Are Consumers Willing to Pay for Security and Privacy of IoT Devices?

Pardis Emami-Naeini, *Duke University;* Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor, *Carnegie Mellon University*

https://www.usenix.org/conference/usenixsecurity23/presentation/emami-naeini

This paper is included in the Proceedings of the 32nd USENIX Security Symposium.

August 9–11, 2023 • Anaheim, CA, USA

978-1-939133-37-3

Open access to the Proceedings of the 32nd USENIX Security Symposium is sponsored by USENIX.

Are Consumers Willing to Pay for Security and Privacy of IoT Devices?

Pardis Emami-Naeini*, Janarth Dheenadhayalan†, Yuvraj Agarwal†, Lorrie Faith Cranor†

*Duke University †Carnegie Mellon University

Abstract

Internet of Things (IoT) device manufacturers provide little information to consumers about their security and data handling practices. Therefore, IoT consumers cannot make informed purchase choices around security and privacy. While prior research has found that consumers would likely consider security and privacy when purchasing IoT devices, past work lacks empirical evidence as to whether they would actually pay more to purchase devices with enhanced security and privacy. To fill this gap, we conducted a two-phase incentive-compatible online study with 180 Prolific participants. We measured the impact of five security and privacy factors (e.g., access control) on participants' purchase behaviors when presented individually or together on an IoT label. Participants were willing to pay a significant premium for devices with better security and privacy practices. The biggest price differential we found was for de-identified rather than identifiable cloud storage. Mainly due to its usability challenges, the least valuable improvement for participants was to have multi-factor authentication as opposed to passwords. Based on our findings, we provide recommendations on creating more effective IoT security and privacy labeling programs.

1 Introduction

Consumers bring home smart devices that have hidden privacy risks and security vulnerabilities. A scan of about 16 million home networks revealed that more than 40% of households worldwide have at least one vulnerable smart device [12]. These security flaws can be exploited, putting users' sensitive data and safety at great harm [10, 41, 44, 86]. At the same time, device manufacturers fail to disclose the sensing capabilities of their devices [64, 74] and sell data to third parties without users' knowledge and consent [58, 99].

Device manufacturers prioritize features that bring evidence-based market value, such as innovative device functionality, deferring security and privacy to the very last stages of product design [70, 84]. Given the current lack of readily available information about device security and privacy, consumers are unable to make purchase decisions based on security and privacy even if they want to [36]. Thus, it can be difficult to make the case that improved security and privacy would lead to improved sales or that consumers might pay a premium for devices with better security and privacy such that device manufacturers' investments towards enhancing such practices can potentially be justified. Empirical evidence that consumers informed about device security and privacy practices would value devices with improved security and privacy could help break this vicious cycle.

Prior research has shown that users are willing to pay more to have enhanced security and privacy, such as improved phishing detection [75], reduced risk of identity theft [83], better privacy when shopping from online vendors [93], and having privacy-enhancing features for social media [88]. In the context of IoT, past research focusing on *hypothetical* purchase decision-making has found that consumers are more willing to purchase and even pay more for devices with better security and privacy practices when that information is made salient [16, 35, 36, 49, 71].

Nevertheless, past efforts on identifying the monetary value of security and privacy for purchasers of IoT devices have two important limitations. The main limitation is their significant *hypothetical bias*, due to participants' lack of incentive to reveal their *true* willingness to pay (WTP). The hypothetical bias could then lead to an overestimation of the monetary value of security and privacy [19]. In our study, we instead used an *incentive-compatible* method to mitigate this prevalent bias. Secondly, prior research does not provide any information on the relative monetary value and significance of IoT security and privacy practices.

In a two-phase online study using a multiple price list (MPL) framework, we measured the monetary value of five security and privacy improvements for 180 Prolific [79] participants considering the purchase of a new IoT device. MPL is a valuation method where participants are asked to specify their preference toward purchase choices that are presented to them in a column of ordered prices [8, 11]. Due to the instrumented randomness in its compensation mechanism, MPL has been shown to be incentive-compatible, meaning

that it would be in participants' best interest to reveal their true WTP in order to maximize their benefits [6, 8, 9, 51].

We generated specifications for two smart devices: a smart speaker with voice assistant and a smart smoke detector. In the first phase, we presented each participant with five security and privacy improvements to one of the tested smart devices, asking them questions to identify how much they value *individual* improvements. In phase two, we explored the monetary value of security and privacy information when presented *together* on an IoT product label. We created three label types for the tested smart device—showing most protective, least protective, or no stated security and privacy attributes. We then presented participants with three MPL tables to elicit how much each label would impact their purchase decisions.

We built statistical models to quantify the value of security and privacy in participants' purchase behavior and used their open-ended responses to further investigate their purchase decision-making. Participants significantly valued almost all security and privacy improvements, both individually (phase one) and holistically (phase two). Based on the regression analysis, having de-identified rather than identified cloud storage was the most valuable improvement for participants. Due to its usability issues, the least valuable improvement was having multi-factor authentication instead of passwords. Additionally, we found that having a correct understanding of security and privacy information significantly impacts the level of risk participants associate with the smart device and the amount of premium they are willing to pay. Our phase-two findings also revealed that compared to a smart device with risky security and privacy practices presented on a label, participants were willing to pay significantly more to purchase a device with *no* information on its security and privacy practices. The open-ended responses show that without transparent disclosure, consumers assume that the device has similar security and privacy practices to other smart devices on the market.

We make the following contributions:

- Through an incentive-compatible methodology, we quantified the monetary value of five smart device security and privacy protections when presented individually and together on a label.
- Using qualitative analysis, we shed further light on how security and privacy protections impact participants' purchase behavior and attitudes.
- We provide actionable recommendations on informing consumers' purchase decision-making by effectively communicating the value of security and privacy.

2 Background and Related Work

We start this section by explaining willingness to pay and the use of multiple price list to calculate it. We then discuss research on the monetary value of security and privacy.

2.1 Measuring Willingness to Pay

The concept of willingness to pay (WTP) first appeared in the economics literature over a century ago [29]. WTP estimates a consumer's perceived value of a product or service by identifying the maximum price they are willing to pay [57, 65]. WTP has been used in many domains, including healthcare [4], food [63], and energy [14].

Methods to gauge consumers' willingness to pay [21, 53] can be categorized based on whether they use a direct or indirect measurement approach and whether they measure consumers' WTP in hypothetical or real contexts [18, 66]. Methods to estimate consumers' hypothetical WTP do so by directly asking participants (e.g., contingent valuation [1, 68]) or by analyzing their preferences to indirectly infer their WTP (e.g., conjoint analysis [42, 59, 101]). By asking participants to imagine making hypothetical purchase decisions, these methods capture stated willingness to pay. However, due to the use of hypothetical purchase scenarios, participants have little incentive to truthfully state their WTP [50, 60]. It has been shown that hypothetical WTP values overestimate the actual WTP [72]. Such disparity has been referred to as hypothetical bias [6, 61, 98]. Hence, these methods have low external validity and, thus, can generate inaccurate results [25, 45, 68, 77].

To avoid hypothetical bias and measure actual WTP, researchers use incentive-compatible methods to elicit participants' real monetary valuation of products with promises to sell the tested products to participants. In addition, incentive-compatible valuation methods communicate to participants that it is in their best interest to reveal their true preferences [6, 61]. Incentive-compatible approaches have high external validity and provide more accurate estimates of WTP as realistic purchase behaviors are observed or simulated in controlled experiments [100]. These methods could either directly ask participants for their WTP (e.g., Becker-DeGroot-Marschak mechanism [15]) or indirectly estimate their WTP using their responses (e.g., multiple price list [6, 8, 55]).

2.2 Multiple Price List

Multiple price list (MPL) is an indirect method to infer consumers' *actual* WTP, where participants are presented with a table in which each row shows a pair of products with corresponding prices. Each participant reviews the table, row by row, and indicates their preferred choice for each row. The experimenter then selects one row at random and implements the participant's purchase choice in that row. When using this method, participants engage with real purchase scenarios, where they are promised to receive a compensation at the end of the experiment [6, 11, 61]. From very early on, participants are instructed about how they will get compensated and why it is in their best interest to reveal their true preferences when answering questions. Specifically, the experiment instruction will explicitly mention that after participants specify their purchase preferences for all the MPL tables, one row among all tables will be *randomly* selected, and the partici-

pant's choice in that row will be used to compensate them. This method is shown to be incentive-compatible and less prone to hypothetical bias, as the promised randomly-selected compensation provides participants with incentives to reveal their true WTP [6, 8, 9, 17, 61].

It has been shown that it is relatively easy for participants to understand that it is in their best interest to truthfully disclose their WTP [8, 9, 11]. In addition, MPL poses a lower cognitive load compared to other incentive-compatible methods [7–9, 11, 22, 31], resulting in potentially more accurate estimation of WTP [8, 9]. We designed a controlled experiment and used the MPL procedure to elicit participants' actual WTP for improved IoT security and privacy practices.

2.3 Monetary Value of Security and Privacy

Using various hypothetical and real purchase settings, researchers have shown that some consumers are willing to pay extra to have better online privacy protections [2, 27, 39, 43, 52, 78, 88, 94]. Svirsky designed an experiment to capture participants' monetary valuation of their privacy. They did so by creating three experimental conditions to analyze how much participants are willing to pay to prevent their Facebook profile data from being shared with the survey taker in exchange for a higher bonus. In one of their conditions, they designed an elicitation treatment and created several MPL tables where participants were asked to review the rows of the tables and select between the proposed level of privacy protection or the amount of bonus, which were increased in equal increments. They concluded that respondents were willing to pay significant amounts to have their privacy protected if the improved privacy is clearly stated at the time of decision making [92].

In the context of IoT, Morgner et al. conducted a survey and used conjoint analysis to explore consumers' perceived relative importance of security update labels. Although their methodology did not allow them to specify consumers' WTP for security update labels, they hypothesized that participants' significant desire for having update labels could indicate their willingness to pay more for having such information at the point of sale [71]. Similarly, Blythe et al. [16] and Johnson et al. [54] conducted contingent valuation and concluded that consumers are willing to pay more for IoT devices with security labels. However, their methods did not allow them to conclude which security or privacy attributes are more valuable to consumers. In a small interview study, Emami-Naeini et al. reported that participants claimed to be willing to pay about 10% to 30% of the base price of a smart device to be provided with security and privacy information before making a purchase [36].

Emami-Naeini et al. designed and evaluated a layered label for consumer IoT devices comprising 47 security, privacy, and general factors [34], where the most critical factors were provided on the primary layer of the label. In their more recent work, Emami-Naeini et al. presented several hypothetical purchase scenarios to participants and quantified the impact of a subset of label security and privacy attributes on survey participants' risk perception and willingness to purchase [35].

Gopavaram et al. used an experimental IoT market to understand how much participants value their privacy when purchasing smart devices. Similarly to our work, they found that consumers are willing to pay to purchase devices with improved privacy. However, unlike our work, they considered privacy in a general sense and tested privacy ratings rather than specific privacy factors (e.g., purpose of data collection). In addition, their work focuses on eliciting the value of privacy and not security practices of smart devices [40].

Our study goes beyond the prior work by measuring the monetary value of IoT *security and privacy* via an *incentive-compatible* study design for increased ecological validity. Our study design enabled us to compare consumers' monetary values for several IoT privacy and security improvements.

3 Methods

We conducted a pre-study with 100 participants and a two-phase survey with 180 participants (130 in phase one and 50 in phase two) on the Prolific crowdsourcing platform. We recruited participants who were at least 18 years old, lived in the US, and had a minimum approval rate of 95%. We launched the pre-study survey and the first phase of the study in May 2021 and the second phase in August 2021. The study protocol was approved by our Institutional Review Board (IRB). We provide the complete list of pre-study, phase-one, and phase-two survey questions in Appendix A. When referring to each survey question, we mention the question number in parenthesis (e.g., PS1).

