The Mirai in reviews: Examining customer reviews of Mirai susceptible IoT devices for Security and Privacy concerns

Swaathi Vetrivel[*], Veerle van Harten [†], Carlos Hernandez Ganan [‡], Michel van Eeten [§], and Simon Parkin [¶]

[1]Delft University of Technology

Abstract

Many consumer Internet-of-Things (IoT) devices lack sufficient security and privacy features that are fit for purpose. This points to a market failure for the security and privacy of consumer IoT. However, there is growing evidence that consumers care about having secure devices. This does not necessarily translate to informed decisions at the point of purchase. We investigate whether and to what extent customer reviews provide security and privacy-related information to consumers at the point of purchase of a new IoT device. We examine 79,052 reviews across four device types within a set of English-language Amazon retail websites. We perform topic modelling to group reviews, then conduct manual coding on these reviews to understand (i) the prevalence of security and privacy issues in consumer reviews, and (ii) the themes that these issues articulate. We find that there are signals in reviews, where around one in ten reviews we analysed includes mention of security and privacy issues. These included technical statements about features, frustrations with specific device use activities, as well as vignettes about trying to use a device in a particular context. We found alongside this that negative judgements on security and privacy were reflected in generally lower overall ratings for devices. Where there are a number of initiatives trying to improve consumer information on IoT security and privacy features, our results point to the value of the already existing mechanism of customer reviews. This suggest initiating complementary activities, such as review ratings focused on specific points of security and privacy interactions (e.g., device setup) and ensuring that there are defaults and shortcuts to meet the expectations of users with varying levels of proficiency. Future directions in this space point to finding ways to support buyers of devices to match available options to their personal expectations.

[*]s.vetrivel@tudelft.nl
[†]v.t.c.vanharten@tudelft.nl
[‡]c.hernandezganan@tudelft.nl
[§]m.j.g.vaneeten@tudelft.nl
[¶]s.e.parkin@tudelft.nl

1 Introduction

Among the range of consumer IoT devices now available – smart doorbells, smart home surveillance systems, etc. – many lack sufficient security and privacy features that are fit for purpose. These shortcomings have been exploited in various ways, most visibly in large-scale Distributed Denial of Service (DDoS) attacks from compromised IoT devices [1, 2]. These emerge as part a broader trend to leverage vulnerable IoT devices for malicious activities, ranging from botnets as criminal infrastructure [3] to cryptojacking [4] , and intimate domestic abuse [5, 6].

Where IoT devices do have security and privacy features, they can be difficult for consumers to use in a meaningful way [7]. The prior dependence on easily-guessed access passwords as default in many IoT devices is evidence of such a mismatch. The vigilance of home users has been put into question, but it is rather that the inherent design and functionality of consumer IoT devices make this an uphill struggle [8]. This all leads back to how the devices are made and features matched to user needs, these being issues which rest with the device manufacturer(s).

Such shortcomings can be seen as a market failure in consumer IoT security and privacy [9]. Historically, this has been attributed to a misalignment of incentives [10], seemingly rooted in consumers not appearing to want secure devices, or not objecting when provided with non-secure devices. Recent research shows instead that users not only care about IoT security but are willing to pay for it [11, 12] . The failure then is one of information asymmetry, a 'market for lemons' [13] where the consumer cannot discern good from bad with the information available to them.

A market correction is required to hasten improvements to the security and privacy ('S&P' from here onwards) conditions of smart homes and the IoT devices within them. There are various efforts, outside of device manufacturers themselves, to introduce a range of *market signals* [14] – more information for informed purchase decisions. This can include evidencing shortcomings in specific models of de-

vices [15], and independent assessment of devices (e.g., [16]), but these rely on consumers proactively seeking this additional information. Self-certified 'labels' are in formative and research stages, where these state assurances, etc. [17–19]. These would standardise the security and privacy information available to consumers, much like food nutrition labelling. However, these are not yet in use. Neither is it not assured that they would reflect actual real-world security device performance, or directly respond to the concerns that consumers have.

What is overlooked is an understanding of how much the consumer base is already recognising – and signalling – a need for security and privacy, and what those needs look like. This would constitute the strength of the existing market signal to others, be it prospective buyers of specific devices or manufacturers. If security issues raised by customers can be characterised and amplified, this would reduce information asymmetry for prospective device buyers, help differentiate secure from insecure devices, and potentially incentivize manufacturers through the impact on brand reputation [20].

Here we examine consumer reviews of smart home devices on a major online retail website. These are reviews written by customers, and so reflect their real concerns and experiences with IoT devices, which can include security and privacy issues. Many online retailers include consumer review content, and purchases are increasingly made online (a trend quickened by the lockdown conditions associated with the Covid-19 pandemic). We address two research questions: (**RQ1**) What fraction of customer reviews for IoT devices articulate security or privacy issues? And: (**RQ2**) When security issues or privacy are mentioned, what themes are being articulated?

We present the first large-scale study of security and privacy concerns in consumer reviews of IoT products. Via distinct sampling strategies, topic modelling and qualitative analysis approaches, we collected and analyzed 79,052 reviews from six country websites of Amazon for four IoT device type categories that have been commonly targeted and infected with the IoT malware because of poor security design choices [1, 15]. This whole scraping process from product link collection and subsequent review data collection after the cleanup was done between May and August 2021.

Our main contributions are as follows:

- We present the first comprehensive evaluation of consumer IoT 'excerpts' as represented in consumer reviews. Consumer reviews are an intrinsic part of purchase deliberation, and as such inform a critical intervention point for improved security purchase decisions. Further, these reviews are the unprompted views of consumers within a mix that includes other preferences.
- We describe a combination of the use of machine learning to categorise IoT product reviews, and manual thematic analysis of the text to understand context and themes.
- We show about 10% of the reviews contain security-related issues, which means that prospective buyers have a reasonable chance of encountering this information when considering various products. Security and privacy information is articulated both in technical and non-technical terms and spans a variety of themes, from firmware updates to worries about data capitalism.
- We discuss several recommendations for strengthening the information and signalling value of customer reviews and leverage this existing mechanism to reduce information asymmetry and the security incentives of manufacturers.

2 Related Work

Here we describe the existing research on user perceptions of IoT security and privacy (S&P), how this fits into a consumer/market context, and the landscape of attacks on consumer IoT devices motivating the urgency to understand and support the market for improved S&P provisioning in consumer IoT.

2.1 User perceptions of IoT security and privacy

Earlier work on user perceptions of IoT S&P has primarily been conducted through surveys, semi-structured interviews and experience sampling. To examine mental models of users of smart devices, Abdi et al. [21] and Zeng et al. [22] conducted semi-structured interviews and report gaps in user's mental models with respect to security. They attribute these gaps primarily to limited technical understanding and point to ad-hoc (and typically non-technical) strategies employed by users in order to protect themselves.

With respect to privacy, a study by Williams et al. [23] observes that more users were deterred by the price of consumer IoT devices than privacy concerns, and note that since purchase of IoT devices are voluntary, privacy was more likely to be sacrificed for functionality rather than necessity. This trade-off is echoed in other studies [24, 25] which observe that users balance the risks from using IoT devices against the convenience and benefits offered. In our analysis of reviews we find issues which essentially revolve around consumers 'not knowing what they were getting into' when purchasing a device, uncovering challenges in using and understanding their new device and how it fits into their smart home environment.

2.2 Assessing consumer perceptions

Prior interviews with retail customers have highlighted the point of purchase as a critical point for informing decisions

about the security of new computing devices [26]. In making decisions about security, home users get their security and privacy advice from various sources, where this can include family, friends, and peers [27, 28], informal technical experts [29, 30], and media such as news stories, blogs, and TV [31, 32].

Other studies have analysed the marketplace for IoT devices to explore methods and mediums to inform consumers of the privacy posture of devices and their corresponding consequences. For instance, Gopavaram et al. [33] investigated customers Willingness-To-Pay (WTP) for privacy vs their Willingness-To-Accept (WTA) a lack thereof through an emulated marketplace study. WTA participants, presented with the highest privacy settings by default, were more likely to pay a premium and purchase devices with higher a privacy rating, thereby indicating that interface design also influences purchase decision.

Blythe, Johnson and Manning [11] found that consumers are willing to pay more for increased security, and that the relative amount of risk reduction has no significant impact on that willingness. The authors also note that providing people with simple security related information prior to making their purchase decision has the potential to encourage purchase of devices which are more secure.

Online customer reviews have been analysed to gain novel insights into the privacy concerns of device owners. Reviews for Intelligent Personal Assistants (IPA) have been analysed through unsupervised machine learning techniques to understand whether consumers are concerned about their privacy when using these devices [34]. Similarly, reviews of children's toys have been used to understand associated safety concerns using text mining techniques [35].

Linden et al. [36] performed a comparative analysis of customer reviews of human and pet wearables and found that very few privacy concerns were expressed with respect to these technologies. Their thematic analysis also revealed that emotional drivers and desired functionality are prioritized over privacy requirements. We explore S&P-related issues raised by IoT device owners, which not only identify concerns but also 'hotspots' in device use where these issues become critical, and expectations around S&P which were not met. Our findings relate issues of awareness and preferences around S&P to the availability of information for an adequately informed purchase.

