list to implement increased control, transparency, strong access controls, consent, security, no data collection, no third parties, deletion, identity protection, offline operation, no sensitive data, and guarantees.

### 5.3.2 Policymakers

Consumers in the affordability-oriented segment may not have privacy as a strong consideration due to not being educated about Internet business models, which is associated with the aforementioned bounded-rationality issue of IoT devices. The

segmentation we presented could serve as a framework for policymakers to approach the design of privacy regulation in ways that meet consumers where they stand. Ideally, a device would be affordable, privacy-preserving, and reliable. However, special consideration must be given to instances where such desirable properties may be at conflict. For instance, it may be ineffective to implement a privacy feature that poses a trade-off with reliability if a device's target audience is largely reliability-oriented, such as making a smart camera work offline for privacy reasons, yet doing so might limit its reliability such as being unable to recognize objects when offline. Another strategy that might prove ineffective is to offer extra privacy features for an additional premium to affordability-oriented consumers. Notwithstanding, more effort should be placed on regulating the communication of risks involved in owning a smart device (e.g., privacy and reliability), especially when the smart version is cheaper or given away for free as part of promotions. Arguably, more clarity about such risks becomes progressively important as consumers are offered increasingly fewer non-smart device alternatives in the future.

It might also be beneficial for policymakers to introduce requirements for developers to practice tailored consumer education about device privacy/security risks according to individuals' corresponding cluster derived from our decision tree. For example, individuals belonging in the privacy-oriented segment could be given privacy-related device details while seeing only summaries on the other cluster-defining aspects.

## 5.4 Limitations

Our survey was conducted in the US. This means that the results from our data analyses may not necessarily represent the state of smart home purchase considerations elsewhere, nor it is representative of the universe of considerations that may exist. As it applies to survey data more generally, the data collected may also be subject to the availability heuristic. In addition, privacy considerations, expectations, and awareness are known to be diverse across different geographies and cultures. We do note, however, that smart home growth has been observed heavily in the US [34].

The responses provided by participants to the question about privacy behaviors, tools, and features may carry a bias from the survey design. This is because participants were exposed to different data collection scenarios throughout the survey where they were asked to express their preferences. Nonetheless, we do not find this bias to compromise the quality of our data since every participant was assigned such data collection scenarios randomly, and thus had comparable experiences. We note, however, that had the question been asked prior to the data collection scenarios, perhaps participants would be less aware about certain features, but would have been primed for privacy when responding to the scenarios analyzed in [5]. The scenario questions may also have influenced the IUIPC questionnaire responses.

While our participant sample was diverse with regard to demographics and socio-economical status, it may not be representative of the US population, as participants on AMT may skew towards people with non-traditional forms of employment and people with heightened technical expertise.

## 5.5 Future Work

We examined consumer considerations about smart home devices in general, but there could be differences in the decision-making across specific devices [14]. For instance, reliability may be more important for a smart lock, whereas privacy may be more important for a camera. Future works could conduct similar analyses considering different devices.

Follow-up experimental studies could be conducted to validate our findings. For example, a potential study may incorporate our decision tree questions to predict which segment consumers belong in (i.e., price, privacy, or reliability) and verify whether the predictions match consumer priorities.

Future studies could explore how interventions could "move" a consumer from one segment to another. In other words, how stable would a person belong to anyone cluster? Would privacy-oriented consumers be more stable than affordability-oriented consumers?

# 6 Conclusion

Smart home device adoption continues to grow steadily, yet privacy and security concerns reportedly remain a roadblock to mass adoption. User-centric qualitative studies have revealed that consumers often consider price, features, and privacy risks when making smart home device purchases, but no studies have attempted to quantify these considerations in a systematic way. Previous studies have also found that many consumers do not consider privacy at all before purchasing a device. We conducted a mixed-method analysis using online survey data collected from 631 participants based in the US. Our analyses show that privacy and security are considered blockers for half of the participants, but more so for participants who reported not owning a device. We found that convenience, ease of use, price, cost-saving, and need are top motivators and price, privacy, security, ease of use, and reliability are top blockers. We conducted a customer segmentation analysis which revealed three clusters: affordability-oriented, privacy-oriented, and reliability-oriented. A decision tree classifier to predict customer segments revealed that even privacy-oriented consumers may be influenced by a device's price if it becomes a motivator. Finally, we present a comprehensive list of desired privacy behaviors, tools, and features reported by our survey participants. From our findings, we define and discuss implications for the targeting, legislation, and privacy design of smart home devices.