3.1 Pre-Study Survey (PS)

The goal of our pre-study was two-fold: to identify what smart devices to consider in the main study, and to determine their appropriate base prices. We designed four 5-minute pre-study surveys, each for a different device: two devices collecting sensitive data (smart speaker with voice assistant and smart security camera) and two collecting less sensitive data (smart smoke detector and motion detector). We recruited 110 Prolific participants for the pre-study. Each participant was randomly assigned to one of the four pre-study surveys for a specific smart device. We asked three attention-check questions and queried participants about devices they had purchased, price-points, and demographics. We excluded 10 participants who failed at least one attention-check question, resulting in 100 participants (25 participants per survey). We compensated each participant with US$1.

We prepared a table of general specifications for each of the four smart devices (see Appendix B) to mimic similar devices available on the website of Best Buy, a popular retailer in the US. For each device type (e.g., smart speaker), we reviewed the specifications of the available alternatives for that device type on Best Buy and created a table showing the features common to all devices of that type. For each feature (e.g., number of speakers), we selected the best option. For example, if the number of speakers ranged from 1 to 3, we chose 3.

In each pre-study survey, after obtaining participants' consent, we asked a few questions to understand participants' experience purchasing the smart device of interest (e.g., smart speaker with voice assistant) (PS1-2). We then presented them with a link to the device specification and asked them three attention-check questions about that specification (PS3-5).

Next, to specify the *optimal price point* for each device, we solicited participants' price sensitivity using the Van Westendorp's Price Sensitivity Meter [62, 97] (PS6-9). This method asks participants to provide four price points [56, 82], namely the price they perceive to be i) *too cheap* so that they would doubt the device quality, ii) *cheap* enough that they would find it to be a bargain, iii) *expensive* but would still consider purchasing it, and iv) *too expensive* to even consider buying it. We use these prices to calculate the optimal price point.[1]

Van Westendorp's Price Sensitivity Meter has an implied assumption that consumers have some knowledge about what the tested product is worth [24]. In the pre-screening surveys, we addressed this assumption by asking participants whether they have purchased the smart device before (PS1). We used the Wilcoxon-Mann-Whitney test to analyze whether those who had previously purchased the smart device had a different price perception for each of the four surveyed price points compared to those who had not previously purchased the device. The results showed that previous purchase did not have a statistically significant impact on participants' price perceptions. Therefore, we considered all 25 responses to each pre-study survey to calculate the optimal price point.

We calculated the optimal price points of $60.26 for the smart speaker, $60.91 for the smart smoke detector, $78 for the smart motion detector, and $94 for the smart security camera. Our goal was to select one smart device from each sensitivity category with similar perceived price points, with the lowest price to decrease the potential hypothetical bias [48, 87]. We chose the smart speaker and the smart smoke detector from the high and low sensitivity categories, respectively, for the main study and we set their prices to $60.

3.2 Main Study Design

We conducted a two-phase incentive-compatible study to assess the value of security and privacy information on IoT consumers' purchase behavior. In the first phase, we explored the impact of *individual* security or privacy attributes (e.g., receiving automatic security updates), while in the second phase, we studied the impact of five security and privacy attributes when presented *together*.

For our study to be incentive-compatible, we designed realistic purchase settings with real smart devices that come with security and privacy labels. However, existing devices do not come with such labels. To make our study as realistic as possible, we used carefully-designed deception elements and simulated real purchase scenarios to elicit participants' true WTP. We introduced our study as a market research study that we were conducting in collaboration with a major retailer (we did not mention any brand). In the initial part of phase-one and phase-two surveys, we asked a few screening questions to assess participants' eligibility. In the recruitment text on Prolific, we told participants that upon completing the screening survey, they will receive $1.50 and if they are eligible and participate in the future marketing research study, they will have the opportunity to get a discount coupon of $10 or more to purchase the soon-to-be-released model (Model X) of a particular type of smart device from a well-known manufacturer. We also told participants that this device would be selected based on their answers to the survey questions, and that we would process the discount coupon within 1 month of study completion. After they completed the screening questions, we used survey logic features to show the main survey questions only to eligible participants. Those who were ineligible only received the promised $1.50. Since the only promised compensation for the main survey questions was the discount coupon, we expected that only those who are in the market to purchase the indicated type of smart device would be incentivized to participate in the study, and that they would reveal their true purchase preferences to get compensated with a discount coupon for their desired smart device. On average, phases one and two took participants 16 and 15 minutes to complete, respectively. In both phases, we compensated each participant with a $10 bonus, directly through Prolific.

3.3 Ethical Considerations

We took several measures to limit the potential harm of the deception element of our study. Our institution's IRB permits using deception only if it is needed to achieve the main purpose of the study. Our goal was to simulate a realistic purchase setting for consumers to reveal their true WTP. Since current smart devices do not come with security/privacy labels, we needed to simulate realistic purchase settings and use deception so that participants would believe they would have an opportunity to actually make a purchase. To minimize potential harm to participants, we debriefed and compensated them after the completion of both phases of our study about the goal of our research and why we used deception (see online Appendix A). After we debriefed participants, no participant contacted us or the IRB to raise any concerns about the study design and compensation. However, several participants expressed interest when they learned about the goal of our study and asked whether we could share our findings upon publication.

4 Phase-One Study (PH1)

We designed a mixed between-subjects and within-subjects study to explore the amount of premium participants are willing to pay for each of the five types of enhanced security and privacy protections presented individually on an IoT label. The randomly-assigned between-subjects factor was the device type with two levels—smart speaker and smart smoke

[1] We used the package 'pricesensitivitymeter' in R to analyze participants' price preferences and calculate the optimal price points.

Attribute	Values			Comparison pairs	
	Low protection	Medium protection	High protection		
Access control	None	Password	Multi-factor authentication	None None Password	vs. Multi-factor authentication vs. Password vs. Multi-factor authentication
Cloud storage	Identifiable	De-identified	None	Identifiable Identifiable De-identified	vs. None vs. De-identified vs. None
Data sharing	Third parties	Manufacturer	None	Third parties Third parties Manufacturer	vs. None vs. Manufacturer vs. None
Purpose	Functionality & Monetization	Functionality & Personalization	Functionality	Functionality & Monetization Functionality & Monetization Functionality & Personalization	vs. Functionality vs. Functionality & Personalization vs. Functionality
Security update	None	Manual	Automatic	None None Manual	vs. Automatic vs. Manual vs. Automatic

Table 1: We considered five security and privacy attributes, each with three values. Comparing these values in pairs led to three distinct comparison pairs per attribute. We presented each participant with one randomly-selected comparison pair for each of the five attributes.

detector (see Section 3.1). Survey questions asked participants to specify their purchase preferences related to the soon-to-be-released Model X using the within-subjects factor of *security and privacy comparison pairs*, described next.

4.1 Study Design

Security and privacy comparison pairs. To mitigate survey fatigue, we aimed to keep the survey around 15 minutes, and our pilot surveys indicated that we could achieve this by asking about a maximum of five attribute-values. Our tested attribute-values were a subset of the most important factors from the primary layer of Emami-Naeini et al.'s IoT label [34]. We included both of the security attributes from the primary layer (`security updates`, `access control`). In addition, we included all the primary layer privacy attributes except for two (`device storage`, `data selling`), which are similar to two other included attributes (`cloud storage`, `data sharing`). For each attribute, we included three levels of protection: low, medium, and high (see Table 1). We selected the low and high protection levels as in Emami-Naeini et al.'s categorization [35]. For the medium level, among the values identified by Emami-Naeini et al. [33], we selected a value that is common for smart devices (e.g., `access control: password`) and/or poses privacy risks (e.g., `purpose: personalization`).

Based on the three selected values, for each attribute (e.g., `data sharing`), we constructed three comparison pairs (e.g., `third parties vs. none`). Each comparison pair indicates an enhancement of the security or privacy attribute, where the value on the left (e.g., `third parties`) represents a lower protection level than the one on the right (e.g., `none`).

Screening questions. We started the phase-one survey by asking screening questions. To keep participants from realizing which two smart devices we were specifically considering in our recruitment criteria (smart speaker and smart smoke detector), we diversified the device types in the screening questions by adding a third smart device (smart motion detector). We then asked participants about their intention to buy each of the three smart devices.

For each device type (presented in a random order), we first presented each participant with the link to the device specification (see Appendix B) and asked them three attention-check questions about it (PH1.1-3). We then asked them for their interest in a future marketing research study with a discount coupon of $10 or more to purchase that smart device (PH1.4). The possible answers were: very interested, moderately interested, slightly interested, and not at all interested. To mitigate potential selection bias [47], we implemented a logic such that only the participants who were at least moderately interested in *both the smart speaker and the smart smoke detector*, and had at least two correct attention-check responses for each tested smart device were invited to the phase-one survey. It took participants around 8 minutes to answer the screening questions. Those who were qualified to participate in the main phase-one survey had the option of either continuing to the main survey or leaving and receiving US$1.50 for their time.

Survey questions. After presenting the consent form, we asked participants questions about five randomly-selected comparison pairs, one per attribute (see Table 1). We presented each comparison pair individually in its own survey section in which we told participants that the smart device comes with a label that only mentions that specific pair. We then asked participants to explain what each side of the comparison pair means (PH1.5-6). Next we provided participants with our own definitions (see Appendix C), and asked them to specify on a five-point Likert scale how and why each comparison pair would impact their risk perception (PH1.7-10) and willingness to purchase the device (PH1.11-14).

We then used MPL to elicit participants' WTP. We provided participants with an introduction to MPL and some illustrative examples (see Figure 1) before asking them to fill out the table for our first comparison pair. Next we presented an MPL table to participants (PH1.15), where the attribute-value on the right was the one for which participants had indicated higher willingness to purchase than the one on the left. For example, Figure 2 shows the MPL table for the comparison pair `security updates: none vs. automatic`, where the participant was more willing to purchase the device with automatic updates than one with no updates.

Based on a small-scale interview study, Emami-Naeini et al.

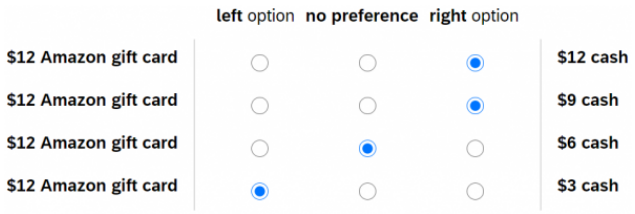

Figure 1: An example MPL table shown to participants. We explained that their compensation depends on how they fill out the table, i.e., one row of the table would be randomly selected, and the participant's preferred choice in that row would be implemented (breaking the tie randomly in case of "no preference"). For example, if the second row was selected, the participant would get $9 in cash, while if the third row was selected, the participant would get either a $12 Amazon gift card or $6 in cash, each with 50% probability.