2.3 Interventions in consumer IoT purchase decisions

Various forms of S&P information label have been proposed, including a graded label, labels indicating S&P features, and labels indicating 'approval' by independent assessment [37]. Blythe et al. [37] found that except for cases where an information label indicated that a device has poor security, consumers were significantly more likely to buy a device with a label. They also found that although functionality was generally more valued, people were willing to pay the same premium for both improved functionality and security.

Emami-Naeini et al. [19] presented participants with labels which were a mix of the aforementioned label types, also reporting positive feedback, from participants who indicated that they struggled to find this kind of information at the time of device purchase. With a focus on the availability and regularity of security updates for IoT devices, Morgner et al. [17] found support among their survey participants for this kind of information, more so for those who perceived higher risks in using such devices.

Broader efforts to both improve and standardise IoT S&P features include nation-level codes of practice (including in the UK [38, 39] and US [40, 41]), where various challenges to such device standardisation efforts have been highlighted in research [42, 43], including agreement on standards and evidencing their effectiveness.

Many of these interventions seek to standardise various assurances, either from manufacturers themselves or from independent experts, that the S&P properties of a device are sufficient to be able to use a newly-purchased home IoT device securely. Here we explore the signals and indicators of S&P issues which emerge from owners themselves, as expressed in reviews provided in the setting of an online shopping platform. This accounts for the diversity of needs and differences in articulation and priorities, related to secure use of IoT devices.

3 Methodology

This section outlines our data collection and analytical approach. Our starting point for collecting the reviews is the online marketplace / shopping website Amazon. Amazon is a dominant e-commerce platform with a large customer base across many different countries. For consistency in the topic modelling and thematic analysis of reviews, we analysed English-language reviews from six Amazon sites: Amazon.com (United States), Amazon.com.au (Australia), Amazon.ca (Canada), Amazon.in (India), Amazon.sg (Singapore), and Amazon.co.uk (United Kingdom).

3.1 Selection of IoT devices

Research on IoT malware, botnets and compromised devices has found specific products that were being compromised at scale because of serious security failures, such as using known factory-default credentials. Many of these devices fall into four categories [15]: surveillance systems (including DVR/NVR), set-top boxes, smart home hubs, and routers. Although routers are often not considered as IoT, they are integral to home networks and also susceptible to IoT attacks.

We approached device selection in two ways. First, we searched for specific products known to be vulnerable to IoT

malware infections (specifically Mirai) [15]. This produces a set of devices with a high likelihood of prior, if not still current, security issues. We were interested to see if the reviews for these products would contain more comments on security or privacy (S&P) than other products for the same device type.

We searched for these once-vulnerable devices on the Amazon websites using a combination of manufacturer, model name/number and device type (e.g., XXX YYY-123 Router, device list added Appendix E) and collected the product links. Of the 53 IoT devices that we searched for, only 16 were still being sold: 14 routers, one DVR, one set-top box, and no smart home hub. The DVR and set-top box did not contain any reviews and so were dropped, resulting in a single category of 14 once-vulnerable routers.

As the second part of device selection, we then expanded our search to additional products in the same four product types (surveillance systems, set-top boxes, smart home hubs, and routers). Since these products do not fall under a single category on Amazon, we used different search terms. For each device type, we searched for a set of commonly used terms. For instance, surveillance systems might be referred to as surveillance camera, IP camera, security camera, etc. The terms used to search for each device type are listed to the Appendix in Table 2, grouped by device category.

3.2 Product page and review retrieval

A Python script was written to search for these terms, and collect all the product links on the first page of search results, along with the partial product title visible in the page in order to help with the manual cleanup in the next step. We repeated this for each of the six country websites. The search accounted for variations in the presentation of results (e.g., whether it is a list or grid of products, as presented on the search results page).

The script returned a total of 3,524 links across all six websites and four device types. The number of product links differs per category, because we used more queries for some categories than for others (as shown in Table 2). For example, for surveillance systems, we also included DVRs and NVRs and thus included queries for those. The different numbers of product links per category has no impact on our analysis, as we analyze each category separately to answer our research questions.

Before scraping the reviews for these products, we first reviewed all product links manually. More than 60% of them were dropped because they did not point to an IoT product. For instance, across IP cameras there were multiple results for fake cameras that merely act as a deterrent for burglars, as well as for cameras that do not connect to the internet. With a focus on internet-enabled devices, we excluded products which do not connect to the Internet or only use their own proprietary mesh network for connectivity. Devices with optional internet connectivity, like IP cameras that could be connected to the internet using a sim card, were kept in the set. Likewise, results for devices like baby monitors, spy cams, and pet cams that were internet enabled were retained. The final set consisted of 1415 product links. The count of product links collected per Amazon website and device type is Table 1.

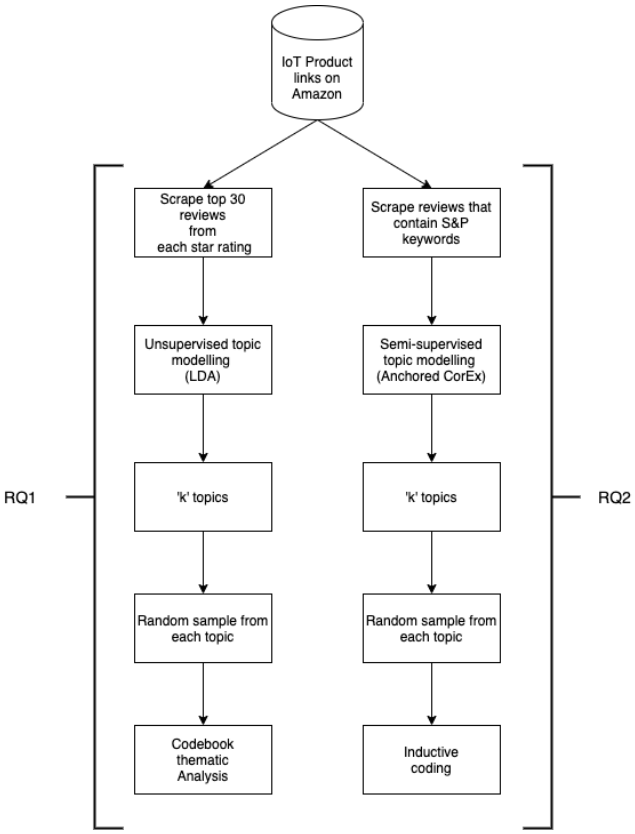

Figure 1: Overview of steps followed for each research question.

3.3 Review dataset construction

Using the product links we collected, we scraped the customer reviews for each product using Python scripts. For each product link, we collected two sets of reviews—one set for each research question (Figure 1).

Our first research question is 'What fraction of customer reviews for IoT devices articulate security or privacy issues?'. Answering this requires us to collect a dataset of reviews that is representative for what a prospective buyer would encounter when looking at product pages of a certain product type. For each product link, we scraped the first 30 reviews (three pages) for each of the five star ratings of the product. This resulted in 150 reviews for most products, though for some products there were less than 30 reviews with a certain star rating, leading to a slightly smaller set. By taking this approach, we mirror both a prospective buyer simply browsing reviews on the bottom of the page, and a buyer picking a star rating and looking at only

Table 1: Count of collected product links for each device type after cleanup.

Amazon website country	Surveillance systems	Routers	Set-top boxes	Smart home hubs	Once-vulnerable routers	Total
Australia	130	72	1	12	5	220
Canada	271	123	24	5	22	445
India	87	45	1	2	1	136
Singapore	76	42		5	3	124
UK	185	40	20	18	10	265
USA	132	50	8	17	9	207
Total	881	372	54	41	68	**1415**

Table 2: Search terms used for each of the four device types.

Surveillance Systems	**Smart home hub**
Surveillance camera	Smart home hub
Network Camera	Smart home control panel
IP Camera	Smart home automation system
Security Camera	
Dome Camera	
DVR/NVR	
Set-top box	**Router**
Digital set top box	Router
IP set top box	WiFi repeater

those reviews. Three pages is our estimate of the maximum number of pages of reviews that most buyers would look through when evaluating purchase options. By design, we collected reviews across different star ratings to avoid bias in the review collection and include reviews associated with different customer experiences.

When filtered by star rating, the reviews are sorted by default by 'Top reviews'. Top reviews are the ones that have been voted as helpful by other customers. However, in some cases Amazon shows most recent reviews, even when other top reviews are available. We collected the reviews in the order that Amazon provided them, since this would also be the order in which a prospective buyer would see them. Despite collecting reviews over all the star ratings, the final results displayed a slight skew towards five star rating.