## 7 Acknowledgments

We thank our survey participants for their insightful and articulated responses. We also thank the anonymous reviewers for their thoughtful comments and suggestions. This work was in part supported by the National Science Foundation (NSF) grant number CNS-1652497.

## References

[1] Alessandro Acquisti and Jens Grossklags. Privacy and rationality in individual decision making. *IEEE security & privacy*, 3(1):26–33, 2005.

[2] Alessandro Acquisti, Leslie K John, and George Loewenstein. What is privacy worth? *The Journal of Legal Studies*, 42(2):249–274, 2013.

[3] Idris Adjerid, Eyal Peer, and Alessandro Acquisti. Beyond the privacy paradox: Objective versus relative risk in privacy decision making. *MIS Q.*, 42(2):465–488, June 2018.

[4] Nor Hazlin Nor Asshidin, Nurazariah Abidin, and Hafizzah Bashira Borhan. Perceived quality and emotional value that influence consumer's purchase intention towards american and local products. *Procedia Economics and Finance*, 35(3):639–643, 2016.

[5] Nata M Barbosa, Joon S Park, Yaxing Yao, and Yang Wang. "what if?" predicting individual users' smart home privacy preferences and their changes. *Proceedings on Privacy Enhancing Technologies*, 2019(4):211–231, 2019.

[6] A.J. Bernheim Brush, Bongshin Lee, Ratul Mahajan, Sharad Agarwal, Stefan Saroiu, and Colin Dixon. Home automation in the wild: Challenges and opportunities. In *Proceedings of the 2011 SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*, 2011.

[7] Karen L Courtney. Privacy and senior willingness to adopt smart home information technology in residential care facilities. *Methods of information in medicine*, 47(01):76–81, 2008.

[8] Karen L Courtney, George Demeris, Marilyn Rantz, and Marjorie Skubic. Needing smart home technologies: the perspectives of older adults in continuing care retirement communities. *Informatics in Primary Care*, 16(3), 2008.

[9] George Danezis, Stephen Lewis, and Ross J Anderson. How much is location privacy worth? In *4th Annual Workshop on the Economics of Information Security (WEIS'05)*, 2005.

[10] George Demiris, Brian K Hensel, Marjorie Skubic, and Marilyn Rantz. Senior residents' perceived need of and preferences for "smart home" sensor technologies. *International journal of technology assessment in health care*, 24(1):120–124, 2008.

[11] Serge Egelman, Adrienne Porter Felt, and David Wagner. Choice architecture and smartphone privacy: There's a price for that. In *11th Annual Workshop on the Economics of Information Security (WEIS'12)*, 2012.

[12] Serge Egelman, Janice Tsai, Lorrie Faith Cranor, and Alessandro Acquisti. Timing is everything? the effects of timing and placement of online privacy indicators. In *Proceedings of the 2009 SIGCHI Conference on Human Factors in Computing Systems (CHI'09)*, 2009.

[13] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. Ask the experts: What should be on an iot privacy and security label? In *41st IEEE Symposium on Security and Privacy (S&P'20)*, 2020.

[14] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. Exploring how privacy and security factor into iot device purchase behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI'19)*, 2019.

[15] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical methods for rates and proportions*. John Wiley & Sons, 2013.

[16] Ftc comment to the national telecommunications information administration on "communicating iot device security update capability to improve transparency for consumers", 2017. https://www.ftc.gov/policy/advocacy/advocacy-filings/2017/06/ftc-comment-national-telecommunications-information.

[17] Google is randomly giving away even more free google home mini speakers, 2019. https://www.forbes.com/sites/johanmoreno/2019/09/29/google-is-randomly-giving-away-even-more-free-google-home-mini-speakers/.