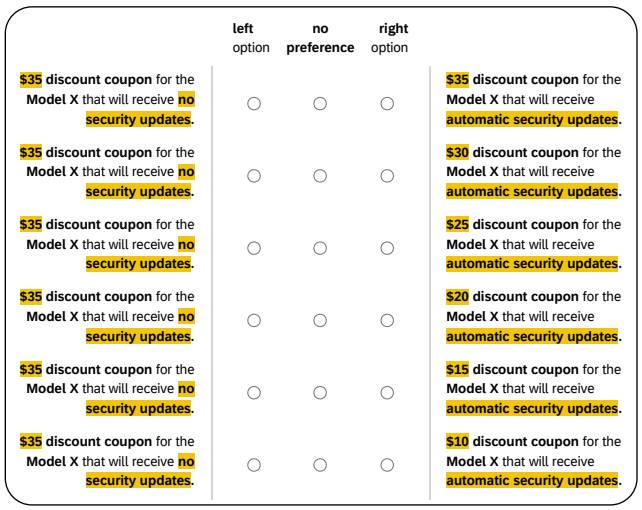

Figure 2: The MPL table for the comparison pair `security update: none vs. automatic` in the case that the participant's willingness to purchase is aligned with our hypothesis. In the reverse case, i.e., when the participant is more willing to purchase a device with no updates, the attribute-values on the left and right columns are swapped.

specified the amount of premium for IoT security and privacy to be 10%–30% of the base price of the device [36]. For the smart devices in our study with the base price of $60, this translates to premiums of $6–$18. To consider participants who did not value security and privacy for their smart devices, we symmetrically widened this range and set the lower limit of premiums to be $0 and the upper limit to be $25. [2]

Each table consisted of six rows (see Figure 2), where each row indicated the discount coupons for the attribute-values on the two sides. For the attribute-value on the left, the discount coupon was fixed at $35, while for the attribute-value on the right, it ranged from $35 (top row) to $10 (bottom row), in $5 decrements. We specified the maximum discount coupon based on the amount of premium (i.e., maximum discount = minimum discount of $10 + upper premium limit of $25).

For each table, we recorded each participant's choices as a sequence of 6 elements. Recall that the table is structured such that the device on the right is the one that the participant had

[2] We did not set the upper limit to $24 due to the $5 step size in our MPL tables.

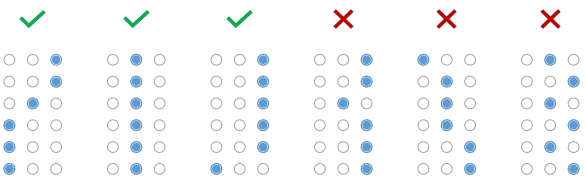

Figure 3: Examples of valid (three on the left) and invalid (three on the right) sequences of participants' choices in an MPL table.

specified a higher willingness to purchase. Therefore, a valid sequence of selected options should have had a general right-to-left pattern from top to bottom, with no switching back. Figure 3 provides examples of valid and invalid sequences. We found no invalid sequences among participants.

Finally, given each participant's selected choices, we determined the lowest and highest premiums the participant was willing to pay for the comparison pair. We found those limits using the switching point(s), i.e., the row(s) where the participant changed their preferences between the three options. The detailed procedure for deriving the premium limits can be found in Appendix E. We ended the phase-one survey with some demographic questions (PH1.16-PH1.20).

4.2 Data Analysis

4.2.1 Quantitative Data Analysis

We built three regression models to statistically describe participants' monetary valuation, willingness to purchase, and risk attitudes. The dependent variables (DVs) in the three models were as follows:

- Monetary valuation (ranging from -$25 to $25 for phase one and -$45 to $45 for phase two).
- Change in willingness to purchase (5 levels): 1 (strong decrease), 2 (slight decrease), 3 (no impact), 4 (slight increase), 5 (strong increase).
- Change in risk perception (5 levels): 1 (strong decrease), 2 (slight decrease), 3 (no impact), 4 (slight increase), 5 (strong increase).

We conducted model selection with backward elimination to find the best models that explain our DVs. Each model had four independent variables (IVs), including:

- `sp_comparison`: The security or privacy comparison pair with 15 levels (see Table 1).
- `device_type`: The type of smart device (smart speaker or smart smoke detector).
- `presented_order`: The ordinal categorical factor with 5 levels, denoting the order in which the comparison pairs were shown to a participant, e.g., first presented comparison pair, second presented comparison pair.
- `correct_definitions_frequency`: This factor has 3 levels and indicates participants' level of correctness when defining the attribute-values for that specific comparison pair. Level 0 denotes that none of the associated attribute-values were correctly defined, level 1 denotes one correct definition, and level 2 implies that all the attribute-values were correctly defined.

We used mixed-effects interval regression [91] to model participants' willingness to pay for improved security and privacy practices. Moreover, we used Cumulative Link Mixed Model (CLMM) regression [26] for the models to describe willingness to purchase and risk perception as their DVs were ordinal. The study had a repeated measures design. Therefore, for all models, we included random effects to count for within-participants data dependencies.

We used AIC to assess the model fit [20] and only included factors that helped improve the models. For all models, in addition to the IVs mentioned, we initially included demographic factors (e.g., age, income) and two-way interactions of the IVs (e.g., interaction of `sp_comparison` and `device_type`). However, these factors got removed in the model selection process as they did not improve model fit.

Mixed-Effects Interval Regression Model. Consider the i^{th} observation with participant p_i, whose willingness to pay lies in the interval $[L_i, U_i)$. The mixed-effects interval regression model first defines a linear predictor,

$$\eta_i = \alpha_0 + \mu_{p_i} + \sum_{f \in \mathcal{F}_i} \alpha_f, \tag{1}$$

where α_0 denotes the intercept, and μ_{p_i} denotes the random effect of participant p_i, modeled as a Gaussian random variable with zero mean and a variance σ_μ^2 determined by the model. Furthermore, α_f denotes the model coefficient for a given factor f, and $\mathcal{F}_i$ denotes the IV levels observed in the i^{th} observation in phase one, i.e.,

$$\mathcal{F}_i := \{\texttt{sp\_comparison}_i, \texttt{device\_type}_i, \texttt{presented\_order}_i, \texttt{correct\_definitions\_frequency}_i\}. \tag{2}$$

We then fit a Gaussian distribution to the WTP lower and upper limits in the i^{th} observation with η_i as the mean [91].

Cumulative Link Mixed Models (CLMMs). In what follows, we provide details on the model to describe participants' willingness to purchase. The description of the risk perception model is similar and is, therefore, omitted.

Consider the i^{th} observation with participant p_i, whose reported willingness to purchase is denoted by a discrete random variable $W_i \in \{1, \ldots, 5\}$. The probability that the willingness to purchase of the participant in this observation is at most $w \in \{1, \ldots, 4\}$ is modeled by the CLMM as

$$\Pr[W_i \leq w] = \sigma\left(\beta_{w|w+1} + \mu_{p_i} - \sum_{f \in \mathcal{F}_i} \gamma_f\right), \tag{3}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ denotes the sigmoid function, $\beta_{w|w+1}$ denotes the threshold parameter between the two consecutive response levels $(w, w+1)$, μ_{p_i} denotes the random effect corresponding to participant p_i, with a similar distribution as in (1), and γ_f denotes the CLMM coefficient corresponding to a given factor f.

Based on the cumulative willingness to purchase probabilities in (3), we denote the *odds ratio of increased willingness to purchase* for a given categorical factor f with respect to its baseline f_{baseline} by $\text{OR}^f_{\text{purchase},+}$. As we prove in Appendix F, this odds ratio can be written in closed form as

$$\text{OR}^f_{\text{purchase},+} = \exp(\gamma_f). \tag{4}$$

We also define the *odds ratio of decreased willingness to purchase* for a given factor f with respect to its baseline as

$$\text{OR}^f_{\text{purchase},-} := 1/\text{OR}^f_{\text{purchase},+} = \exp(-\gamma_f). \tag{5}$$

Baseline selection. In phase one, when we asked participants what "data to provide personalization" means, several participants (39/87) referred to it as a desirable feature of the smart device, due to its functionality and convenience. Given participants' desire to have this feature, we hypothesized that it would have a small impact on participants' risk perception. Hence, we selected `functionality & personalization vs. functionality` as the baseline for `sp_comparison`. For `device_type`, we selected the smart smoke detector as the baseline as it collects less sensitive data than a smart speaker [76, 103]. For `presented_order`, we selected 1st pair as the baseline. For `correct_definitions_frequency`, we selected `no correct definition` as the baseline as it was the first level of the factor. Note that in a regression model, the coefficient of a categorical factor should be interpreted compared to the baseline of that factor. Hence, any level of factors can be selected as the baseline, without any impact on how the model fits the data.

4.2.2 Qualitative Data Analysis

We conducted content analysis to qualitatively code participants' open-ended responses [85]. The first author created the codebook and kept it updated throughout the analysis. For each open-ended question, the first author and another researcher used the codebook to first code 10% of the responses. Then, the coders met to discuss and resolve their disagreements in the codebook. Afterwards, each coder independently coded the rest of the responses and held several discussion meetings with the other coder to go over the coded responses and resolve any remaining disagreements.

4.3 Findings

To have enough statistical power to construct our regression models (see Table 2), we needed at least 20 participant responses per security or privacy comparison pair for each device type. Since each participant answered questions about five comparison pairs, we needed at least 120 participants in total. 737 participants finished the screening survey, among which only 159 completed the main survey. Out of those, we excluded 29 participants who had at least one incorrect answer to the attention-check questions. Thus, we ended up with 130 participants. In our final dataset, for each smart device, each comparison pair was answered by 21–28 participants. The participant demographics are provided in Appendix D.

4.3.1 Label Security and Privacy Definitions

For each tested comparison pair (e.g., security update: none vs. automatic), we asked participants to define the lower (e.g., no security update) and higher protection (e.g., automatic

Row	Model Factor	Willingness to Pay (AIC = 3393.7)				Willingness to Purchase (AIC = 1614.3)					Risk Perception (AIC = 1600.3)				
		Estimate ($)	CI	SE	p-value	Estimate	SE	OR_+	OR_-	p-value	Estimate	SE	OR_+	OR_-	p-value
	`sp_comparison` (baseline = Purpose: Functionality & Personalization vs. Functionality)														
1	Cloud storage: Identifiable vs. De-identified	13.31	[8.29, 18.33]	2.56	***	1.40	0.42	4.05	0.25	***	−1.78	0.43	0.17	5.95	***
2	Access control: None vs. Password	12.74	[7.71, 17.78]	2.57	***	2.25	0.46	9.50	0.11	***	−2.38	0.46	0.09	10.83	***
3	Access control: None vs. Multi-factor	12.66	[7.75, 17.57]	2.51	***	1.68	0.43	5.37	0.19	***	−2.33	0.44	0.10	10.25	***
4	Security update: None vs. Manual	12.53	[7.59, 17.48]	2.52	***	1.72	0.44	5.61	0.18	***	−1.31	0.43	0.27	3.70	**
5	Security update: None vs. Automatic	12.26	[7.47, 17.04]	2.44	***	1.81	0.42	6.09	0.16	***	−2.14	0.44	0.12	8.50	***
6	Data sharing: Third parties vs. None	11.80	[6.83, 16.78]	2.54	***	1.86	0.45	6.42	0.00	***	−2.20	0.44	0.11	9.01	***
7	Purpose: Functionality & Monetization vs. Functionality	11.79	[6.83, 16.75]	2.53	***	1.96	0.45	7.09	0.14	***	−1.42	0.43	0.24	4.15	***
8	Data sharing: Third parties vs. Manufacturer	11.73	[6.80, 16.67]	2.52	***	0.72	0.41	2.05	0.49	0.08	−0.40	0.40	0.67	1.49	0.32
9	Purpose: Functionality & Monetization vs. Functionality & Personalization	9.48	[4.52, 14.44]	2.53	***	0.88	0.40	2.40	0.42	*	0.17	0.41	1.19	0.84	0.68
10	Data sharing: Manufacturer vs. None	6.99	[2.00, 11.98]	2.54	**	1.02	0.42	2.77	0.36	*	−1.73	0.42	0.18	5.63	***
11	Cloud storage: Identifiable vs. None	6.41	[1.45, 11.38]	2.53	*	0.44	0.42	1.56	0.64	0.29	−1.44	0.44	0.24	4.21	***
12	Security update: Manual vs. Automatic	5.75	[0.79, 10.71]	2.53	*	0.68	0.41	1.98	0.50	0.09	−1.01	0.42	0.36	2.75	*
13	Access control: Password vs. Multi-factor	−2.63	[−7.65, 2.40]	2.56	0.31	0.02	0.41	1.02	0.98	0.95	−1.54	0.43	0.21	4.68	***
14	Cloud storage: De-identified vs. None	−7.54	[−12.71, −2.38]	2.64	**	−0.86	0.41	0.42	2.37	*	−1.01	0.42	0.36	2.74	*
	`device_type` (baseline = smart smoke and carbon monoxide (CO) detector)														
15	Smart speaker with voice assistant	−1.35	[−3.82, 1.12]	1.36	0.29	−0.26	0.32	0.77	1.29	0.43	−0.03	0.20	0.97	1.03	0.88
	`correct_definitions_frequency` (baseline = No correct definition)														
16	One correct definition	4.18	[0.67, 7.69]	1.79	*	−0.12	0.31	0.89	1.13	0.70	0.79	0.35	2.21	0.45	*
17	All correct definitions	4.26	[0.82, 7.71]	1.76	*	0.00	0.19	1.00	1.00	0.98	0.91	0.34	2.50	0.40	**
	`presented_order` (baseline = First presented comparison pair)														
18	Second presented comparison pair	2.32	[−0.35, 5.00]	1.37	0.09	0.15	0.24	1.16	0.86	0.54	−0.05	0.24	0.95	1.05	0.83
19	Third presented comparison pair	1.92	[−0.75, 4.59]	0.36	0.16	0.24	0.24	1.27	0.79	0.31	0.01	0.24	1.01	0.99	0.95
20	Forth presented comparison pair	3.12	[0.43, 5.80]	1.37	*	0.47	0.24	1.60	0.63	0.05	−0.65	0.25	0.52	1.92	**
21	Fifth presented comparison pair	2.48	[−0.24, 5.14]	1.37	0.08	0.14	0.24	1.15	0.87	0.57	−0.47	0.25	0.63	1.60	0.06
	threshold coefficients														
22	1|2	-	-	-	-	−2.69	0.49	-	-	-	−1.11	0.48	-	-	-
23	2|3	-	-	-	-	−1.92	0.47	-	-	-	0.19	0.48	-	-	-
24	3|4	-	-	-	-	0.08	0.46	-	-	-	2.26	0.49	-	-	-
25	4|5	-	-	-	-	1.42	0.46	-	-	-	3.29	0.52	-	-	-
	intercept														
26	α_0	−0.22	[−0.24, 5.14]	2.72	0.94	-	-	-	-	-	-	-	-	-	-
	random effects														
27	σ^2_μ	27.62	-	-	-	0.36	-	-	-	-	0.56	-	-	-	-