Our second research question is 'When security issues or privacy are mentioned, what themes are being articulated?'. For this question, we wanted to focus in on reviews that explicitly mention security and privacy issues. For each product link, we used the 'search customer reviews' option provided by Amazon. The keywords used for searching were informed by user studies papers [21–25, 44]; this was part of an effort to use not only 'tech-savvy' words describing S&P, but to also account for the ways in which users articulate these concerns (e.g., a user may say 'setup' instead of 'configuration'). Even though terms like "get into", "always listening" and "big data" were used by users in those studies, we did not include them in our queries as Amazon does not support concatenated search terms. The list of keywords is presented in Appendix A, grouped by category. We checked the distribution of ratings for these search results. This is shown in figure 2. We can see that among the search results, five-star reviews form the largest category, but other rating levels are also present in the dataset, even though we collected the reviews in the way Amazon presented them in response to each query, just like a potential buyer would see them.

For all scraped reviews, we collected the title, content, date the review was posted, country it was posted from and the number of people that voted the review helpful. The username was not collected because of privacy considerations.

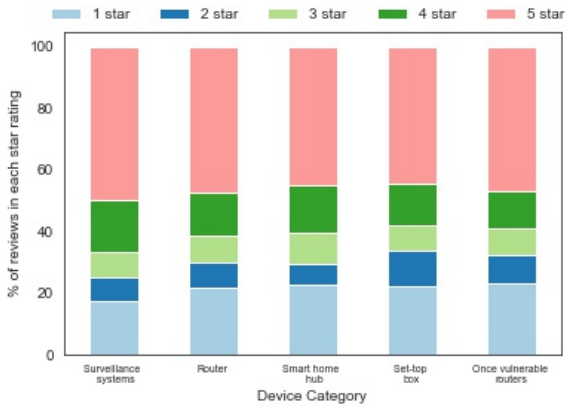

Figure 2: Distribution of ratings for reviews collected by searching for S&P keywords

3.4 Quantifying presence of IoT S&P issues

To answer our first research question, to what extent S&P issues were being discussed in reviews, we take a two-step

approach. We first conduct unsupervised topic modelling on the reviews in each device type category, in order to arrive at coherent clusters of reviews around certain topics. We then draw a random sample of 50 reviews from each topic and manually classify each review as to whether it discusses security or privacy, or not. This provides us with a quantification of what portion of the reviews for a certain device type mention issues of security or privacy.

We used Latent Dirichlet Allocation (LDA) topic modelling to discover 'topics' in the review dataset [45]. Stop words lacking semantic meaning, such as 'a' and 'is', were removed – here, this was expanded to include words such as 'buy' and 'find'. We then used a lemmatizer from the spacy library in Python[1] to remove inflections and return the dictionary form of words [46] to arrive at a standardised basic form, e.g., 'configuring' and 'configured' return 'configure'. In line with results of Martin and Johnson [47] we focused on nouns, based on speech tags (noun, adjective, verb, adverb) attached by the lemmatizer.

We estimated an LDA model for each of the five product types. Following the coherence metric, each model ended up with a specific number of topics. In total, the modelling identified 22 topics—12 for Surveillance Systems, 6 for routers and 2 each for the remaining device types. Next, to quantify the number of reviews that relate to security or privacy in each product category, we randomly sampled 50 reviews from each review topic, which amounted to a total of 1100 reviews.

Two researchers manually labelled the reviews according to whether they discussed security or privacy-related issues or not. This activity followed a thematic analysis approach [48], and more precisely, 'codebook' thematic analysis, to manually identify content which would fit under the categories [49]. Initial disagreements in the classification were resolved through discussion and clarification of what was within the scope of S&P. In addition to obvious statements about security or privacy of a product, reviews which referred to e.g., firmware updates and those with mentions of authentication during setup were also considered within scope.

3.5 Determining context of IoT S&P themes

After quantifying the presence of security and privacy (S&P) issues, we moved toward a more in-depth sense of the themes being articulated in these reviews. For this purpose, we collected 15,174 reviews via search queries with security and privacy-related keywords. We then ran topic modelling technique to identify clusters from which we could sample reviews for qualitative thematic analysis. For this, we chose semi-supervised Anchored Correlation Explanation (CorEx) topic modelling [50]. Unlike LDA, CorEx topic model makes few assumptions about the latent structure of the data, and flexibly incorporates domain knowledge through the anchor words that are fed to the model.

[1] https://spacy.io/api/lemmatizer

The highest coherence value was obtained for a 'k' value of 8, and we therefore ran the Anchored CorEx for eight topics. Since we had six categories of search words, we anchored each topic to two search words from each category that had the highest number of search results. This was done to guide the model towards these search groups while also leaving room for other related words to be picked up as part of the same topic. This allowed us to group the reviews into eight clusters based on the topics.

In the next step, a random sample of 100 reviews from each of these topic clusters was taken for thematic analysis. This sampling technique ensured that the samples taken for the thematic analysis are representative of each cluster. Moreover, thematic analysis allowed us to get a deeper understanding of the contexts in which S&P feature in customer reviews. The methodology outlined by Braun and Clarke [48] was followed for the thematic analysis for the context of S&P issues. Specifically, one coder analysed the dataset that was produced, performing inductive coding to identify themes emerging from the review content [49].

The Atlas.ti qualitative data analysis software was used, to code certain portions of text. Using this, the portions relevant for security and privacy in each review were assigned a code based on what it represented. These codes are included in Appendix D. Once this was done for all 800 reviews, in an iterative process, we grouped the codes into groups based on the underlying theme. These themes and the corresponding review counts are shown in Table 4.

3.6 Research ethics

The study design and data management protocol was approved by the Ethics Review Board of our institution. We evaluated our research design against the principles of the Menlo Report for ethical practices in computing studies [51]. Data collection was conducted on publicly available customer ratings and reviews on Amazon websites, and the associated usernames were not collected. Moreover, the scraping process was distributed over a longer duration through added delays in the script to ease the load on Amazon servers. Further, the scripts were fed specific pages and were not crawlers.

With respect to 'justice', our study design aims to contribute to reducing asymmetry for all consumers, not specific groups.

4 Results

Here we present the results of our combined topic modelling and thematic analysis activities. Reviewers are indicated by R###, and a device type classification indicated by an additional letter: Set-top Box (B), Router (R), Once-vulnerable Router (RO), Surveillance system (S), Home Hub (H). All quotes from reviews are included verbatim, including potential textual idiosyncrasies and errors.

4.1 S&P prevalence in reviews

To answer our first research question, on the fraction of security and privacy (S&P) issues in device reviews, we collected the top 30 reviews from each star rating for 1415 products (see Section 3.4). This set contained 68,684 reviews over the five device types. For each device type, we ran unsupervised topic modelling (LDA) on the reviews. This allows us to compare and contrast the results between device types. An overview of the 22 topics identified by the LDA models is provided in Appendix B.

We randomly sampled 50 reviews from each of these topics, so 1100 reviews in total, and manually classified them whether they mentioned security or privacy issues or not. By drawing the manual samples from the topics, rather than from the total set of reviews, we avoid certain larger topics from dominating the random sample. This way, we get a better sense of the diversity of reviews. This approach gets us a percentage per topic of the fraction of reviews that contain security and privacy issues. To get an overall percentage across topics, we calculate a weighted average that takes into account the size of each topic cluster in terms of the number

During classification of the reviews, we encountered mostly straightforward references to security or privacy properties of the devices. There were also borderline cases, such as customers mentioning the availability or lack of firmware updates to get new features for the device, rather than for security purposes. This is where a 'codebook'-oriented approach to analysis was utilised, adjusting the definitions of what was classified. To avoid undercounting the presence of relevant security and privacy information in the reviews, we also classified borderline cases as containing relevant security and privacy information.

Another area of borderline cases were the reviews that refer to the setup process. Many reviews comment on setup being easy or, difficult e.g., "*I haven't bought a router in a while to be honest but this was staggeringly easy to set up* (R20188-R)". We only classified as security-related those reviews that refer to security steps during the setup process, like login, password and authentication. These security actions might actually be evaluated negatively by the reviewer: "*dvr forces you to put a password we dont want passwords the software could have major revisions done to it to make easier to use* (R19314-S)".

In the end, classification resulted in varying numbers of reviews per topic sample referring to security and privacy. The highest references was within the topics for routers (15), while the only topic which didn't have any reference was from surveillance systems. On average, each sample had five reviews relating to security and privacy. This shows that security and privacy issues are not silo-ed in specific conversations and instead emerge in most contexts. We explore these contexts further in the next subsection.

The overall results of the classification for each device type category are shown in Table 3. Across all top reviews, about one in ten reviews (9.8%) articulate security and privacy-related issues. On the one hand, this is a minor fraction of all reviews. On the other hand, it does mean that potential buyers browsing reviews stand a decent chance of encountering comments on the security and privacy properties of the devices they are looking at. It also means that review writers feel these aspects are important enough to mention them one in ten times, which is a non-trivial amount given the brevity of most reviews: the average review in this dataset contains 100 words (median: 65). That is less than the length of this paragraph (125 words).