[18] Shakthidhar Reddy Gopavaram, Jayati Dev, Sanchari Das, and Jean Camp. Iotmarketplace: Informing purchase decisions with risk communication. 2019.

[19] Jens Grossklags and Alessandro Acquisti. When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. In *6th Annual Workshop on the Economics of Information Security (WEIS'06)*, 2007.

[20] The trust opportunity: Exploring consumer attitudes to the internet of things - internet society, 2019. https://www.internetsociety.org/resources/doc/2019/trust-opportunity-exploring-consumer-attitudes-to-iot/.

[21] 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), November 2018.

[22] Naresh K Malhotra, Sung S Kim, and James Agarwal. Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. *Information systems research*, 15(4):336–355, 2004.

[23] 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 *European Workshop on Usable Security (Euro USEC'18)*, 2018.

[24] Shrirang Mare, Franziska Roesner, and Tadayoshi Kohno. Smart devices in airbnbs: Considering privacy and security for both guests and hosts. *Proceedings on Privacy Enhancing Technologies*, 2:436–458, 2020.

[25] Emily McReynolds, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. Toys that listen: A study of parents, children, and internet-connected toys. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI'17)*.

[26] Patricia A Norberg, Daniel R Horne, and David A Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of consumer affairs*, 41(1):100–126, 2007.

[27] Pixel 2 comes with a free google home mini, 2019. https://www.cnet.com/news/pixel-2-comes-with-a-free-google-home-mini/.

[28] Stefanie Pötzsch. Privacy awareness: A means to solve the privacy paradox? In *IFIP Summer School on the Future of Identity in the Information Society*.

[29] Yu Pu and Jens Grossklags. Valuating friends' privacy: Does anonymity of sharing personal data matter? In *13th Symposium on Usable Privacy and Security (SOUPS'17)*.

[30] Yu Pu and Jens Grossklags. Using conjoint analysis to investigate the value of interdependent privacy in social app adoption scenarios. In *36th International Conference on Information Systems (ICIS'15)*, 2015.

[31] Yu Pu and Jens Grossklags. Towards a model on the factors influencing social app users' valuation of interdependent privacy. *Proceedings on privacy enhancing technologies*, 2016(2):61–81, 2016.

[32] Daniel J Solove. The myth of the privacy paradox. *George Washington Law Review*, 89, 2020.

[33] Spotify premium subscribers can get a free google home mini, 2019. https://9to5google.com/2019/12/11/spotify-free-google-home-mini/.

[34] Smart home - united states - statista, 2020. https://www.statista.com/outlook/279/109/smart-home/united-states.

[35] Madiha Tabassum, Tomasz Kosinski, and Heather Richter Lipford. " i don't own the data": End user perceptions of smart home device data practices and risks. In *15th Symposium on Usable Privacy and Security (SOUPS'19)*, 2019.

[36] Janice Y Tsai, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti. The effect of online privacy information on purchasing behavior: An experimental study. *Information systems research*, 22(2):254–268, 2011.

[37] Plans announced to introduce new laws for internet connected devices, 2019. https://www.gov.uk/government/news/plans-announced-to-introduce-new-laws-for-internet-connected-devices.

[38] 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)*.

[39] Yaxing Yao, Justin Reed Basdeo, Oriana Rosata Mcdonough, and Yang Wang. Privacy perceptions and designs of bystanders in smart homes. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), November 2019.

[40] Eric Zeng, Shrirang Mare, and Franziska Roesner. End user security and privacy concerns with smart homes. In *13th Symposium on Usable Privacy and Security (SOUPS'17)*, pages 65–80, 2017.

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

[42] 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 *Mensch und Computer 2018-Workshopband*, 2018.

# A Appendix

## A.1 Survey Protocol

**[CONSENT FORM]**

**The Smart Home**

The concept of the smart home involves the control and automation of lighting, heating (such as smart thermostats), ventilation, air conditioning (HVAC), and security, as well as home appliances such as washer/dryers, ovens or refrigerators/freezers.