Note: $*p < 0.05$ $**p < 0.01$ $***p < 0.001$

Table 2: Regression results of the first phase of our study, corresponding to the interval regression model to describe participants' WTP, as well as the CLMMs to describe participants' willingness to purchase and risk perception. Each row corresponds to a single factor, and shows the resulting model estimates, i.e., coefficients, for that factor, alongside the standard error (SE), and p-value for all the three models. For the willingness to pay model, we also include the confidence interval (CI) of the monetary estimate of each factor. Furthermore, for the willingness to purchase (resp., risk perception) model, we include the odds ratios of increased and decreased willingness to purchase (resp., risk perception) for all the factors, as defined in (6)-(5). The security and privacy comparison pairs (i.e., rows 1-14) are ranked in descending order according to the premium dollar amount participants were willing to pay for them, represented by the 'Estimate ($)' column under the willingness to pay model. Note that negative estimates in rows 13 and 14 of the willingness to pay model imply that participants were willing to pay more for the attribute-value that we hypothesized to be less protective (i.e., password access control in row 13 and de-identified cloud storage in row 14). We also include the AIC values for all the three models, which represent the models' goodness of fit.

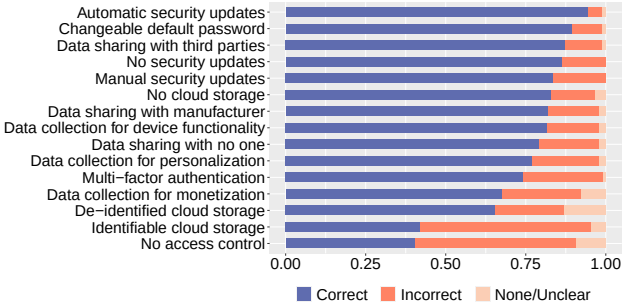

Figure 4: Distribution of the correctness of definitions provided by participants for each presented attribute-value. The y-axis shows the exact wording of the attribute-values we asked participants to define. After asking them with their own definitions, we provided all participants with the correct definitions.

security update) values (see Table 1) and then provided them with our own definitions (see Table 5 in Appendix C).

We qualitatively coded participants' definitions into three categories: correct, incorrect, and none/unclear. Figure 4 shows the fraction of responses in each category across all attribute-values, most of which (11/15) were correctly defined by at least 74% of participants. The definitions given for two attribute-values were mostly incorrect. The first one was having *no access control* with 60% incorrect responses. Open-ended responses indicated that for almost all participants who provided an incorrect answer, this attribute-value implied that the manufacturer had no control over the device (while in reality, it meant that the user has no control over who can access their device). P15 said:

> *It means that I completely control how the device works. The company has no autonomy over my settings.*

The second attribute-value was *cloud storage: identifiable* with 53% incorrect responses. To more than half of the participants (46/88), identifiable cloud storage implied accessible only by the primary user by providing identity information to log in. P79 reported:

> *This means I am the only one that could log-in with some personal credentials and retrieve the data the device has recorded/transmitted.*

Based on our regression analysis, the type of smart device or any demographic factors, including level of education, had no statistically significant impact on participants' number of correct definitions for security and privacy attribute-values.

4.3.2 Value of IoT Security and Privacy

We included five security and privacy attributes in our study: access control, cloud storage, data sharing, purpose, and security update. For each attribute, we considered three protection levels, leading to three comparison pairs (see Table 1). Each phase-one study participant was randomly assigned to one of the two smart devices, and presented with five randomly-selected comparison pairs, ensuring that exactly one pair from each of the five security and privacy attributes was shown.

We constructed three statistical models to describe the value of improved security and privacy on participants' willingness to pay, willingness to purchase, and risk perception. Table 2 shows the outputs of the three statistical models in phase one, where the comparison pairs are ranked according to the USD amount participants were willing to pay for them, denoted by the 'Estimate' under the willingness to pay model. In all models, a positive estimate of a factor indicates an increase in the dependent variable (DV) compared to the baseline of that factor, while a negative estimate shows a decrease in the DV. Because of implied enhanced security and privacy protection, we expected that each comparison pair would lower participants' perceived risk and increase their WTP and their willingness to purchase the smart device.

Cloud storage. We tested three comparison pairs: `cloud storage: identifiable vs. de-identified`, `cloud storage: identifiable vs. none`, and `cloud storage: de-identified vs. none`. Among those, we hypothesized that `cloud storage: identifiable vs. none` would most increase the level of protection and have the largest impact on increasing participants' WTP and willingness to purchase the smart device. Contrary to our hypothesis, participants found de-identified cloud storage to be more desirable than no cloud storage. In fact, they were significantly more willing to purchase (row 14, estimate $= -0.86$, p-value < 0.05) and pay for (row 14, estimate $= -\$7.54$, p-value < 0.01) smart devices with de-identified storage, mainly to be provided with the cloud functionality, although no cloud storage significantly lowered the perceived risk (row 14, estimate $= -1.01$, p-value < 0.05) as compared to de-identified storage. P106 stated:

> *A device with de-identified storage is much better and functional than one with no storage because it takes away personal identifiers with the added bonus of saving information away from the device.*

Among all security and privacy comparison pairs, participants were willing to pay the highest premium for having de-identified cloud storage as opposed to an identifiable one (row 1, estimate $= \$13.31$, p-value < 0.001). They were also significantly more willing to purchase a device (row 1, estimate $= 1.40$, p-value < 0.001) with de-identified cloud storage than one with identifiable storage. Moreover, model coefficients indicated that de-identified cloud storage significantly reduced (row 1, estimate $= -1.78$, p-value < 0.001) participants' risk perception compared to identifiable storage.

For all the presented security and privacy attribute-values, we used the same definitions that Emami-Naeini et al. provided in their IoT security and privacy label specification [33]. When defining the possible values (e.g., identifiable, de-identified) for cloud storage, they used a passive voice and did not explicitly specify whether the user or the service provider/-manufacturer would store data on the cloud. We hypothesized that depending on who would store data on the cloud, the value of having cloud storage might be perceived differently. To test our hypothesis, we looked into participants' definitions of de-identified, identifiable, and no cloud storage. For each definition, we then qualitatively coded it into three possible categories: 1) the definition does not explicitly mention who stores the data on the cloud (e.g., P59: "The information that is stored in the cloud is not linked to you or it does not include any personal identification information."), 2) whether it explicitly mentions that the user would store data on the cloud (e.g., P103: "This means that the consumer is not able to store the data on the cloud."), and 3) whether it explicitly mentions that the device manufacturer/service provider would store data on the cloud (e.g., P69: "The Model X manufacturer has some kind of cloud storage which is anonymous and is not linked to you."). All of the definitions fell into one of these categories. We then constructed a mixed-effects interval regression model with random intercept per participant to analyze whether the definition category as the independent variable has any impact on participants' WTP, willingness to purchase, and risk perception (each as a model dependent variable). Model coefficients indicated that the category did not have a statistically significant impact on the monetary value of security and privacy, as well as participants' risk perception and willingness to purchase the smart device.

Access control. As we expected, compared to having no control over access, participants were willing to pay a significantly higher amount to purchase a device that uses changeable default password (row 2, estimate $= \$12.74$, p-value < 0.001) or multi-factor authentication (MFA) (row 3, estimate $= \$12.66$, p-value < 0.001). In addition, both MFA and password significantly reduced participants' risk perception (`access control: none vs. MFA` (row 3, estimate $= -2.33$, p-value < 0.001), `access control: none vs. password` (row 2, estimate $= -2.38$, p-value < 0.001)) and increased their desire to purchase the device (`access control: none vs. MFA` (row 3, estimate $= 1.68$, p-value < 0.001), `access control: none vs. password` (row 2, estimate $= 2.25$, p-value < 0.001)).

Due to its enhanced security protection, we hypothesized that participants would be willing to pay more for a device with MFA than one with password. Our analysis, however, showed otherwise. The risk perception model coefficients indicated that as we expected, participants perceived MFA as a significantly more secure access control option than passwords (row 13, estimate $= -1.54$, p-value < 0.001). However, the risk reduction was not enough of a reason to increase our participants' willingness to purchase or willingness to pay for having MFA over passwords, primarily due to usability challenges associated with MFA. P27 stated:

> *I know that multi-factor is probably more secure, but being able to change the default password is easier.*

Security updates. As we hypothesized, all three comparison pairs for security updates implied enhanced protection and,

thus, significantly decreased participants' risk perception and increased the premium they were willing to pay.

The model coefficients showed that compared to having no security updates, participants had a significantly higher desire to purchase and were willing to pay significantly more to be provided with manual (row 4, willingness to purchase: estimate = 1.72, p-value < 0.001, WTP: estimate = \$12.53, p-value < 0.001) or automatic (row 5, willingness to purchase: estimate = 1.81, p-value < 0.001, WTP: estimate = \$12.26, p-value < 0.001) updates.