The fraction varies across device types: only 5.9% of the reviews for surveillance systems, while the fraction for router reviews is almost three times larger (16.3%). These differences cannot be explained from significant thematic differences across devices. The themes we will discuss in the next section are present among all device types. For some reason, routers trigger more reviewer comments on security and privacy than the other device types. This might reflect the awareness of reviewers of how crucial a router is to the overall security of the home network. Routers have also been targeted and compromised by IoT malware, but the same holds for surveillance systems. So it is not clear that the ongoing attacks explain the differences. In fact, the set of once-vulnerable routers, which have been compromised at scale, have a slightly lower prevalence of S&P (13.6%) than the general router category (16.3%). This difference is not statistically significant, which suggests that the known issues with these routers did not cause a substantial increase in security-related comments in those reviews.

Table 3: Percentage of reviews referring to security and privacy issues for each device type along with the average rating

Device type	**% of S&P reviews**	**S&P average rating**	**Non S&P average rating**
Surveillance systems	5.9	3.32	3.51
Routers	16.3	3.44	3.19
Hubs	8.7	2.38	3.49
Set-top boxes	6.4	3.33	3.24
Once vulnerable routers	13.6	3.33	3.27
Total	9.8	3.16	3.39

4.2 Inductive thematic analysis results

To answer the second research question, what themes are being articulated in reviews with security and privacy comments, 100 reviews were sampled at random from each of the eight

topics output by the semi-supervised algorithm (see Section 3.5). Thematic analysis was then conducted on these 800 reviews. The first step in thematic analysis involved in inductively defining and adding a single code to reviews based on their content. We defined 99 granular codes before reaching saturation. Of the 800 reviews that we analysed, 485 (60.6%) did not contain any references to security or privacy, even though they contained one of our search terms. This was an expected side-effect of using organic terms of real users, such as "record", "log", etc. This inevitably selects many reviews that are not related to security or privacy. These reviews did not receive a code. These were not considered for further analysis. In the final step, we condensed the large set of codes into a smaller set of themes, based on high-level commonalities among the codes. Appendix D contains the full list of codes and themes.

After this analysis, two additional properties of the reviews stood out that were orthogonal to the substantive themes. First, some reviews were written in quite technical language, referencing specific protocols or technical artefacts. For example: *"... NOTE: If you value your privacy you should put these cameras in their own vlan with NO outgoing access to any other vlans or network..."(R640-S)*. In some cases, reviewers using technical language mention being in, or having experience with, IT. Other reviews try to explain issues without using technical terms, as in: *...its IP address page.. has no option to log out from page...only one option is to close window and after again visit to its IP page.. it does not ask to fill password for log in.... and router page is opened automatically... anybody can change or do anything...only option is to clear history every time after log in...(R10280-R)*

The second distinction which we observed was whether reviews expressed personal frustration and friction with the *steps* involved in security configurations, e.g., *"It is app control. Every device [connect] need to open apps & new password setup which is so bother me.thanks"R11708-R*, or not, e.g., *"...It just prompts you to scan a code that allows your phone to download the app. Then scan code again, you are up and watching your cameras on the phone..."(R11708-R)* This distinction is about sentiment, as separate from whether the review has a positive or negative evaluation.

We decided to add two additional codes to all reviews, to complement the thematic code: whether they contained technical statements (yes or no) and to whether they expressed friction with the security features (yes or no). Interestingly, only 20% of the reviews with S&P issues express frictions with security steps, and none of these were written in technical language. This does not indicate that none of technical reviews experienced troubles with security features, because they do. Rather, they tend to articulate these problems not as personal frustration for not being able to achieve the desired result, but as specific technical critiques of the usability the devices – one of our thematic labels.

Table 4 presents the distribution of reviews across the seven substantive themes and across the two distinctions of friction and technical nature. It also include the average star rating of the reviews in each subset. The distribution of these two distinctions over the seven themes is shown in Figures 3 and 4.

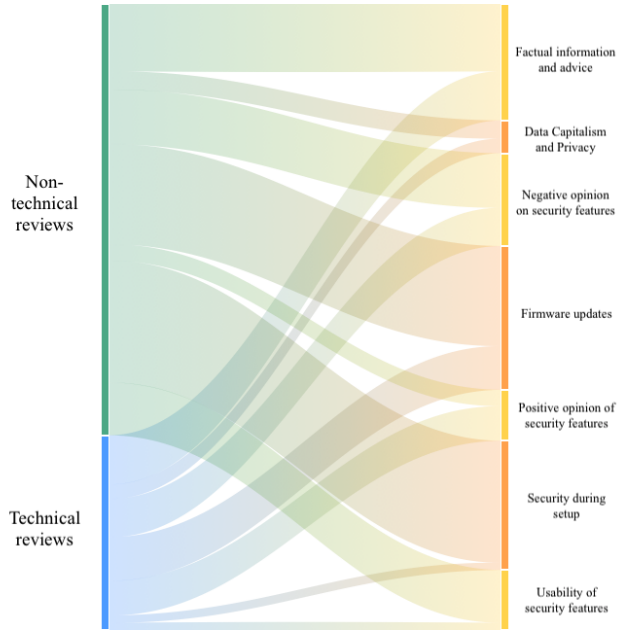

Figure 3: Sankey diagram depicting relation between technical and non-technical reviews over the other themes

Table 4: Distribution of reviews over themes, friction, technical nature, and security-related versus not security-related, as well as average review score per subset

Theme	Avg. rating	Count	%
Firmware updates	3.23	78	24.8
Security during setup	4.08	71	22.6
Factual information and advice	3.55	62	19.7
Negative opinion on security features	2.16	49	15.6
Usability of security features	2.68	33	10.5
Positive opinion of security features	4.75	27	8.6
Data capitalism and privacy	3.00	18	5.7
Friction with security steps	2.36	63	20.0
No friction with security steps	4.52	252	80.0
Technical reviews	3.52	102	33.7
Non-technical reviews	3.63	213	66.4
Security-related	3.41	315	39.4
Not security-related	3.56	485	60.6

4.2.1 Firmware updates

We now take a closer look at each of the substantive themes, starting with Firmware updates. These updates are the primary medium for installing security patches on devices. Of 315 reviews with S&P issues, 78 refer to firmware updates. Only

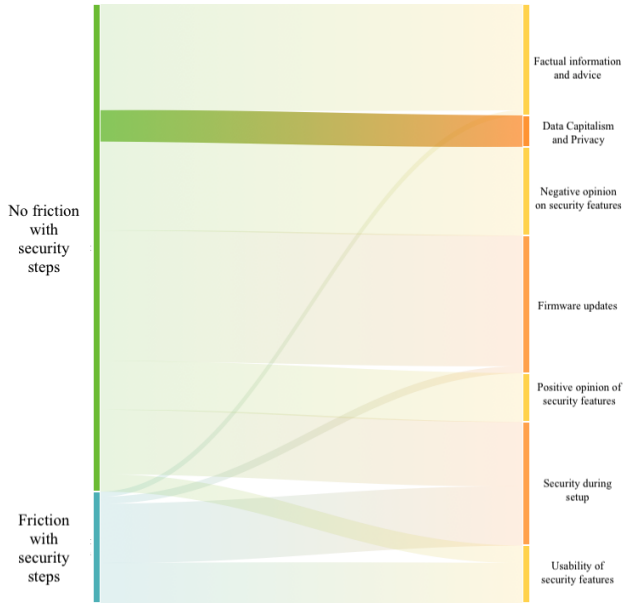

Figure 4: Sankey diagram depicting relation between reviews that express friction with security steps and those that don't over the other themes

14 of these talk about it positively. The rest are complaints about the update process.

As discussed in Section 4.1, we included all reviews discussing firmware given its crucial role in device security and its potential relevance to prospective buyers. Of the 78 reviews in this theme, 23 talk about firmware updates in relation to issues with the device—complaints about updates not solving an issue, causing it or not helping with it. A couple of reviews mention being annoyed with frequent updates "*too many interruptions for firmware patches (R11035-R)"* while on the other hand, a handful are appreciative of it "*Pros: incredibly fast, frequent updates (R8927-B)"*. There were also a few reviews about firmware being "*tooo buggy even after updating (R13798-R)*", firmware update process being *"unnecessarily kludgy" (R11624-R)* and firmware updates containing feature updates as well. Most of the reviews about issues during firmware update mention reaching out to customer service for support with the process, but aren't always able to resolve them:

> *"...After going to the [manufacturer] website, I found that the solution was to update the firmware. [...] somehow during the update it got stuck or corrupted and the power light just blinked red not stop. I sat on the phone for over an hour with support..." (R10576-R)*

In assessing the usability of the update process for itself for consumer IoT devices, Haney & Furman [52] found that participants experienced a lack of transparency in how updates worked and how. Where they noted a disconnect between updates and security, here we see a similar disconnect with firmware updates specifically, around expectations of their role in resolving issues with devices.

Only eight reviews discuss firmware updates explicitly in relation to security. Some mention auto-update feature as a security benefit (*"...Self updating. This system keeps itself up to date with the latest firmare, and software patches for stability and security..."(R13681-R)*), although one complained that *"...Autoupdate did not work..."(R12665-R)*. Some reviews mention particular security vulnerabilities that they would like a patch for. However, only in one case was the patch available:

> *"Was looking for a cheap router with updated Firmware available that included KRACK Patch ... I had to manually download and update the firmware..." (R11045-R)*

A similar desire for timely updates has been reported in a survey study elsewhere, focused on the information users would want to see on a product 'security label' [17].