Wi-Fi is often used for remote monitoring and control. Home devices, when remotely monitored and controlled via the Internet, are an important constituent of the Internet of Things.

Modern systems generally consist of switches and sensors connected to a central hub sometimes called a "gateway" from which the system is controlled with a user interface that is interacted either with a wall-mounted terminal, mobile phone software, tablet computer or a web interface, often but not always via Internet cloud services.

- **Please select 3 images containing a smart home device [QUALIFICATION]**
  o Camera
  o Desk Lamp
  o Blender
  o Smart Thermostat
  o Voice Assistant
  o Smart Bulbs

- What factors do you consider when making decisions about adopting smart home devices? Please answer below.
- **Q1 For example, what are factors that could motivate you to purchase smart home products?** [text entry]
- **Q2 Similarly, what are factors that could keep you from adopting the technology?** [text entry]

The following questions are related to your preferences about collection of certain data by smart home devices. Please proceed when you are ready.

**[SCENARIOS 1-4]** [repeated four times, half of times purpose was omitted, two of the scenarios included a random device]

- The manufacturer/developer of your ["smart home device" or random device for scenarios 3 and 4] is accessing or inferring [random attribute], for example, [attribute explanation].

  They are using this information for [random purpose], for example, [purpose explanation]

- How do you feel about the data collection in the scenario described above if you were given no additional information about the scenario?
  o Very uncomfortable
  o Somewhat uncomfortable
  o Neither uncomfortable
  nor comfortable
  o Somewhat comfortable
  o Very comfortable

- If you had the choice, would you allow or deny this data collection?
  o Allow
  o Deny

- If you had the choice, when would you like to be notified about this data collection?
  o Never
  o Only the first time
  o Once in a while
  o Every time

- The manufacturer is sharing the data described in the scenario above with third parties (e.g., advertising companies, business affiliates). How do you feel about this?
  o Very uncomfortable
  o Somewhat uncomfortable
  o Neither uncomfortable nor comfortable
  o Somewhat comfortable
  o Very comfortable

- Finally, given the scenario described above, how do you feel about the government having access to this information?
  o Very uncomfortable
  o Somewhat uncomfortable
  o Neither uncomfortable
  nor comfortable
  o Somewhat comfortable
  o Very comfortable

- Please explain the rationale behind your answers [text entry]
- Was there anything unclear in this scenario? Is there a way we can improve the presentation of this scenario? (optional) [text entry]
- **[REVIEW SCENARIOS 1-4]**
- This was the scenario described earlier:

  The manufacturer/developer of your smart home device is accessing or inferring [attribute from respective scenario], for example, [attribute example].

  They are using this information for [purpose from respective scenario], for example, [purpose example]

  You indicated being [comfortable or uncomfortable] with this scenario.

- From the list below, please select the circumstances that could make you [more or less] comfortable. Please select up to three.
  o If the manufacturer was [well known or unknown]
  o If I [gave or did not give] consent to collect data
  o If information was collected [less or more] frequently o If the information involved was [not] sensitive
  o If I could [not] benefit from it (e.g., discounts, serendipitous opportunities)
  o If the information was stored for a [short or longer] period of time, ["then" or "or never"] deleted
  o If the information was [only used for or used beyond] the intended purpose
  o If I was [not] aware of how the data were being used
  o If the data collection was [not] useful for personal and home

safety
o If the data were [not] used for improving products and services
o If the data were [not] used for the common good (e.g., benefit the society at large)
o If I could [not] control the data (e.g., access, copy, and delete)
o If data [not] were handled and secured properly
o Other (please specify) [text entry]

- Please explain why you selected the circumstances above. [text entry]
- Was there anything unclear in this review stage? Is there a way we can improve the presentation of this review stage? (optional) [text entry]
- Please indicate your level of comfort in case your Identity (i.e., who you are) is included along with the data in this scenario. Your original level of comfort was [original comfort level for manufacturer]
  o Very uncomfortable
  o Somewhat uncomfortable
  o Neither uncomfortable nor comfortable
  o Somewhat comfortable
  o Very comfortable

- Please explain the rationale behind the answer above [text entry]
- **Q3 What privacy behaviors you would like to be able to adopt in the context of smart home devices? For example, would there be any privacy-protecting tools, configurations, and techniques you would like to use?** [text entry]

**[ECONOMICS-RELATED QUESTION]**[random assigment to one of four conditions]

Voice assistants take voice commands from users, enabling them to perform various tasks such as listen to music, control video/photo playbacks, and receive news updates. Voice assistants can also enable home automation, allowing users to control smart home appliances through voice commands.