Based on the regression analysis, participants' WTP was significantly higher for having automatic security updates compared to manual updates (row 12, estimate = \$5.75, p-value < 0.05). Increased convenience was the main reason for participants' higher interest in receiving automatic security updates. 32% of participants, however, preferred manual updates and were willing to pay significantly more to have them compared to automatic updates. The two most commonly mentioned reasons for these participants were: 1) the desire to have control over updates, and 2) lack of trust in companies pushing the automatic security updates. Prior work shows that trust in the manufacturer's brand impacts consumers' purchase behavior and risk perception [5, 102]. P43 attributed their interest in having manual security updates to the increase autonomy over the update process:

> *I prefer not to be dictated to. I purchased the item, so it should be my choice. Instead the updates should highlight the benefits to propel to install them.*

P123 expressed distrust in device manufacturers issuing security updates and said:

> *I believe automatic security updates can lead to higher risk because nowadays, you don't know what big tech is doing with their devices which can result in intrusion of privacy.*

Data sharing. Most participants reported being significantly concerned when their data was being shared with anyone (row 6, estimate = -2.2, p-value < 0.001 and row 10, estimate = -1.73, p-value < 0.001) and therefore were significantly more willing to purchase (row 6, estimate = 1.86, p-value < 0.001 and row 10, estimate = 1.02, p-value < 0.05) and pay (row 6, estimate = \$11.8, p-value < 0.001 and row 10, estimate = \$6.99, p-value < 0.01) when having the assurance that no one will have access to the collected data. However, when offered two alternatives—data shared with third parties vs. the device manufacturer—despite being concerned about both, participants were willing to pay significantly more (row 8, estimate = \$11.73, p-value < 0.001) for a device that does not share their data with third parties, even with their lack of trust in manufacturers. P62 mentioned:

> *I don't know how much faith I have in the device companies, but sharing with the company only would be better than the data being released possibly to multiple parties.*

Most participants (31/45) reported that their main concern with third-party data sharing is the lack of transparency and control over the third parties the data is shared with and being uninformed about their security and privacy practices. P14 stated:

> *I don't trust what third parties will do with my information. They may use that information to try to market their products to me.*

Purpose of data collection. We considered three values for the purpose of data collection: providing device functionality only, providing device functionality and personalization, and providing device functionality and monetization. To be more realistic, we included the purpose of providing main device functionality in all the three values as we expected this purpose to be common across smart devices.

Participants were strongly against data monetization. Regression coefficients indicated that compared to monetization, participants had a significantly higher desire to purchase and were willing to pay significantly more to have their data used only to provide and improve main device functionality (row 7, willingness to purchase: estimate = 1.96, p-value < 0.001, WTP: estimate = \$11.79, p-value < 0.001) or for personalization (row 9, willingness to purchase: estimate = 0.88, p-value < 0.05, WTP: estimate = \$9.48, p-value < 0.001).

We expected participants to pay more for data not being used for personalization. However, many participants reported that they would like to have personalization and, contrary to our hypothesis, were willing to pay more to be provided with it. P98 mentioned:

> *If it is going to collect data anyway, I would prefer to pay more and purchase one that also uses it to better provide relevant and personalized service.*

Nevertheless, a few participants (5/43) expressed concern about personalization. They reported that for accurate personalization, the device needs detailed user data and that would increase its associated risks. P64 stated:

> *I like the devices with personalization as they are more useful, but I also know that there is an increased risk associated with personalization as the device needs to obtain more information about the user specifically.*

Other model factors. Based on the regression analysis, the type of smart device (row 15) did not have a significant impact on participants' purchase behavior. However, having more accurate definitions (rows 16-17) for the attribute-values of the comparison pairs led to a significant decrease in risk perception and increase in the amount of premium participants were willing to pay. In our models, we controlled for the presented order of the comparison pairs (rows 18-21). In general, the risk gradually decreased and willingness to pay increased as participants were presented with more scenarios. The decrease in risk perception and increase in willingness to pay were most significant when assessing the fourth scenario

Label Type	Attribute				
	Access control	**Cloud storage**	**Data sharing**	**Purpose**	**Security update**
Most Protective	Multi-factor authentication	De-identified	None	Functionality	Automatic
Least Protective	None	Identifiable	Third parties	Functionality & Monetization	None
No Information	Not disclosed	Not disclosed	Not disclosed	Not disclosed	Not disclosed

Table 3: Based on the phase-one study, for the second phase, we considered three types of security and privacy label to communicate the most protective information (most protective label), least protective information (least protective label), and no information (no information label) to participants.

(row 20).

Our qualitative analysis indicated that when assessing the value of *privacy* improvements (`cloud storage`, `data sharing`, and `purpose`), the majority of participants did not explicitly mention the IoT device in their responses, suggesting that their valuation could be general regardless of the device (e.g., mobile phone) or platform (e.g., social media). On the other hand, most participants mentioned the smart device in question when assessing the value of *security* attributes (`security update`, `access control`). For example, when comparing manual security updates vs. automatic updates, several participants reported that they are concerned about the functionality of their smart home devices and prefer to manually update them not to encounter unintended device behavior instead of them being automatically updated. In addition, when the device was smart speaker, most participants attributed their high monetary value for having access control to the shared usage of the device in their household.

5 Phase-Two Study (PH2)

In the phase-one study, we examined the impact of security and privacy improvements on participants' purchase behavior when presented *individually*. In phase two, we examined the impact of five security and privacy improvements on participants' purchase behavior when presented *together* as they might be on an IoT label.

5.1 Study Design

Label comparison pairs. We designed a within-subjects survey for phase two. Based on the phase-one findings, we created three types of security and privacy labels: *least protective*, *most protective*, and *no information label*. We specified the attributes on the *most protective* and *least protective* labels based on how each tested security and privacy comparison pair in phase one influenced participants' perceived risk (see Section 5.3). We specified the security and privacy attributes on the *no information label* to "not disclosed." We constructed three *label comparison pairs* (e.g., *least protective* label vs. *most protective* label), which we used as the within-subjects factor. From phase one, we found that device type is not a significant factor in changing participants' purchase behavior. Thus, we only included smart speaker in phase two.

Screening questions. In our screening survey, we first asked attention-check questions about the device specifications (PH2.1-3) and then asked screening questions on participants' interest in purchasing the smart device (PH2.4). We aimed to recruit participants who had at least two correct responses to the attention-check questions and were also moderately or very interested in participating in our study and purchasing a soon-to-be-released smart speaker. The screening questions took on average 7 minutes to answer. Participants who did not continue after the screening questions were compensated with US$1.50.

Survey questions. To capture participants' understanding of the label security and privacy information, we asked them multiple-choice questions about all the elements of the least protective and most protective labels (PH2.5). In each question, we asked participants to select the correct answer related to the definition of the presented attribute-value. Three of the answer choices were based on the common misunderstandings we found from participants' definitions in the first phase and the fourth choice was the correct answer. To ensure participants had a correct understanding of the security and privacy attribute-values, we only analyzed data from participants who correctly answered all the definition-related questions.

For each pair, we asked how and why the pair would impact participants' risk perception (PH2.6-9) and willingness to purchase (PH2.10-13). We then used MPL to elicit participants' willingness to pay for each label comparison pair (PH2.14). We extended the MPL premium range to $0–$45 (compared to $0–$25 in phase one) due to expected added value of having multiple security and privacy improvements on the label as opposed to individual improvements in phase one. We ended the second-phase survey with demographic questions (PH2.15-19).

5.2 Data Analysis

Similarly to the phase-one study, we constructed three regression models to quantitatively analyze participants' responses in phase two. The dependent variables (DVs) in the three models were as follows:

- Monetary valuation (ranging from -$45 to $45).
- Change in willingness to purchase (5 levels): 1 (strong decrease), 2 (slight decrease), 3 (no impact), 4 (slight increase), 5 (strong increase).
- Change in risk perception (5 levels): 1 (strong decrease), 2 (slight decrease), 3 (no impact), 4 (slight increase), 5 (strong increase).

We conducted model selection with backward elimination to find the models that best fit phase-two close-ended responses. Here, each model had a single independent variable (IV), `label_comparison`, with 3 levels: 1) least protective label vs. most protective label, 2) no information label vs. most protective label, and 3) no information label vs. least protective label. Similar to phase one, the second phase of our study had a repeated measures design. Therefore, for all models, we included random effects to count for within-participants data

Row	Model Factor	Willingness to Pay (AIC = 809.1)				Willingness to Purchase (AIC = 213.8)					Risk Perception (AIC = 228.7)				
		Estimate	CI	SE	p-value	Estimate	SE	OR_+	OR_-	p-value	Estimate	SE	OR_+	OR_-	p-value
	`label_comparison` (baseline = No information label vs. Least protective label)														
1	Least protective label vs. Most protective label	61.19	[54.69, 67.68]	3.31	***	9.01	1.54	8183.70	0.00	***	−7.31	1.21	0.00	1494.73	***
2	No information label vs. Most protective label	56.46	[49.96, 62.96]	3.32	***	7.61	1.34	2026.37	0.00	***	−5.26	0.90	0.01	192.08	***
	threshold coefficients														
3	1|2	-	-	-	-	−1.30	0.47	-	-	-	−3.73	0.72	-	-	-
4	2|3	-	-	-	-	−0.31	0.41	-	-	-	−2.86	0.63	-	-	-
5	3|4	-	-	-	-	3.26	0.75	-	-	-	0.38	0.34	-	-	-
6	4|5	-	-	-	-	5.39	1.05	-	-	-	1.66	0.45	-	-	-
	intercept														
7	α_0	−22.83	[−27.77, −17.90]	2.52	***	-	-	-	-	-	-	-	-	-	-
	random effects														
8	σ_u^2	0.00	-	-	-	2.35	-	-	-	-	0.95	-	-	-	-

Note: *$p < 0.05$ **$p < 0.01$ ***$p < 0.001$

Table 4: Regression results of the phase-two study, corresponding to the mixed-effects interval regression model to describe participants' willingness to pay, as well as the CLMMs to describe participants' willingness to purchase and risk perception. Each row corresponds to a single factor, and shows the resulting model estimates, i.e., coefficients, for that factor, alongside the standard error (SE), and p-value for all the three models. For the willingness to pay model, we also include the confidence interval (CI) of the monetary estimate of each factor. Furthermore, for the willingness to purchase (resp., risk perception) model, we include the odds ratios of increased and decreased willingness to purchase (resp., risk perception) for both the label comparison pairs. We also include the AIC values for all the three models, which represent the models' goodness of fit.

dependencies. We used the same regression analysis methods as in phase one to construct CLMM and mixed interval regression models (see Section 4.2.1). We also followed the same qualitative analysis method as in phase one to qualitatively code the open-ended responses (see Section 4.2.2).

5.3 Findings

We conducted the power analysis based on the mixed-effects regression model we planned to run for the phase-two study. The analysis showed that we need at least 47 responses for each label comparison pair. We reached our goal after recruiting 250 participants in the screening survey. Among those, 68 answered the main survey. We further excluded 18 participants who had at least one incorrect answer to the attention-check questions. Thus, we ended up with 50 participants, leading to 50 observations for each label comparison pair. Participant demographics are provided in Appendix D.

In our phase-two study, we explored how much participants value having transparency over security and privacy practices of smart devices when presented holistically on an IoT label. Given the findings of phase one, we created three label types: most protective, least protective, and no information (see Table 3). We chose the attribute-values to use on the most and least protective labels by considering the regression coefficients of the risk perception model in Table 2. For each attribute, we selected the comparison pair that had the largest impact on participants' risk perception. The resulting comparison pairs for the five attributes are:

- Cloud storage: Identifiable vs. De-identified
- Access control: None vs. Multi-factor authentication
- Security update: None vs. Automatic
- Data sharing: Third parties vs. None
- Purpose: Functionality & Monetization vs. Functionality

For each comparison pair, we added the left-side value (i.e., the less protective one) to the least protective label and its right-side value (i.e., the more protective one) to the most protective label. We marked all the values in the no information label scenario as 'not disclosed' (see Table 3).