4.2.2 Security during setup

Device setup is also a phase of key importance in the security of purchased devices. 21.5% (172) of the 800 reviews refer to setup. Interestingly, most reviews that refer to setup express polar opinions on the spectrum of it being very easy to frustratingly complicatedThis is more likely a reflection of user's expectations regarding the setup process than a direct indication of its difficulty, i.e., people experienced setting up as much simpler than anticipated or more difficult than expected.

Of the 172 reviews, 71 of these explicitly refer to security, e.g., *"enter a name for the SSID and create a password and that was it" (R11280-R)*. The rest do not refer to security steps like setting up username and passwords as an explicit step in setup, e.g., *"Easy to setup and configure"(R11107-R)*. Interestingly, almost all reviews that refer to security steps during setup are non-technical reviews (see Figure 4).

More than half of the reviews that refer to security discuss problems with passwords, including one review which mentions having written the password on a masking tape on top of the router. Of the rest, half a dozen were references to the relative ease of setting up using WPS and QRCode: *"Very easy setup using the WPS button on my router and the WPS button on the Extender" (R12684-R)*. Others experience frustration with the same *"The robot will not scan the qr code when attempting to pair" (R2976-S)*. Some of these reviews also mention returning devices because of frictions during the setup process *"Every time, I tried to set it up, it said that it had failed!!!This router is getting returned!!!" (R9692-R)*.

4.2.3 Factual information and advice

In total, 62 (19.68%) reviews provided (purportedly) factual information, sometimes coupled with security advice

to other users. Nearly half of these (43.5%) contain technical terms and details. Some of the technical reviews merely list the security protocols as part of the device specifications, e.g., *The Range Extender supports n/g/b wireless with WEP, WPA/WPA2-PSK encryptions (R9436-R)*. This does not articulate whether these features are good or bad, but may be useful for technically-literate consumers with matching expectations. Eight technical reviews draw some conclusion about the security of the device, but even these might be harder for a non-technical audience to grasp the implications *"The continuous video on the SD card is accessible via the app (and the app servers are in the cloud like everything else) but is supposed to be end-to-end encrypted."(R467-S)*

Nine reviews outline potential security issues in devices and contain technical advice on overcoming them. This is akin to 'informal technical support' normally provided by a 'local expert' to a device user [29, 30]. Four of these reviews ask users to isolate the devices on their network:

> *"...Based on what I saw in the software I would want this camera completely isolated from the world. Don't use their app, don't scan the QR code, don't let it phone home to the internet. Put it on a completely isolated network with your NVR equipment..." (R1654-S)*

Of the reviews that contain security advice in non-technical language, two advise users to change password of their devices *"I'm sure it's a default manufacturer password so I changed it, as I suggest anyone do" (R9144-R)*. Three warn users against buying products due to associated security and privacy issues.

> *"Several stories in the media about the poor security of [device name] devices. If you value your privacy and security, and would prefer your personal data and camera feed information not to be sold to the highest bidder please avoid these cameras..." (R7518-S)*

A few of the non-technical reviews make a generic comment that a device is *secure* without providing any details, like in this cause for a travel router:

> *"This is a great way to be more secure when using the Internet. Especially when traveling." (R12323-R)*

4.2.4 Negative opinion on security features

Unlike the previous theme where users commented and advised on security features, in this theme users express strong negative sentiment about the perceived lack of security of privacy. This negative sentiment is about the security of the device as a whole and not about the steps involved in configuration of security settings. The dissatisfaction includes general security concerns, notes about vulnerabilities, stories about devices being hacked, devices being flagged as non-secure, complaints about limited security and discomfort with providing customer service remote access. That is, these are limitations of the security of the device, which in general cannot be remedied through any amount of configuration effort – the specification is unsatisfactory.

The security concerns raised vary based on the device type and while some reviews raise these concerns in simple terms, others use more technical language. For instance, both of the reviews below express concern about access to the video feed of a security camera.

> *"... Cons: [...] didnt even see an option to change username which is a big negative as these days privacy is of utmost importance..." (R5398-S)*

The second review is technically more articulate about the negative evaluation:

> *"... The mobile device app [can't] access a camera off the local network. So if you segregate unclean IOT devices to their own subnet behind your firewall, then your mobile device will not be able to access the video feed.- Did I mention your username:password credentials are passed in plaintext as part of the URL when interacting with the camera?..." (R5998-S)*

Security concerns expressed about routers include lack of encryption, packet sniffing, lack of option to change the username 'admin', and vulnerabilities associated remotely configuration of router without an user account. Here as well, we observed a difference between technical and non-technical reviews, the former describe the issue, the latter merely state it:

> *"The description says "Advanced Security" but it doesn't have WPA3 available nor is this device compatible with WPA3, maybe 10yrs ago it was "Advanced"..... There is vulnerability in WPA2..." (R9257-R)*

> *"It has not increased the speed of signals.showing security problems" (R12360-R)*

Eight reviews provide accounts of devices the reviewer thought had been hacked; one was for a router and the rest were surveillance systems. For instance:

> *"... I logged in 1 morning (to the app) while hearing the clicks, and you can actually see a flash of light and a quick pause go off in the monitor with each click (as if someone is taking pictures).. it's as if this monitor's access is being live fed to perverted viewers who have access. Another thing I realized is that it only happened on the camera labeled "bedroom." (R1738-S)*

Of note here is that some reviews situate S&P issues, where such 'stories' about security or privacy experiences have been seen to resonate with technology users in similar contexts [28], where stories are typically of perceived security incidents.

4.2.5 Usability of security features

Most (88%) of the 33 reviews within this theme are non-technical and can be broadly classified into security features hindering usability. A clarifying example of the former is a reviewer who is annoyed with 2FA authentication since it interferes with functionality:

> *"Tonight the [device name] alarm went off whilst I was out. I accessed the app to see if I was about to be burgled or needed to speak to a visitor. And what did I get? "You need to setup 2 factor authentication before we will let you access your system"..." (R5901-S)*

Other reviews complain about the lack of support for password management software and talk about difficulty in entering strong passwords:

> *"...With the 4.0 line of firmware, stronger passwords are required. This is a shame in my opinion, and should maybe only be required to enable P2P or use the internet connected features. It's also a bit annoying that you can't plug a standard keyboard into the USB port for typing the password. Or even use a touch screen (see above!) Using the mouse is not a friendly way to create a strong password, especially the tiny mouse the unit comes with..." (R2487-S)*

On the other end, we find reviews that comment on usability that is not security-enhancing, e.g., on how easy it was to connect to a router since there was no password required. One review notes a design complaint:

> *..."I bought this router primarily for the WPA3 security. Problem is, when that's enabled, the onboard software disables the Wifi Protected Setup (WPS) button. (R9196-R)*

Another review complains about the lack of multi-user support:

> *"... I tried to set up my wife with the app to access the camera and all she can do is view the camera/s or make them record, nothing else, which is not acceptable at all, every one I give access to must have the ability to do everything I can do..." (R7471-S)*

4.2.6 Positive opinion about security features

The 28 reviews within this theme express a positive sentiment about the security features of the devices. Interestingly, the nine reviews in this theme that do not talk in technical terms offer details on why they like it—a handful mention encryption on the device while others provide more details:

> *"...I like the option of being able to share the video with other people this part I was worried about, I wanted it to be secure in that no one was able to view it without my permission. [...] you can share the cam by sending a request direct to the other person's email, it also shows you on the app who has permission to view it [...]" (R4991-S)*

Several reviews indicate that users trust 2FA to be more secure and safe: *"The phone app also has 2 factor authentication (YEAH!! All apps should!! don't let the hackers into your IoT because they stole or guessed your password!)" (R467-S)*. The other reviews are satisfied with the security of the device because they trust their own configuration rather than the device itself:

> *"...Personally I'm running super secure WiFi behind an awesome firewall and a VPN, there is noway some "hacker" is going to try that hard to get into this camera..." (R4409-S)*

The reviews for routers refer to the encryption settings, guest networks and built-in VPN, with one review mentioning that the built-in VPN was the reason for choosing a particular brand. In addition, a couple of reviews appreciated the DDoS and malware protection, firewall and networking monitoring tool that alerts them when a new device joins their network. Such reviews may indicate what options are available in the market, for prospective buyers to then challenge where they see such options *not* being offered.

4.2.7 Data capitalism and privacy

The 18 reviews in this theme are nearly evenly split between technical and non-technical reviews. Eight reviews express exhaustion at having to register with an account for usage and consent to user agreements:

> *"...the app will not work at all unless you sign up for an account with [manufacturer name]. Not only is this completely unnecessary to operate the router (My 5-year old [manufacturer name] has a great iPhone app and the only thing I have to sign into to use it is the router itself), but I'm completely fed up with this behavior from companies. [...] I'm tired of "agreeing" that companies can do basically whatever they want without any legal repercussions or responsibilities..." (R10200-R)*

Another review talks about an app for IP Camera asking for permission to access location and microphone of the users phone while another requests permissions to tweet, comment and (un)follow on Twitter when logged in through a Twitter account. A couple of reviews resent having to sign up with an e-mail id to manage a router and the underlying consensus across most of these reviews seems to be that their data is being 'sold' by the companies:

> *"...The app asks for extra permissions like location data and even body biometrics. They are obviously selling this data..." (R1043-S)*

Research looking specifically at smart speakers [21] has found that users have insufficient understanding of the data that is collected and processed by these IoT devices. In this context, this translates to reviews demonstrating concerns about having insufficient information about privacy-related matters to have made an informed purchase decision.