Below is a picture of a popular voice assistant. [voice assistant photo]

In the next step, you will be given a scenario about voice assistants. In this scenario, "personal information" may involve data about your identity, lifestyle, habits, and personal background.

- Consider a scenario where you [are looking to purchase a voice assistant that costs OR had a voice assistant for which you paid] $49. The voice assistant [has OR has little to no] privacy controls and protections against collection and sharing of your personal information
- How much would you be willing to take as a discount off the price tag in exchange for allowing the manufacturer to collect and share personal information in the future? Please specify the amount in dollars [number entry] OR
- How much would you be willing to take as a refund in order to allow the manufacturer to collect and share your personal information? Please specify the amount in dollars [number entry]. OR
- How much would you be willing to pay as a one-time additional fee to add such privacy controls and protections? Please specify the amount in dollars [number entry]. OR
- How much would you be willing to pay extra in order to have more privacy controls and protections such as limited collection and sharing of your personal information? Please specify the amount in dollars (number entry). [number entry]
- Please explain why you would chose this amount. [text entry]

**[IUIPC QUESTIONNAIRE (7-POINT LIKERT GRID)]**

**[DEMOGRAPHICS]** We are almost done! Please answer the following questions regarding your demographics.

- What is your gender?
  o Male
  o Female
  o Or specify [text entry]

- What is your age?
  o 18-25
  o 26-35
  o 36-45
  o 46-55
  o 56-65
  o >65

- Do you live in the US?
  Yes
  No

- How many hours do you spend using the Internet every week? [slider entry from 0 to 168]

- Do you currently own a smart home device?
  o Yes
  o No

- How many smart home devices do you currently own? [if answer is yes to previous answer]
  o 1
  o 2
  o 3
  o 4
  o 5
  o 6
  o 7
  o 8
  o 9
  o 10 or more

- What types of smart home devices do you currently own? Please check all that apply [if answer is yes to owning device]
  o Security camera
  o Doorbell camera
  o Baby monitor
  o Pet technology
  o Motion sensor

o Smoke detector
o Leak sensor (water consumption)
o Smart lock
o Door/window alarm
o Garage door
o Smart lighting
o Switch/Plug
o Voice assistant
o Audio/speakers
o Thermostat
o Smart/automation hub
o Other (please specify)

- What is your occupation? [text entry]
- What is the highest level of school you have completed or the highest degree you have received?
  o Less than high school degree
  o High school graduate (high school diploma or equivalent including GED)
  o Some college but no degree
  o Associate degree in college (2-year)
  o Bachelor's degree in college (4-year)
  o Master's degree
  o Doctoral degree
  o Professional degree (JD, MD)

- Information about income is very important to understand. Would you please give your best guess? Please indicate the answer that includes your entire household income in (previous year) before taxes.
  o Less than $10,000
  o $10,000 to $19,999
  o $20,000 to $29,999
  o $30,000 to $39,999
  o $40,000 to $49,999
  o $50,000 to $59,999
  o $60,000 to $69,999
  o $70,000 to $79,999
  o $80,000 to $89,999
  o $90,000 to $99,999
  o $100,000 to $149,999
  o $150,000 or more

- What is the size of your household?
  o 1
  o 2
  o 3
  o 4
  o 5
  o 6
  o 7 or more

- Do you have children?
  o Yes
  o No

- Are you now married, widowed, divorced, separated or never married?
  o Married
  o Widowed
  o Divorced
  o Separated
  o Single

- Do you have any comments or suggestions for this survey? Thanks! [text entry]