In the survey, we presented three smart speaker models to participants: Model X with the most protective label, Model Y with no information label, and Model Z with the least protective label. These three models had identical technical specifications (see Appendix B) and only differed in their security and privacy attributes. Using the three label types, we created three label comparison pairs (`label_comparison`): no information label vs. most protective label, no information label vs. least protective label, and least protective label vs. most protective label. For each label comparison pair (e.g., no information label vs. most protective label), we expected the left component (e.g., no information label) to be less desired than the right component (e.g., most protective label). Therefore, we hypothesized that participants would be willing to pay significantly more for having the smart device with the right component compared to the left component.

Our results showed that participants were indeed willing to pay more for multiple security and privacy improvements than any individual one. However, the premium they were willing to pay for multiple improvements was less than the sum of premiums for individual improvements. The regression analysis (see Table 4) showed that compared to having risky security and privacy practices or no transparency, participants were significantly more willing to purchase (row 1, estimate = 9.01, p-value < 0.001 and row 2, estimate = 7.61, p-value < 0.001) and willing to pay significantly higher premiums (row 1, premium= \$38.36, estimate = \$61.19, p-value < 0.001 and row 2, premium= \$33.63, estimate = \$56.46, p-value < 0.001) to have a smart device with improved security and privacy practices. Note that the difference between the model estimates and the premiums is due to the model intercept (row 7, $\alpha_0 = -\$22.83$). P28 compared Model X (most protective label) and Model Z (least protective label):

> *I would definitely pay more for Model X that comes with good practices vs Z that has really bad ones. Model X has privacy and security that seems to have the user's safety in mind. It protects me and my information, by not sharing my data, providing automatic security updates, and so on. Model Z seems to exist to benefit everyone but*

the user, since its main focus is profit, it shares data with third parties, provides no security update, etc.

Contrary to our hypothesis, participants preferred purchasing a smart device with no security or privacy information compared to a smart device with least protective label. The regression analysis showed that compared to having no transparency, participants were willing to pay significantly less for a device that has the least protective label (row 7, estimate $= -\$22.83$, p-value < 0.001). Several participants (15/50) reported that without a label, they would assume the device is similar to other devices on the market and follows similar security and privacy practices. P5 compared Model Y (no information label) and Model Z (least protective label):

Model Y's information is not disclosed, which makes me think it is likely similar to other models in the market ... which are probably less risky [than Model Z].

6 Limitations

We recruited participants through Prolific. Although commonly used in user research, crowdsourcing platforms are not representative of the average population [80]. In addition, our study only examined online purchase behavior of US participants. Surveys have shown that the majority of US consumers of smart devices make their device purchases online [46]. However, it is still important to study the valuation of in-person consumers and from different countries. Despite these limitations, online crowdsourcing platforms are commonly used to elicit consumers' IoT purchase behavior [16, 71]. In addition, we recruited only US participants, limiting our conclusions to the US context.

We used multiple price list (MPL) to elicit participants' true monetary valuation for security and privacy, which introduces limitations. In real-world purchase scenarios, consumers usually do not fill out a survey. We used a survey design to control for study factors without introducing confounding variables of real purchase settings. In addition, the premium that participants could pay and the step size were limited by the MPL table [3, 9]. Although the amount of premium could depend on several factors, including the base price of the device, we believe our study accurately provides the *relative importance* of security and privacy comparison pairs and labels.

Moreover, in our study, we used the same security and privacy practices identified by Emami-Naeini et al. [34]. However, participants could have various familiarity levels with the tested practices. The familiarity level and how realistic and useful they perceived the practices to be could impact their risk perception and purchase behavior, which should be considered in future work.

7 Discussion

Through our two-phase incentive-compatible design, we quantified the monetary value of IoT security and privacy factors communicated to consumers at the point of sale, either individually or holistically. Our statistical analysis showed that consumers are willing to pay a significant dollar amount for purchasing the smart device with improved security and privacy practices when considering the two devices together. However, increased protection was not always enough: our qualitative analysis showed that concerns about usability (password vs. multi-factor authentication, row 13 in Table 2), and convenience (de-identified cloud storage vs. no cloud storage, row 14 in Table 2) could push consumers away from purchasing the more secure smart devices, even when consumers perceived them as lower risk.

Willingness to purchase leads to willingness to pay. In their recent work [35], Emami-Naeini et al. explored the role of IoT security and privacy practices on participants' hypothetical purchase behavior, studying how the most and least protective values of each security or privacy attribute would impact participants' risk perception and willingness to purchase a device. We build on [35] by considering three protection values (least protective, medium protective, and most protective) for each security or privacy attribute and use an incentive-compatible method to quantify the premium consumers are willing to pay to have improved security and privacy. We found similar risk perception and willingness to purchase results for the attribute-values as in [35]. However, our current study also found that participants' willingness to purchase was aligned with their willingness to pay for almost all improved security and privacy practices (see Table 2).

Need for mandatory IoT labels. In phase two of our study, we studied the monetary value of security and privacy features when presented holistically on an IoT label (see Section 5). Our quantitative analysis showed that consumers are willing to pay a significant premium for smart devices with improved security and privacy (see rows 1 and 2 in Table 4). However, consumers also indicated a significantly higher willingness to pay for a smart device with no security and privacy information than a device with a label indicating risky security and privacy practices (see row 7 in Table 4). Our qualitative analysis showed that when security and privacy information was not mentioned, participants assumed that the device's practices were not that risky (see Section 5.3). This finding suggests that if not required to adopt the IoT label, device manufacturers with insecure and privacy-invasive practices can leverage this and avoid disclosing their practices.

Currently, IoT security and privacy labels have been deployed as a voluntary program in a few countries, including Singapore [28] and Finland [38], and is on their way to be deployed in Australia [90]. Similarly, in the US, policymakers in collaboration with the National Institute of Standards and Technology (NIST) and other organizations are currently working on specifying how the labeling program should look like [73]. In order to put less pressure on the IoT market, current national and international IoT labeling efforts are focused on incentivizing IoT device manufacturers to voluntarily adopt the security and privacy labels. Confirming past research [13], our findings (see Section 5) suggest that voluntary labeling programs might not be as effective in in-

creasing consumers' awareness and protection in the long run. To help consumers make informed purchase decisions when doing comparison shopping, we recommend that policymakers implement mandatory labeling programs. Enforcing IoT labels, however, has its own challenges for stakeholders, including consumers and device manufacturers, that need to be carefully studied and addressed. One of the challenges of enforcing product labels could be information overload [32, 37]. Although this challenge is not unique to IoT security and privacy labels [89], the complex nature of security and privacy information could further lead to consumer confusion when making device purchases. Moreover, IoT labeling programs could pose a large financial cost for device manufacturers. The cost of labeling has been estimated to be $4,000 on the lower end and up to $700,000 for large device companies [30].

Education needs to accompany the label. Prior to presenting our definitions to participants, we asked them to define the attribute-values of each security and privacy comparison pair (see Section 4.3.1). Our regression models (see rows 16 and 17 in Table 2) indicated that those who correctly defined the security and privacy attribute-values were willing to pay a significant premium to purchase a more protective smart device, implying that prior knowledge about the attributes is important. This accords with the food domain, where prior knowledge about the ingredients has been shown to significantly impact the effectiveness of nutrition labels [23, 67, 69].

In addition, our regression models showed that participants were willing to pay more for having security and privacy improvements that decreased their risk perception (see models Willingness to Pay and Risk Perception in Table 2). This suggests that consumers who have a better understanding of the risks associated with security and privacy practices have a higher appreciation of the improved protection provided by their smart devices. Due to lack of usability and convenience, our qualitative analysis (see Section 4.3.2) showed that some participants would decline to take additional steps to improve their security and privacy (e.g., using multi-factor authentication, as opposed to having password). Knowing about the potential risks could be especially useful for communicating the value of this type of practices.

Prior research has shown that people are prone to loss aversion [95] and that a loss-framed message leads to more secure behavior [81]. Therefore, when people have a better understanding of the risks, they would be more willing to protect themselves. This suggests that to further inform consumers and help them value the offered security and privacy protections, IoT labels should disclose the potential risks of not having the offered security or privacy practices. For example, in the case of security updates, the manufacturer could use the device label to disclose the potential risks of not receiving the updates automatically or forgetting to install them manually.

To further inform consumers' purchase decision-making, IoT labels should come with an educational component. The significance of having a robust consumer education program has been highlighted in the NIST proposal to design IoT security and privacy labels [73]. The NIST document provided a few factors that should be included in such an educational component, but presented no empirical evidence to justify their efficacy. Based on our research findings, we recommend that designers of such educational programs include usable information on the definitions of the devices' security and privacy practices and the risks associated with each practice.

8 Conclusion

Due to lack of information, consumers are unable to consider security and privacy when purchasing smart devices and it is unclear how much they would value such information if being presented at the point of sale. We recruited 180 participants and conducted a two-phase incentive-compatible study on Prolific to explore how much participants value purchasing smart devices with enhanced security and privacy practices when being communicated on an IoT device label. Our findings showed that participants were willing to pay significant premiums to purchase a smart device with a single improved security and privacy practice, such as having de-identified vs. identifiable cloud storage. Moreover, we found that presenting multiple security and privacy protections together drives the premiums consumers would be willing to pay even higher.

Acknowledgments

We thank our reviewers and our shepherd for their invaluable feedback. This work was supported in part by NSF award SaTC-1801472, and the Carnegie Mellon University CyLab Security and Privacy Institute.

References

[1] Jack Abrams. A new method for testing pricing decisions. *Journal of Marketing*, 28(3):6–9, 1964.

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

[3] Wiktor Adamowicz, Jordan Louviere, and Michael Williams. Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management*, 26(3):271–292, 1994.

[4] Azimatun Noor Aizuddin, Saperi Sulong, and Syed Mohamed Aljunid. Factors influencing willingness to pay for healthcare. In *BMC Public Health*, volume 12, pages 1–1. Springer, 2012.

[5] Ulas Akkucuk and Javed Esmaeili. The impact of brands on consumer buying behavior: An empirical study on smartphone buyers. *International Journal of Research in Business and Social Science (2147-4478)*, 5(4):1–16, 2016.

[6] Frode Alfnes, Kyrre Rickertsen, et al. Non-market valuation: experimental methods. *The Oxford Handbook of the Economics of Food Consumption and Policy*, 215:242, 2011.

[7] Roselyne Alphonce and Frode Alfnes. Eliciting consumer WTP for food characteristics in a developing context: Application of four valuation methods in an African market. *Journal of Agricultural Economics*, 68(1):123–142, 2017.

[8] Steffen Andersen, Glenn W Harrison, Morten Igel Lau, and E Elisabet Rutström. Elicitation using multiple price list formats. *Experimental Economics*, 9(4):383–405, 2006.

[9] Steffen Anderson, Glenn W Harrison, Morten I Lau, and Rutstrom E Elisabet. Valuation using multiple price list formats. *Applied Economics*, 39(6):675–682, 2007.

[10] India Ashok. Hackers leave finnish residents cold after ddos attack knocks out heating systems. https://www.ibtimes.co.uk/hackers-leave-finnish-residents-cold-after-ddos-attack-knocks-out-heating-systems-1590639, November 2016.

[11] Daniele Asioli, Adriana Mignani, and Frode Alfnes. Quick and easy? respondent evaluations of the Becker–DeGroot–Marschak and multiple price list valuation mechanisms. *Agribusiness*, 37(2):215–234, 2021.

[12] Avast Antivirus. Two out of five digital households worldwide at cyber risk, Avast reveals. https://press.avast.com/two-out-of-five-digital-households-worldwide-at-cyber-risk-avast-reveals, February 2019.

[13] Hosein Badran. IoT security and consumer trust. In *Proceedings of the 20th Annual International Conference on Digital Government Research*, pages 133–140, 2019.

[14] Laura Bakkensen and Paul Schuler. A preference for power: Willingness to pay for energy reliability versus fuel type in Vietnam. *Energy Policy*, 144:111696, 2020.