On the other hand, when their privacy is protected, especially in surveillance systems, some users note and appreciate it. Two reviews refer to a privacy mode in their camera positively:

> *"...What peaked my interest in this camera was the addition of face recognition, which means you are able to ignore family members if you want to. This was an important factor for me as the rest of the family were not too happy about being filmed all the time..." (R4721-S)*

However, one review mentions concern about how privacy is being handled:

> *"...Security feels quite questionable. I don't see any promise your video/pics/data/camera is safe and secure. I can put a PIN on the camera and that's good, but other than that, I haven't noticed any real mention of how they are protecting my privacy..." (R4323-S)*

Aside from privacy issues, these reviews generally indicate a weariness of having to create new, multiple accounts for a range of home IoT devices, rather than this being consolidated. This then has parallels to the history of password usability research.

5 Discussion

The overarching result for RQ1 is that across all of the reviews analysed, 9.8% of reviews refer to security and privacy. There is some variance across device types, from surveillance systems category 5.9% to routers with 16.3%. Interestingly, once-vulnerable routers have a slightly lower percentage (13.6%), so the security problems that have plagued these devices have not emerged in Amazon reviews.

Overall, it illustrates that there is a notable portion of reviews which discuss issues related to security and privacy. Further, the results from the thematic analysis for RQ2 show that customers talk about their negative personal experiences with these devices as evidenced by the reviews that discuss frustrations the setup, the hacking stories and exhaustion with what is perceived as unwarranted data collection from companies. Moreover, we see security advice and privacy concerns also being discussed.

Our results indicate that at least a small percentage of consumers in the market for IoT devices *do* care about security and privacy, and are able to articulate where these concerns arise in the lifecycle [53] of device ownership. There is then information in the market, within these reviews, that can be leveraged to ensure that other prospective buyers can make more informed decisions. This is to acknowledge that there is information asymmetry in consumer knowledge and how informed purchase decisions are; qualifying this, the asymmetry is experienced differently by different users, but information is available as part of the existing process of considering a purchase and points to approaches to amplifying this information.

Looking further at RQ2, we find that some reviewers express concerns which inform a negative view of the device overall - their personal preferences were not met. This represents *dissatisfaction* with a newly-purchased device. Others express a range of positive and negative sentiment toward the setup phase for a newly-purchased device, identifying a critical point in the ownership of a smart home device – preferences should be met here too.

The experiences of reviewers may point to a discounting of security and privacy risks in the market itself. It may be that a customer cannot return/exchange a device solely on the grounds of their S&P preferences not being met; we saw many reviews of owners being disappointed after purchase, when new information about a device's S&P properties becomes apparent to them, after some not-inconsequential period of attempted use; arguably those preferences are not being treated seriously at present, outside of regular no-questions-asked return policies. These misgivings are however also indicated in associated review ratings, where S&P issues in devices were reflected in the overall ratings. However, this points to a challenge for prospective buyers to *find* that information within reviews which may also discuss non-S&P issues and preferences.

A clear issue from a lot of the thematic analysis is that the customer did not get what they were expecting, or uncovered something that they had not thought about which then became a problem that they were stuck with. Any narrative of consumers choosing devices with inferior security must then also acknowledge those consumer who *know about* that inferior security, but do not know what to do with that new knowledge (e.g., newly-discovered vulnerabilities, failings or oversights in S&P configuration).

5.1 Limitations

As the Amazon website presents product reviews in a range of different ways, we designed our review sampling strategy to best account for a range of different review search strategies; this approach does not optimise for any one 'typical' approach for review presentation or search. In our analysis, we did, for instance, uncover specific S&P preferences which can aid in accounting for more directed review search strategies. It is acknowledged in security awareness research elsewhere that the potential for various forms of media to inform security behaviours is under-researched (including TV and news [32,54]); we then add product reviews to that list, where here we found potential to leverage this additional source of information.

Many of the reviews we analysed focused on the initial experience of device ownership. Where other research has only recently begun to map out extended periods of IoT device (e.g., [55], we found that reviews analysed here pinpointed specific activities where S&P issues were found, such as device setup.

5.2 Recommendations

Based on our analysis of customer reviews of home IoT devices, we provide the following initial recommendations:

- **Targeted S&P sub-category ratings.** What would seemingly be the most approachable change would be to have a sub-category of review score on shopping websites (and on display in physical stores), specifically for security and privacy. Amazon reviews already include sub-categories such as 'easy to use' and 'ergonomic', which appear to be tailored to the product type. We found that technical statements in reviews might only express sentiment and advice by association with a less-than-favourable review score. Reading such reviews may not adequately inform a less knowledgeable buyer, though the rating can serve as a shorthand indicator of device quality. More specifically, however, S&P-focused sub-categories could point to specific activities, such as 'ease of setup' and 'troubleshooting'.

- **Use the review system to match advice to emergent concerns.** Consumers may be willing to follow S&P advice, if it addresses their existing concerns. Our analysis provides evidence of strong themes in the concerns customers may have at the time of purchase, or with a recently-purchased device. On an online shopping platform such as Amazon, this could be addressed in the Q&A part of the product listing, to incorporate 'signals of interest' [14]. Buying a device then finding it unusable due to S&P concerns means the market does not know it is not being used – a further indication of 'interest' is then necessary. Where answers for security can be indicated as useful, it 'matches' a user to a solution that they believe restores the intention to use the product, beyond the signal signified by the initial act of purchasing the product. This also indicates specific issues to address in existing devices, rather than standardised expectations that completely new devices should meet (as indicated in formative security labels initiatives [17]).

- **Prioritise shortcuts in security design.** We noticed a distinction between reviews framed in technical details, and others not, with similar concerns, around shared device use activities (such as setup, adding a new user, etc.). There is then scope to balance the needs of users who want configuration options and novice users who want ease of setup. This relates to an established design principle of "Flexibility and efficiency of use" [56], which refers to shortcuts for skilled users. It may be that a device can be secured, but that not all owners experience use of those features in the same way.

6 Conclusions

We investigated to what extent customer reviews of IoT products provide security and privacy information to consumers at the point of purchase. Where there were security and privacy signals in reviews, these included technical statements about features, frustrations with specific device use activities, as well as vignettes about trying to use a device in a particular context. Negative views on IoT devices were reflected in generally lower overall ratings for devices. All in all, we find that customer reviews provide a valuable and widely-used mechanism for conveying S&P information to consumers—prior to, and complementary with, potential future labelling schemes for IoT.

Our findings indicate that tangible options for S&P may be of interest as much as the features which participants can 'imagine', allowing users to compare meaningful options and offerings to choose from what is available, rather than what is imaginable. This indicates that surveys of real device features in the market are useful. Future work will also include leveraging our manually-labelled reviews to train a classifier, to analyze a review dataset for S&P prevalence and themes.

References

[1] M. Antonakakis, T. April, M. Bailey, E. Bursztein, J. Cochran, Z. Durumeric, J. Alex Halderman, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, Y. Zhou, M. Antonakakis Tim April, M. Bernhard Elie Bursztein, J. J. Cochran Zakir Durumeric Alex Halderman Luca Invernizzi, M. Kallitsis, D. Kumar, C. Lever Zane Ma, J. Mason, and N. Sullivan Kurt Thomas, "Understanding the Mirai Botnet,"

USENIX Security '17, 2017. [Online]. Available: https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/antonakakis

[2] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "Ddos in the iot: Mirai and other botnets," *Computer*, vol. 50, no. 7, pp. 80–84, 2017.

[3] K. Thomas, D. Huang, D. Wang, E. Bursztein, C. Grier, T. J. Holt, C. Kruegel, D. McCoy, S. Savage, and G. Vigna, "Framing dependencies introduced by underground commoditization," *Workshop on the Economics of Information Security (WEIS)*, 2015.

[4] H. L. Bijmans, T. M. Booij, and C. Doerr, "Just the tip of the iceberg: Internet-scale exploitation of routers for cryptojacking," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 2019, pp. 449–464.

[5] N. Bowles, "Thermostats, locks and lights: Digital tools of domestic abuse," *The New York Times*, vol. 23, 2018.

[6] L. M. Tanczer, I. López-Neira, and S. Parkin, "'I feel like we're really behind the game': perspectives of the united kingdom's intimate partner violence support sector on the rise of technology-facilitated abuse," *Journal of Gender-Based Violence*, 2021.