## A.2 List of Codes

| Code | Motivator Example | % | Blocker Example | % |
|---|---|---|---|---|
| **convenience** | If I have a need for something to maximize my work output or it makes my life more convenient, I could look into smart devices. If a smart device can help me function more I would use it. | 41.20 | If the smart device is something that won't drastically improve my life then I would not buy it. | 2.55 |
| **privacy** | If it is non-intrusive | 4.12 | Mostly, companies obtaining information on my personal life. Just because consumers buy from a company doesn't mean the company can own them. | 36.61 |
| **price** | good prices | 26.94 | Probably the costs if they're far too high as well as complicated technology. | 41.36 |
| **security** | The security of the system and how protected it is from outside tampering. | 8.56 | The security of the item, could it be hacked? Could it have a camera that could turn on and be hacked? Would my personal information be safe? | 23.61 |
| **cost-saving** | The money I would save over time on utility bills | 20.76 | If it is not going to be saving in any large chunk of money in the long run. | 3.01 |
| **risk** | - | - | Safety of personal data and lack of assurance devices are not video taping and recording audio of my every move like I am on reality TV. | 13.47 |
| **reliability** | I want to make sure that there is a "dumb" fallback in case the cloud fails, and I want security. | | Single use products. Inexperienced brands. Behavior if they lose connection. | 16.80 |
| **control** | Freedom, meaning that I want these devices to control everything and give me the satisfaction that I want to have. | 3.33 | - | - |
| **remote control** | I'm most motivated by smart home products that can inform me of something in my home while I am away – those products that monitor my home | 10.30 | - | - |
| **fun** | Making life simpler. They're cool and cutting edge. They are fun. | 3.01 | - | - |
| **time-saving** | If it will save me a lot of time. If it can do things for me that I would have never imagined possible. If it reduces the amount of things I have to remember to do on my own. | 6.34 | Something that doesn't integrate well or reall doesn't save me either time or money | 0.32 |
| **functionality** | Some factors are their ability. I would like to have them be able to generally listen and complete my commands. | 7.61 | If the technology has a too many unnecessary features that make it an annoyance. | 2.54 |
| **ease of use** | Convenience and ease of use are the definite priorities in getting a smart home device. I want to make sure that it's easy for me to use and that it will save me time and money in my everyday life. | 28.37 | I would likely not purchase technology if it was exceedingly complicated or difficult to use. I also wouldn't purchase anything too expensive. | 19.65 |
| **need** | If they will fulfill a need that I am currently in need of. If it allows me to be more efficient with my time and money. | 10.94 | Doesn't do anything I need it to do, is too expensive, is not secure. | 9.67 |
| **novelty** | The novelty factor might motivate me to purchase a smart home device. | 2.69 | - | - |
| **nothing** | - | - | None | 2.54 |
| **online** | - | - | I do not like products that have no need to connect to a network...lights, Thermostats..etc. | 2.06 |
| **environ.-friendly** | For example the biggest factor would be conservation, trying to be efficient and making the best of resources by limiting my use as much as I can. | 2.06 | - | - |
| **reputation** | The company that maintains the data must be trustworthy. The mustn't have a poor reputation for cyber security. | 3.01 | The reviews being bad | 3.80 |
| **personalization** | Being able to program them to operate when I wanted them to, especially thermostatic products and appliances. | 1.58 | - | - |
| **safety** | Safety. I like the one that works for the oven. sometimes people forget to turn off the oven when going on a trip or think they forgot so it gives you the alleviation of knowing your house is safe. | 5.23 | Whether their use makes me more vulnerable to home invasion, unfavorable cost-to-savings ratio, whether their use compromises the security of my personal information... | 3.01 |
| **fear of tech** | - | - | Artificial intelligence becoming too smart, if power goes out or internet fails, technology dependant, paranoia or knowing people could hack me | 1.27 |
| **interoperability** | It has to be compatible with my smartphone | 9.51 | If it's not compatible with other devices. For example if I purchased an Apple product and I currently have devices that aren't compatible with Apple. Everything would need to work together. | 5.71 |
| **offline** | The product would need be able to be used locally and not rely on an internet connection. It would need to keep working if the company went out of business. | 0.79 | - | - |
| **aesthetics** | I would like the product to have a sleek design so that it looks nice in my home. I don't want the product to be an eyesore. | 2.85 | Lack of physical appeal, don't match my style or decor and too many steps to use | 2.22 |
| **durability** | The durability of the equipment. I would want it to last as long as possible. | 1.27 | Cheaply made, too complicated to use, not on the market long enough | 1.27 |
| **other** | Better broadband availability in my area. I can't do it before that happens. | 2.69 | How busy my life is going | 1.58 |