[15] Gordon M Becker, Morris H DeGroot, and Jacob Marschak. Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3):226–232, 1964.

[16] John M Blythe, Shane D Johnson, and Matthew Manning. What is security worth to consumers? investigating willingness to pay for secure internet of things devices. *Crime Science*, 9(1):1–9, 2020.

[17] Sarah Brebner and Joep Sonnemans. Does the elicitation method impact the WTA/WTP disparity? *Journal of Behavioral and Experimental Economics*, 73:40–45, 2018.

[18] Christoph Breidert, Michael Hahsler, and Thomas Reutterer. A review of methods for measuring willingness-to-pay. *Innovative Marketing*, 2(4):8–32, 2006.

[19] Magdalena Brzozowicz et al. Hypothetical bias and framing effect in the valuation of private consumer goods. *Central European Economic Journal*, 5(52):260–269, 2018.

[20] Kenneth P Burnham and David R Anderson. Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2):261–304, 2004.

[21] Trudy A Cameron and Michelle D James. Estimating willingness to pay from survey data: an alternative pre-test-market evaluation procedure. *Journal of Marketing Research*, 24(4):389–395, 1987.

[22] Maurizio Canavari, Andreas C Drichoutis, Jayson L Lusk, and Rodolfo M Nayga Jr. How to run an experimental auction: A review of recent advances. *European Review of Agricultural Economics*, 46(5):862–922, 2019.

[23] E Carrillo, P Varela, and S Fiszman. Influence of nutritional knowledge on the use and interpretation of spanish nutritional food labels. *Journal of Food Science*, 77(1):H1–H8, 2012.

[24] Hasan Huseyin Ceylana, Bekir Koseb, and Mufit Aydin. Value based pricing: A research on service sector using Van Westendorp price sensitivity scale. *Procedia-Social and Behavioral Sciences*, 148:1–6, 2014.

[25] Alexander Chernev. Reverse pricing and online price elicitation strategies in consumer choice. *Journal of Consumer Psychology*, 13(1-2):51–62, 2003.

[26] Rune Haubo B Christensen. Cumulative link models for ordinal regression with the R package ordinal. *Submitted in J. Stat. Software*, 2018.

[27] Lorrie Faith Cranor, Joseph Reagle, and Mark S Ackerman. Beyond concern: Understanding net users' attitudes about online privacy. *The Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy*, pages 47–70, 2000.

[28] Cyber Security Agency. Cybersecurity labelling scheme. https://www.csa.gov.sg/programmes/cybersecurity-labelling.

[29] Herbert Joseph Davenport. Proposed modifications in Austrian theory and terminology. *The Quarterly Journal of Economics*, 16(3):355–384, 1902.

[30] DCMS. Evidencing the Cost of the UK Government's Proposed Regulatory Interventions for Consumer IoT. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/900330/Evidencing_the_cost_of_the_UK_government_s_proposed_regulatory_interventions_for_consumer_internet_of_things__IoT__products.pdf. Accessed: 2022-6-6.

[31] Andreas C Drichoutis and Jayson L Lusk. What can multiple price lists really tell us about risk preferences? *Journal of Risk and Uncertainty*, 53(2-3):89–106, 2016.

[32] Angela Edmunds and Anne Morris. The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management*, 20(1):17–28, 2000.

[33] Pardis Emami-Naeini, Yuvraj Agarwal, and Lorrie Faith Cranor. Specification for CMU IoT security and privacy label.

[34] 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 *2020 IEEE Symposium on Security and Privacy (SP)*, pages 447–464. IEEE, 2020.

[35] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. Which privacy and security attributes most impact consumers' risk perception and willingness to purchase IoT devices? In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 1937–1954, 2021.

[36] 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*, pages 1–12, 2019.

[37] Martin J Eppler and Jeanne Mengis. The concept of information overload-a review of literature from organization science, accounting, marketing, MIS, and related disciplines (2004). *Kommunikationsmanagement im Wandel*, pages 271–305, 2008.

[38] Finnish Transport and Communication Agency. Finland becomes the first European country to certify safe smart devices – new cybersecurity label helps consumers buy safer products. https://www.traficom.fi/en/news/finland-becomes-first-european-country-certify-safe-smart-devices-new-cybersecurity-label.

[39] Ajit Ghuman. Research: A market where consumers can pay for privacy is emerging. https://venturebeat.com/2021/04/30/research-a-market-where-consumers-can-pay-for-privacy-is-emerging/, April 2021.

[40] Shakthidhar Gopavaram, Jayati Dev, Sanchari Das, and L Jean 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.

[41] Emily Green. Hacker terrorizes family by hijacking baby monitor. https://nordvpn.com/blog/baby-monitor-iot-hacking/, December 2018.

[42] Paul E Green, Abba M Krieger, and Yoram Wind. Thirty years of conjoint analysis: Reflections and prospects. *Interfaces*, 31(3_supplement):S56–S73, 2001.

[43] Jens Grossklags and Alessandro Acquisti. When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. In *WEIS*, 2007.

[44] Jasper Hamill. 'Panty Buster' sex toys can be hacked to 'remotely pleasure people without their consent', researchers claim. https://metro.co.uk/2018/02/01/panty-buster-sex-toys-can-hacked-remotely-pleasure-people-without-consent-researchers-claim-7279177/, February 2018.

[45] Glenn W Harrison and E Elisabet Rutström. Experimental evidence on the existence of hypothetical bias in value elicitation methods. *Handbook of experimental economics results*, 1:752–767, 2008.

[46] Andrea Harvey. 3 in 4 Americans bought a smart home tech device this past year. https://www.safewise.com/blog/smart-home-tech-spending/. Accessed: 2022-6-6.

[47] James J Heckman. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161, 1979.

[48] David A Hensher. Hypothetical bias, choice experiments and willingness to pay. *transportation research part B: methodological*, 44(6):735–752, 2010.

[49] Nick Ho-Sam-Sooi, Wolter Pieters, and Maarten Kroesen. Investigating the effect of security and privacy on IoT device purchase behaviour. *computers & security*, 102:102132, 2021.

[50] Elizabeth Hoffman, Dale J Menkhaus, Dipankar Chakravarti, Ray A Field, and Glen D Whipple. Using laboratory experimental auctions in marketing research: a case study of new packaging for fresh beef. *Marketing Science*, 12(3):318–338, 1993.

[51] Leonid Hurwicz. On informationally decentralized systems. *Decision and organization: A volume in Honor of J. Marschak*, 1972.

[52] Harris Interactive. A survey of consumer privacy attitudes and behaviors. *Rochester, NY*, 47, 2000.

[53] Kamel Jedidi and Z John Zhang. Augmenting conjoint analysis to estimate consumer reservation price. *Management Science*, 48(10):1350–1368, 2002.

[54] Shane D Johnson, John M Blythe, Matthew Manning, and Gabriel TW Wong. The impact of IoT security labelling on consumer product choice and willingness to pay. *PloS one*, 15(1):e0227800, 2020.

[55] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. Experimental tests of the endowment effect and the coase theorem. *Journal of political Economy*, 98(6):1325–1348, 1990.

[56] Beata Kupiec and Brian Revell. Measuring consumer quality judgements. *British Food Journal*, 2001.

[57] Marine Le Gall-Ely. Definition, measurement and determinants of the consumer's willingness to pay: a critical synthesis and avenues for further research. *Recherche et Applications en Marketing (English Edition)*, 24(2):91–112, 2009.

[58] Nicole Lindsey. Smart devices leaking data to tech giants raises new IoT privacy issues. https://www.cpomagazine.com/data-privacy/smart-devices-leaking-data-to-tech-giants-raises-new-iot-privacy-issues/, October 2019.

[59] Jordan J Louviere and George Woodworth. Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of Marketing Research*, 20(4):350–367, 1983.

[60] Jayson L Lusk, Deacue Fields, and Walt Prevatt. An incentive compatible conjoint ranking mechanism. *American Journal of Agricultural Economics*, 90(2):487–498, 2008.

[61] Jayson L Lusk, Jason F Shogren, et al. Experimental auctions. *Methods and Applications in Economic and Marketing Research*, pages 46–94, 2007.

[62] David W Lyon. The price is right (or is it?). *Marketing Research*, 14(4):8, 2002.

[63] Giuseppina Migliore, Massimiliano Borrello, Alessia Lombardi, and Giorgio Schifani. Consumers' willingness to pay for natural food: evidence from an artefactual field experiment. *Agricultural and Food Economics*, 6(1):1–10, 2018.

[64] Carrie Mihalcik. Apple HomePod mini reportedly has a secret sensor for temperature, humidity. https://www.cnet.com/home/smart-home/apple-homepod-mini-reportedly-has-a-secret-sensor-for-temperature-humidity/, March 2021.

[65] Klaus Miller, Reto Hofstetter, Harley Krohmer, and John Zhang. Measuring consumers' willingness to pay. which method fits best? *GfK Marketing Intelligence Review*, 4(1):42–49, 2012.

[66] Klaus M Miller, Reto Hofstetter, Harley Krohmer, and Z John Zhang. How should consumers' willingness to pay be measured? an empirical comparison of state-of-the-art approaches. *Journal of Marketing Research*, 48(1):172–184, 2011.

[67] Lisa M Soederberg Miller and Diana L Cassady. The effects of nutrition knowledge on food label use. a review of the literature. *Appetite*, 92:207–216, 2015.

[68] Robert Cameron Mitchell and Richard T Carson. *Using surveys to value public goods: the contingent valuation method*. RFF Press, 2013.

[69] Sally G Moore, Judy K Donnelly, Steve Jones, and Janet E Cade. Effect of educational interventions on understanding and use of nutrition labels: A systematic review. *Nutrients*, 10(10):1432, 2018.

[70] Philipp Morgner and Zinaida Benenson. Exploring security economics in IoT standardization efforts. *arXiv preprint arXiv:1810.12035*, 2018.

[71] Philipp Morgner, Christoph Mai, Nicole Koschate-Fischer, Felix Freiling, and Zinaida Benenson. Security update labels: establishing economic incentives for security patching of IoT consumer products. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 429–446. IEEE, 2020.

[72] James J Murphy, P Geoffrey Allen, Thomas H Stevens, and Darryl Weatherhead. A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, 30(3):313–325, 2005.

[73] National Institute of Standards and Technology. Recommended Criteria for Cybersecurity Labeling of Consumer Software. https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.02042022-2.pdf. Accessed: 2022-6-6.

[74] Alfred Ng and Megan Wollerton. Google calls Nest's hidden microphone an "error". https://www.cnet.com/news/google-calls-nests-hidden-microphone-an-error/, February 2019.

[75] Kenneth D Nguyen, Heather Rosoff, and Richard S John. Valuing information security from a phishing attack. *Journal of Cybersecurity*, 3(3):159–171, 2017.

[76] Christi Olson. New report tackles tough questions on voice and AI. https://about.ads.microsoft.com/en-us/blog/post/april-2019/new-report-tackles-tough-questions-on-voice-and-ai. Accessed: 2022-6-6.

[77] Bryan Orme. Which Conjoint Method Should I Use?, Sawtooth Software research paper series: Sawtooth Software, 2003.

[78] Jeffrey Prince and Scott Wallsten. How much is privacy worth around the world and across platforms? *Available at SSRN 3528386*, 2020.

[79] Prolific. Quickly find research participants you can trust. https://www.prolific.co/. Accessed: 2022-6-6.

[80] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. How well do my results generalize? comparing security and privacy survey results from MTurk, web, and telephone samples. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 1326–1343. IEEE, 2019.

[81] Nuria Rodríguez-Priego, René van Bavel, José Vila, and Pam Briggs. Framing effects on online security behavior. *Frontiers in Psychology*, 11:2833, 2020.