[7] E. Zeng and F. Roesner, "Understanding and improving security and privacy in multi-user smart homes: a design exploration and in-home user study," in *28th USENIX Security Symposium (USENIX Security '19)*, 2019, pp. 159–176.

[8] B. Bouwmeester, E. Rodríguez, C. Gañán, M. van Eeten, and S. Parkin, ""The thing doesn't have a name": Learning from emergent real-world interventions in smart home security," in *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, 2021, pp. 493–512.

[9] I. Brass, L. Tanczer, M. Carr, and J. Blackstock, "Regulating iot: enabling or disabling the capacity of the Internet of Things?" *Risk & Regulation*, vol. 33, pp. 12–15, 2017.

[10] B. Schneier, ""regulation of the internet of things"," Schneier on Security (blog), 2017.

[11] J. M. Blythe, S. D. Johnson, and M. Manning, "What is security worth to consumers? investigating willingness to pay for secure Internet of Things devices," *Crime Science*, vol. 9, no. 1, pp. 1–9, 2020.

[12] S. Gopavaram, J. Dev, S. Das, and L. J. Camp, "Iot marketplace: Willingness-to-pay vs. willingness-to-accept," in *Proceedings of the 20th Annual Workshop on the Economics of Information Security (WEIS 2021)*, 2021.

[13] G. A. Akerlof, "The market for "lemons": Quality uncertainty and the market mechanism," in *Uncertainty in economics*. Elsevier, 1978, pp. 235–251.

[14] A. E. Roth, *Who gets what—and why: The new economics of matchmaking and market design*. Houghton Mifflin Harcourt, 2015.

[15] E. Rodríguez, A. Noroozian, M. van Eeten, and C. Gañán, "Superspreaders: Quantifying the role of iot manufacturers in device infections," in *20th Workshop on the Economics of Information Security (WEIS)*, 2021.

[16] Mozilla, "*Privacy not included," https://foundation.mozilla.org/en/privacynotincluded/, 2021, accessed: 2021-05-25.

[17] P. Morgner, C. Mai, N. Koschate-Fischer, F. Freiling, and Z. Benenson, "Security update labels: establishing economic incentives for security patching of IoT consumer products," in *2020 IEEE Symposium on Security and Privacy (S&P)*. IEEE, 2020, pp. 429–446.

[18] P. Emami-Naeini, Y. Agarwal, L. F. Cranor, and H. Hibshi, "Ask the experts: What should be on an IoT privacy and security label?" in *2020 IEEE Symposium on Security and Privacy (S&P)*. IEEE, 2020, pp. 447–464.

[19] P. Emami-Naeini, H. Dixon, Y. Agarwal, and L. F. 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*, 2019, pp. 1–12.

[20] D. K. Basdeo, K. G. Smith, C. M. Grimm, V. P. Rindova, and P. J. Derfus, "The impact of market actions on firm reputation," *Strategic Management Journal*, vol. 27, no. 12, pp. 1205–1219, 2006. [Online]. Available: http://www.jstor.org/stable/20142408

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

[22] E. Zeng, S. Mare, and F. Roesner, "End user security and privacy concerns with smart homes," in *thirteenth symposium on usable privacy and security (SOUPS 2017)*, 2017, pp. 65–80.

[23] M. Williams, J. R. Nurse, and S. Creese, "Privacy is the boring bit: user perceptions and behaviour in the Internet-of-Things," in *2017 15th Annual Conference on Privacy, Security and Trust (PST)*. IEEE, 2017, pp. 181–18 109.

[24] S. Zheng, N. Apthorpe, M. Chetty, and N. Feamster, "User perceptions of smart home IoT privacy," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, pp. 1–20, 2018.

[25] J. M. Haney, S. M. Furman, and Y. Acar, "Smart home security and privacy mitigations: Consumer perceptions, practices, and challenges," in *International Conference on Human-Computer Interaction*. Springer, 2020, pp. 393–411.

[26] S. Parkin, E. M. Redmiles, L. Coventry, and M. A. Sasse, "Security when it is welcome: Exploring device purchase as an opportune moment for security behavior change," in *Proceedings of the Workshop on Usable Security and Privacy (USEC'19)*. Internet Society, 2019.

[27] S. Das, L. A. Dabbish, and J. I. Hong, "A typology of perceived triggers for end-user security and privacy behaviors," in *Fifteenth Symposium on Usable Privacy and Security ({SOUPS} 2019)*, 2019, pp. 97–115.

[28] E. Rader, R. Wash, and B. Brooks, "Stories as informal lessons about security," in *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS)*, 2012, pp. 1–17.

[29] E. S. Poole, M. Chetty, T. Morgan, R. E. Grinter, and W. K. Edwards, "Computer help at home: methods and motivations for informal technical support," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 739–748.

[30] N. Nthala and I. Flechais, "Informal support networks: an investigation into home data security practices," in *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*, 2018, pp. 63–82.

[31] S. Das, J. Lo, L. Dabbish, and J. I. Hong, "Breaking! a typology of security and privacy news and how it's shared," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–12.

[32] E. Rader and R. Wash, "Identifying patterns in informal sources of security information," *Journal of Cybersecurity*, vol. 1, no. 1, pp. 121–144, 2015.

[33] S. R. Gopavaram, J. Dev, S. Das, and J. Camp, "Iot-marketplace: Informing purchase decisions with risk communication," 2019.

[34] L. Manikonda, A. Deotale, and S. Kambhampati, "What's up with privacy? user preferences and privacy concerns in intelligent personal assistants," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 2018, pp. 229–235.

[35] M. Winkler, A. S. Abrahams, R. Gruss, and J. P. Ehsani, "Toy safety surveillance from online reviews," *Decision support systems*, vol. 90, pp. 23–32, 2016.

[36] D. van der Linden, M. Edwards, I. Hadar, and A. Zamansky, "Pets without pets: on pet owners' under-estimation of privacy concerns in pet wearables." *Proc. Priv. Enhancing Technol.*, vol. 2020, no. 1, pp. 143–164, 2020.

[37] S. D. Johnson, J. M. Blythe, M. Manning, and G. T. Wong, "The impact of IoT security labelling on consumer product choice and willingness to pay," *PloS one*, vol. 15, no. 1, p. e0227800, 2020.

[38] UK Department for Digital, Culture, Media & Sport (DCMS), "Code of practice for consumer IoT security," https://www.gov.uk/government/publications/code-of-practice-for-consumer-iot-security, UK Department for Digital, Culture, Media & Sport (DCMS), 2018.

[39] ——, "Regulating consumer smart product cyber security - government response," https://www.gov.uk/government/publications/regulating-consumer-smart-product-cyber-security-government-response, UK Department for Digital, Culture, Media & Sport (DCMS), 2021.

[40] M. Fagan, M. Yang, A. Tan, L. Randolph, and K. Scarfone, "Security review of consumer home Internet of Things (IoT) products," https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8267-draft.pdf, US National Institute of Standards and Technology (NIST), 2019.

[41] K. Megas, B. Cuthill, and S. Gupta, "Establishing confidence in iot device security: How do we get there?(draft)," National Institute of Standards and Technology, Tech. Rep., 2021.

[42] S. L. Keoh, S. S. Kumar, and H. Tschofenig, "Securing the Internet of things: A standardization perspective," *IEEE Internet of things Journal*, vol. 1, no. 3, pp. 265–275, 2014.

[43] E. Leverett, R. Clayton, and R. Anderson, "Standardisation and certification of the 'Internet of Things'," in *Proceedings of the Workshop on Economics of Information Security (WEIS)*, vol. 2017, 2017.

[44] C. Geeng and F. Roesner, "Who's in control? interactions in multi-user smart homes," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.

[45] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[46] M. Sanderson, "Introduction to information retrieval," *Natural Language Engineering*, vol. 16, no. 1, pp. 100–103, 2010.

[47] F. Martin and M. Johnson, "More efficient topic modelling through a noun only approach," in *ALTA*, 2015.

[48] V. Clarke and V. Braun, *Thematic Analysis*. New York, NY: Springer New York, 2014, pp. 1947–1952. [Online]. Available: https://doi.org/10.1007/978-1-4614-5583-7_311

[49] V. Braun and V. Clarke, "One size fits all? what counts as quality practice in (reflexive) thematic analysis?" *Qualitative research in psychology*, pp. 1–25, 2020.

[50] R. J. Gallagher, K. Reing, D. Kale, and G. Ver Steeg, "Anchored correlation explanation: Topic modeling with minimal domain knowledge," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 529–542, 2017.

[51] E. Kenneally and D. Dittrich, "The Menlo report: Ethical principles guiding information and communication technology research," *Available at SSRN 2445102*, 2012.

[52] J. M. Haney and S. M. Furman, "Work in progress: Towards usable updates for smart home devices," in *International Workshop on Socio-Technical Aspects in Security and Trust*. Springer, 2020, pp. 107–117.

[53] D. Kotz and T. Peters, "Challenges to ensuring human safety throughout the life-cycle of smart environments," in *Proceedings of the 1st ACM Workshop on the Internet of Safe Things*, 2017, pp. 1–7.