Table 1: List of codes along with examples of when they were considered a motivator or a blocker, ordered by absolute average values of each factor for all of the data when each factor is encoded as 1 for motivator, -1 for blocker, and 0 otherwise. "Other" was assigned to answers where a participant made a statement which we could not assign to any of the codes.

| Code | Example | % |
|---|---|---|
| **control** | I would like the choice to control what gets shared and why. And I don't want it to be underhanded. | 30.43 |
| **transparency** | I would like full reports on where my data is going from smart home devices sent daily or weekly | 20.76 |
| **access control** | More password, 2 factor configuration, anything that can make security better | 9.83 |
| **consent** | Letting the customer know ahead of time and asking for the customers consent and just keeping the customer involved as much as possible. i know for me it would make me feel better. Trustworthy companies, in my opinion will communicate with the customer on a regular basis and that in turn would make me feel the safest. i really cannot think of any of behaviors | 9.67 |
| **security** | I would most likely buy a firewall or some type of security to ensure my privacy and safety | 9.51 |
| **no data collection** | A legitimate way to block any data collection, though I doubt that would ever occur. | 8.56 |
| **no third parties** | I would like the smart home devices to work in my home without giving info to third parties. However this can be done should be done. | 7.29 |
| **deletion** | I'd like to be able to delete things regularly. Permanent deletion. | 7.29 |
| **identity protection** | If there was a way to de-identify the information by changing our voices or not attaching a location to the information, I would feel safer using these devices | 7.13 |
| **offline** | I would like for smart home tools to not transmit any user data out of the device. I would like them to be able to download new software but not to upload any of their own collected data. | 4.12 |
| **no sensitive data** | Anything that would block personal data from being shared. If I have to enter anything personal to use the equipment, I would want to be able to lock it and that it be never stored. And that I have the say so of what data about me I considers private and personal. | 3.65 |
| **guarantees** | Opt-outs, control of data, contractual obligations to not use my data | 3.01 |
| **encryption** | Yes, I'd like to have all of my data encrypted, scrambled and rendered useless by any third parties, and the data that is being used should only be used for my benefit and no one else's. | 2.69 |
| **not purchasing** | I wouldn't use smart home devices at all. | 2.06 |
| **no secondary use** | I would like to be sure I could deny or allow any collection or use of 3rd party data collection. No one needs to know that info. For any reason. Other than to enhance my smart home experience. | 1.9 |
| **personalization** | I would like to be able to run my own server for the devices to communicate with – if not all the time, at least to be able to do so as a backup in case the central server gets shut down. | 1.11 |
| **unlinkability** | Most of all, I do not want my smart devices linked to an already existing account, i.e., I want the smart devices to have a separate account from which I can control my devices, and those devices will not have access to personal or sensitive information. | 1.11 |
| **age restrictions** | Yes. There would be privacy locks for children, privacy feautures that only adults can access, and privacy features having to do with keeping the home more safe, and keeping our data secured. | 0.32 |
| **remote control** | I would like to be able to remotely check on my house without anyone having access to those data. | 0.32 |
| **other** | Smart home devices can make my life easier. They do not malfunction and can last a long time without being replaced in the future. As for configuration of the device, I would just make it easier for the user to have and practice with. | 2.69 |

Table 2: Codes ordered by frequency from privacy tools/behaviors question with percentage of respondents who mentioned them. "Other" was assigned to answers where a participant made a statement which we could not assign to any of the codes.