[82] Oliver Roll, Lars-Hendrik Achterberg, and Karl-Georg Herbert. Innovative approaches to analyzing the price sensitivity meter: Results of an international comparative study. *Laurea Publications A• 72*, page 181, 2010.

[83] Brent Rowe and Dallas Wood. Are home internet users willing to pay ISPs for improvements in cyber security? In *Economics of information security and privacy III*, pages 193–212. Springer, 2013.

[84] Martin Sadler. Securing our connected world. https://dcmsblog.uk/2017/10/securing-connected-world/, October 2017.

[85] Johnny Saldaña. *The coding manual for qualitative researchers*. Sage, 2015.

[86] Alex Schiffer. How a fish tank helped hack a casino. https://www.washingtonpost.com/news/innovations/wp/2017/07/21/how-a-fish-tank-helped-hack-a-casino/?noredirect=on, July 2017.

[87] Jonas Schmidt and Tammo HA Bijmolt. Accurately measuring willingness to pay for consumer goods: a meta-analysis of the hypothetical bias. *Journal of the Academy of Marketing Science*, 48(3):499–518, 2020.

[88] Michel Schreiner and Thomas Hess. On the willingness to pay for privacy as a freemium model: First empirical evidence. 2013.

[89] Lara Spiteri Cornish and Caroline Moraes. The impact of consumer confusion on nutrition literacy and subsequent dietary behavior. *Psychology & Marketing*, 32(5):558–574, 2015.

[90] Standards Australia. Iconic Nation. https://standards.org.au/getattachment/22868f05-90a0-4ade-a1ea-99153908dea5/H_1870-The-Standards-Australia. Accessed: 2022-6-6.

[91] LP StataCorp. Stata multilevel mixed-effects reference manual. 2021.

[92] Dan Svirsky. Why are privacy preferences inconsistent? *The Harvard John M. Olin Fellow's Discussion Paper Series*, 81, 2018.

[93] 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.

[94] Joseph Turow, Lauren Feldman, and Kimberly Meltzer. Open to exploitation: America's shoppers online and offline. *Departmental Papers (ASC)*, page 35, 2005.

[95] Amos Tversky and Daniel Kahneman. Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4):1039–1061, 1991.

[96] United States Census Bureau. Educational attainment in the United States: 2020. https://www.census.gov/data/tables/2020/demo/educational-attainment/cps-detailed-tables.html. Accessed: 2022-6-6.

[97] Peter H Van Westendorp. NSS Price Sensitivity Meter (PSM)–A new approach to study consumer perception of prices. In *Proceedings of the 29th ESOMAR Congress*, volume 139167, 1976.

[98] Riccardo Vecchio and Massimiliano Borrello. Measuring food preferences through experimental auctions: A review. *Food Research International*, 116:1113–1120, 2019.

[99] Kaveh Waddell. Connected devices share more data than needed, study says. https://www.consumerreports.org/privacy/connected-devices-share-more-data-than-needed-study-says-a7015033345/, May 2021.

[100] Klaus Wertenbroch and Bernd Skiera. Measuring consumers' willingness to pay at the point of purchase. *Journal of marketing research*, 39(2):228–241, 2002.

[101] Dick R Wittink, Marco Vriens, and Wim Burhenne. Commercial use of conjoint analysis in Europe: Results and critical reflections. *International journal of Research in Marketing*, 11(1):41–52, 1994.

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

[103] 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.

A Survey Procedure

The survey questions and debriefing statements can be found at https://anonymous.4open.science/r/MonteryValueSPSurvey-673D/USENIX2023_MonetaryValueofSP_CameraReady_SurveyAppendix.pdf.

B Smart Device Specifications

We prepared a specification table for each of the four smart devices we included in the study (we selected the smart speaker and smart smoke detector for phase-one study and smart speaker for phase-two study). Each sheet in the following link shows a copy of the specification of one of the smart devices: https://docs.google.com/spreadsheets/d/1KE-ALWH_bZGHSVTMhJDf1RDE3GcCLwPXvvmbOHfVTR0/edit?usp=sharing.

C Attribute-Value Definitions

The attribute-value definitions can be found in Table 5.

D Demographic Information

Table 6 contains participants' demographic information.

E Procedure to Derive Participants' Willingness to Pay Limits

For each of the five comparison pairs shown to each participant, we categorized their willingness to purchase response into one of the following three buckets:

1. Slightly/strongly increased willingness to purchase: This was aligned with our hypothesis, in which case, we set aligned = True, $L_{\text{init}} = \$0$, and $U_{\text{init}} = \$25$.
2. Slightly/strongly decreased willingness to purchase: This was contrary to our hypothesis, in which case, we set aligned = False, $L_{\text{init}} = \$0$, and $U_{\text{init}} = \$25$.
3. No impact on willingness to purchase: In this case, we set aligned = True or aligned = False, each with 50% probability, and we set $L_{\text{init}} = -\$25$ and $U_{\text{init}} = \$25$.

In the above scenarios, aligned is a Boolean variable indicating whether the response matched our hypothesis, and $(L_{\text{init}}, U_{\text{init}})$ respectively denote the potential lowest and highest premiums the participant was willing to pay. Then, denoting the participant's sequence of selected elements by seq, we used Algorithm 1 to determine the lowest and highest premiums the participant was willing to pay for the comparison pair. This algorithm finds the switching point(s) between the three options (R, N, and L, corresponding to the right, no preference, and left options, respectively), and sets the lower and upper limits of the participant's willingness to pay based on the premiums at those switching point(s). It finally negates the resulting limits if the participant's willingness to purchase was contrary to our hypothesis.

F Proof of (4)

For a given factor f, $\text{OR}^{f}_{\text{purchase},+}$ is defined as

$$\text{OR}^{f}_{\text{purchase},+} := \frac{\left(\frac{\Pr(\text{increased w.t. purchase} \mid f)}{1-\Pr(\text{increased w.t. purchase} \mid f)}\right)}{\left(\frac{\Pr(\text{increased w.t. purchase} \mid f_{\text{baseline}})}{1-\Pr(\text{increased w.t. purchase} \mid f_{\text{baseline}})}\right)}, \tag{6}$$

where $\Pr(\text{increased w.t. purchase} \mid f)$ and $\Pr(\text{increased w.t. purchase} \mid f_{\text{baseline}})$ denote the probabilities that the willingness to purchase slightly/strongly increases given the factor f and the factor baseline f_{baseline}, respectively.

Attribute-Value	Definition
No access control	Anyone can access the device without a password or other authentication method.
Changeable default password	Password is required to access the device settings or data and user may change that password.
Multi-factor authentication	At least two factors are required to access the device settings or data, for example a password and a one-time code sent to a previously registered phone number.
Identifiable cloud storage	User's identity could be revealed from the data stored in the cloud.
De-identified cloud storage	The data stored in the cloud does not contain any personal identifiers that reveal a user's identity.
No cloud storage	The collected data will not be stored in the cloud.
Data sharing with third parties	The collected data will be shared with at least one third party.
Data sharing with manufacturer	The collected data will be shared with the device manufacturer.
Data sharing with no one	The collected data will not be shared with anyone.
Data collection for monetization	Data is collected to provide main device features, improve services, and help develop new features, and the manufacturer receives income from sending user tailored advertisements or selling user's data to third parties.
Data collection for personalization	Data is collected to provide main device features, improve services, and help develop new features, and to provide user with personally relevant features and customized content.
Data collection for device functionality	Data is collected to provide main device features, improve services, and help develop new features.
No security updates	The device will not receive any security updates.
Manual security updates	The user needs to manually install security updates.
Automatic security updates	The device will automatically receive security updates.

Table 5: Attribute-value definitions that we presented to participants. The first column contains the exact wordings that we used for attribute-values when we asked participants to define them.

Metric	Levels	Pre-Study (%)				Phase-One Study (%)	Phase-Two Study (%)	Census (%)
		Speaker	Security camera	Smoke detector	Motion detector			
Gender	Male	31	54	37	48	52	46	48
	Female	65	46	63	48	46	48	52
	Non-binary	4	0	0	4	2	6	–
Age	18-29 years	52	42	42	56	42	42	21
	30-49 years	39	46	46	40	48	38	33
	50-64 years	9	12	8	4	9	18	24
	65+ years	0	0	4	0	1	2	22
Education	No high school	0	0	4	0	1	0	10
	High school	4	17	8	12	7	6	29
	Some college (no degree)	31	25	21	24	25	28	17
	Associate	0	0	8	4	7	8	10
	Bachelor	39	42	39	32	39	40	22
	Master	13	12	12	16	18	16	9
	Professional	0	0	0	0	0	0	1
	Doctoral	13	4	8	12	3	2	2
Income	< $10,000	9	0	4	4	4	6	5
	$10,000–$19,999	4	4	4	12	6	4	8
	$20,000–$29,999	13	4	0	16	8	6	8
	$30,000–$39,999	13	4	8	8	13	18	8
	$40,000–$49,999	17	17	30	12	12	6	8
	$50,000–$59,999	0	8	8	8	8	8	7
	$60,000–$69,999	4	0	8	4	9	6	6
	$70,000–$79,999	9	17	4	8	7	4	6
	$80,000–$89,999	4	4	0	4	4	8	5
	$90,000–$99,999	0	0	4	0	5	4	5
	$100,000–$149,999	18	21	17	8	13	14	15
	$150,000 or more	0	12	4	12	7	10	19
	Prefer not to answer	9	9	9	4	4	6	–
Tech Background	Yes	26	33	25	16	40	36	–
	No	74	67	75	84	60	64	–

Table 6: Demographic information of our participants in pre-study, phase-one and phase-two surveys. Compared to the 2020 US Census data [96], our study participants were younger and with higher education levels.

Algorithm 1 Lower and upper limits of willingness to pay

Input: aligned, L_{init}, U_{init}, seq
Output: L, U

if seq includes R **then**
  $L \leftarrow$ premium in the last row with response R
else
  $L \leftarrow L_{\text{init}}$
end if
if seq includes L **then**
  $U \leftarrow$ premium in the first row with response L
else
  $U \leftarrow U_{\text{init}}$
end if
if aligned = True **then**
  return L, U
else
  return $-U$, $-L$
end if

Given (3), the probability of increased willingness to purchase given a categorical factor f for a typical participant in phase one can be written as

$$\begin{aligned}&\Pr(\text{increased w.t. purchase} \mid f) = \Pr(W \geq 4 \mid f)\\ &= 1 - \Pr(W \leq 3 \mid f) \overset{\text{(a)}}{=} 1 - \sigma\left(\beta_{3|4} - \gamma_f\right), \end{aligned} \tag{7}$$

where in (a), we freeze the typical participant random effect at its mean, i.e., zero. Combining (7) with (6) implies that the odds ratio of increased willingness to purchase can be written in closed-form as

$$\begin{aligned}\text{OR}^{f}_{\text{purchase},+} &= \left(\frac{1-\sigma\left(\beta_{3|4}-\gamma_f\right)}{\sigma\left(\beta_{3|4}-\gamma_f\right)}\right) \Big/ \left(\frac{1-\sigma\left(\beta_{3|4}-\gamma_{f_{\text{baseline}}}\right)}{\sigma\left(\beta_{3|4}-\gamma_{f_{\text{baseline}}}\right)}\right)\\ &\overset{\text{(b)}}{=} \frac{\exp\left(\gamma_f - \beta_{3|4}\right)}{\exp\left(\gamma_{f_{\text{baseline}}} - \beta_{3|4}\right)} \overset{\text{(c)}}{=} \exp\left(\gamma_f\right), \end{aligned} \tag{8}$$

where (b) follows from the definition of the sigmoid function, and (c) holds because the CLMM coefficient corresponding to the factor baseline is zero, i.e., $\gamma_{f_{\text{baseline}}} = 0$. This completes the proof. The proof for the `label_comparison` factor in phase two follows the same lines as above. □