[54] E. M. Redmiles, S. Kross, and M. L. Mazurek, "How I learned to be secure: a census-representative survey of security advice sources and behavior," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 666–677.

[55] G. Chalhoub, M. J. Kraemer, N. Nthala, and I. Flechais, ""it did not give me an option to decline": A longitudinal analysis of the user experience of security and privacy in smart home products," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–16.

[56] J. Nielsen, *Usability engineering*. Morgan Kaufmann, 1994.

A Search terms used for identifying S&P related customer reviews

Table 5: Security and privacy-related keywords used for searching customer reviews, grouped by category

Configuration and Authentication	**Access and Storage**	**Encryption and Security**	**Privacy**	**Attack**	**Patches and Updates**
setup	access	encrypt	privacy	hack	patch
configure	manipulate	encryption	private	attack	uptodate
profile	watch	ssl	personal	fraudsters	update
default	track	protocol	trust	spy	firmware
control	record	secure	open	steal	vulnerability
2FA	log			blackmail	risk
authentication	data			criminals	safe
password				cutoff	protect

B Results of LDA

LDA for Surveillance Systems

- **Topic 1**: night, vision, product, quality, picture, light, day, time, image, work
- **Topic 2**: app, phone, time, monitor, quality, video, home, view, picture, work
- **Topic 3**: system, quality, nvr, image, security, poe, setup, setting, video, picture
- **Topic 4**: network, device, connection, work, setup, router, app, internet, issue, access
- **Topic 5**: unit, battery, ring, model, number, year, doorbell, resolution, zone, month
- **Topic 6**: cloud, storage, subscription, video, window, stream, service, option, plan, year
- **Topic 7**: motion, detection, notification, alert, time, sensitivity, record, video, setting, alarm
- **Topic 8**: cam, brand, contact, color, noise, today, stuff, good, audio, condition
- **Topic 9**: cable, power, wire, ethernet, wall, box, plug, screw, plastic, plate
- **Topic 10**: card, video, sd, record, app, footage, recording, memory, playback, file
- **Topic 11**: software, car, pc, door, computer, web, hardware, people, foot, interface
- **Topic 12**: support, customer, service, issue, problem, email, help, tech, replacement, update

LDA for Routers

- **Topic 1**: extender, unit, range, room, instruction, wifi, work, light, install, plug
- **Topic 2**: network, setup, system, app, home, mesh, point, access, port, cable
- **Topic 3**: connection, performance, laptop, bit, set, video, quality, eero, phone, work
- **Topic 4**: speed, signal, house, internet, coverage, floor, strength, wifi, drop, test
- **Topic 5**: support, time, work, day, money, service, hour, tech, week, customer
- **Topic 6**: issue, price, month, review, year, band, problem, time, model, day

LDA for Hubs

- **Topic 1**: device, time, app, work, product, tv, control, button, setup, hub
- **Topic 2**: music, speaker, sound, play, love, sound_quality, room, alarm, quality, question

LDA for Set-top Boxes

- **Topic 1**: tv, box, fire, device, app, stick, work, product, control, issue
- **Topic 2**: channel, tv, time, record, program, device, unit, guide, cable, recording

LDA for once vulnerable routers

- **Topic 1**: network, setup, connection, speed, work, range, feature, access, signal, option
- **Topic 2**: issue, time, internet, connection, day, problem, cable, work, unit, support

C Results of Anchored CorEx

LDA for Surveillance Systems

- **Topic 1** (23.8%) : *anchor words: setup, configure, control*
  setup, password, easy setup, setup easy, easy, initial setup, configure, username password, username, camera setup
- **Topic 2** (19.8%) : *(anchor words: access, watch, record)*
  record, camera, motion, night, detection, quality, vision, night vision, motion detection, video
- **Topic 3** 11.1% : *(anchor words: encrypt, secure, protocol)*
  secure, protocol, encrypt, address, iris, blue iris, onvif, 168, 192 168, 192
- **Topic 4** 15.6% : *(anchor words: open, trust, personal)*
  open, trust, personal, open source, time open, open camera, personal data, camera open, open door, seal
- **Topic 5** 11.5% : *(anchor words: steal, hack, spy)*
  hack, steal, price, easily, attach, small, hole, recommend, outside, white
- **Topic 6** 8.9% : *(anchor words: protect, update, firmware)*
  update, firmware, firmware update, update firmware, latest, latest firmware, version, update review, upgrade, firmware upgrade
- **Topic 7** 6.5% : *no anchors*
  support, work, time, issue, review, try, problem, email, reset, contact
- **Topic 8** 2.8% : *no anchors*
  network, connect, setting, devices, cable, power, connection, feature, point, plug

D Codebook from Inductive Coding

- Firmware updates: fw_security_patch fw_auto_update exclusive_fw_update fw_available fw_buggy fw_comparision fw_language fw_latest fw_update_caused_issue fw_update_didn't_fix_issue fw_update_difficult fw_update_for_feature fw_update_frequent_good fw_update_had_feature fw_update_kludgy fw_update_might_cause_issue fw_update_might_fix fw_update_not_available fw_update_timing fw_update_to_solve_issue fw_update_took_longer fw_update_when_setting_up fw_updates_back_to_back_annoying
- Security during setup: setup_password setup_qrcode_easy setup_qrcode_strange setup_ssid_not_hidden setup_too_simple setup_wps setup_wps_easy
- Factual Information and advice: chinese_servers_distrust contacts_mfg_server fw_update_not_available_chinese modem_too_secure remote_access_cust_support rtsp_pw_plaintext security_advice security_comment security_protocol unconcerned_about_vulnerabilities used_device_pwd_already_set vulnerable_to_chinese_hackers woods_tablet
- Negative opinion on security features: 2fa_missing guest_network_doesn't_isolate has_security_vulnerability I_was_hacked_stories limited_security remote_access_uncomfortable security_concern unhappy_about_security unsecure_device_alert
- Usability of security features: 2FA_useless can't_change_password can't_setup_security conflict_with_wpa3_wps_settings encrypted_affects_playback encryption_program_marketing extra_security_on_trial locked_out_cos_password no_password_good password_not_available password_reset_thru_customer_support passwords_cumbersome security_options_at_launch share_admin_priveleges share_password_with_qr unhappy_about_update_for_privacy
- Positive opinion of security features: 2fa_better_security configured_for_security_so_not_worried ddos_protection_good firewall_configuration_not_needed firewall_good good_securitywise guest_network_for_security malicious_activity_monitoring_valuable notification_when_new_device_joins secure_video_sharing
- Data capitalism and privacy: app_permissions cloud_not_needed face_recognition_good_for_privacy not_always_listening_is_plus privacy_concern privacy_mode_in_camera

E List of once vulnerable devices we searched for

Table 6: List of once vulnerable IoT devices searched

Manufacturer	Device	Device Type	Manufacturer	Device	Device Type
AMIT	WIP-300	Router	DrayTek	Vigor 2862	Router
ASUS	RT-AC5300	Router	Dream Multimedia	Dreambox DVB Satellite	SetTopBox
ASUS	RT-N10U	Router	Flying Voice Technology	FWR9601 VoIP	Router
ASUS	RT-AC58U	Router	GNSS	Net-G5 GNSS	Receiver
ASUS	RT-N10.B1	Router	Grandstream	UCM6202 IP	PBX
ASUS	RT-AC54U	Router	Hichan Technology	WiDisk	Router
ASUS	RT-AC87U	Router	Hisilicon	Hi3798MV300	SetTopBox
ASUS	RT-N14U	Router	Huawei	HG659	Router
ASUS	RT-N13U.B1	Router	Interlogix	TruVision	NVR
ASUS	RT-G32	Router	Level One	WBR-6005	Router
ASUS	RT-N10	Router	Lifetrons	FG1060N	Router
ASUS	DSL-N10	Router	Linksys	Smart Wifi	Router
ASUS	WIRELESS-AC1200	Router	Linksys	LRT214	Router
AirTies	Air4920-2	SetTopBox	MAGINON	IPC-250HDC	IP Camera
AirTies	Air7120	SetTopBox	Rockchip	RK3228	TV Box
Amlogic	S905L	SetTopBox	Siera	Panther	DVR
Ceru Co	vu+solo2	SetTopBox	Sony	Ipela SNC-CH160	Security Camera
Ceru Co	Vu+ DVB	SetTopBox	Strong	1600	Extender
Cisco	Docsis	Gateway	Tecom	AH2322 ADSL	Router
Devolo	Microlink Dlan Wireless	Router	Ubiquiti	Aircube AC	Router
Digicom	RAW300L-A05	Router	Upvel	UR 313N4G	Router
DrayTek	Vigor 2860	Router	Upvel	UR-321BN	Router
DrayTek	Vigor 2925	Router	X10 Wireless Technology Inc	AirSight Xx34A	IP Camera
DrayTek	Vigor 2760	Router	ZTE	F620V2	Router
DrayTek	Vigor 2960	Router	Zhone Technologies	ZNID-GPON-2426A-NA	Router
DrayTek	Vigor 2926	Router	Zyxel	ADSL gateway	Router
DrayTek	Vigor 2133F	Router			