"I would not install an app with this label": Privacy Label Impact on Risk Perception and Willingness to Install iOS Apps

David G. Balash, *University of Richmond;* Mir Masood Ali and Chris Kanich, *University of Illinois Chicago;* Adam J. Aviv, *The George Washington University*

https://www.usenix.org/conference/soups2024/presentation/balash

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

August 12–13, 2024 • Philadelphia, PA, USA

978-1-939133-42-7

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

"I would not install an app with this label": Privacy Label Impact on Risk Perception and Willingness to Install iOS Apps

David G. Balash
University of Richmond

Mir Masood Ali
University of Illinois Chicago

Chris Kanich
University of Illinois Chicago

Adam J. Aviv
The George Washington University

Abstract

Starting December 2020, all new and updated iOS apps must display app-based privacy labels. As the first large-scale implementation of privacy nutrition labels in a real-world setting, we aim to understand how these labels affect perceptions of app behavior. Replicating the methodology of Emani-Naeini et al. [IEEE S&P '21] in the space of IoT privacy nutrition labels, we conducted an online study in January 2023 on *Prolific* with $n = 1,505$ participants to investigate the impact of privacy labels on users' risk perception and willingness to install apps. We found that many privacy label attributes raise participants' risk perception and lower their willingness to install an app. For example, when the app privacy label indicates that *financial info* will be collected and linked to their identities, participants were 15 times more likely to report increased privacy and security risks associated with the app. Likewise, when a label shows that *sensitive info* will be collected and used for cross-app/website tracking, participants were 304 times more likely to report a decrease in their willingness to install. However, participants had difficulty understanding privacy label jargon such as "diagnostics", "identifiers", "track" and "linked". We provide recommendations for enhancing privacy label transparency, the importance of label clarity and accuracy, and how labels can impact consumer choice when suitable alternative apps are available.

1 Introduction

Smartphone applications (apps) have become a necessary part of most people's daily lives [17, 46, 47], and app marketplaces such as the Apple App Store [5] provide smartphone users the ability to quickly install a plethora of apps to meet their needs. Today's smartphones come with an impressive array of sensors, such as microphones, cameras, GPS, gyroscopes, and accelerometers. These sensors allow apps to collect more types and larger amounts of data from users of smartphones [44], increasing the privacy risks within the mobile environment [2]. Research has shown that smartphone users are concerned about their privacy when it comes to their mobile apps [4, 29, 39, 56], but are often unaware of the extent of app data collection [25, 37, 40, 48].

To help people overcome the burdens associated with reading privacy policies [20, 33, 49, 54], researchers designed privacy nutrition labels [9, 19, 22, 23, 35, 36, 38, 55, 57] to improve privacy communication and do away with natural language presentations of privacy behavior. Apple privacy labels were introduced in December 2020 [6, 13] to provide users with more transparency about the data being collected by apps [32]. The labels present users with a standardized set of information about the data being collected, such as the type of data (e.g., location, search history), the purpose of the data collection (e.g., targeted advertising, app functionality), and whether the data is linked to the user's identity [1]. These labels aim to help users make more informed decisions about which apps to use and increase trust in the app ecosystem. Labels have the potential to help users make informed choices when selecting an application to install. Therefore, it is important to understand whether privacy labels lead to better privacy outcomes for users such that users' privacy expectations align with the actual behavior of the apps they use.

Our study replicates the methodology and extends the results from Emami-Naeini et al. [23] on IoT device privacy labels into the ecosystem of Apple's iOS privacy labels, a real-world, large-scale (over a million apps) deployment of privacy labels. The methodology across both studies emphasizes comparing consumer reactions to hypothetical products/apps with differing designs and intuitive privacy implications, one with an expectation of higher privacy invasion and one with a lower expectation. Emani-Naeini et al. considered a hypothetical smart lightbulb (lower expectations) and a smart speaker (higher expectations); we compare a hypothetical note-taking app (lower) to a social media app (higher). By extending the prior IoT study to the iOS privacy label ecosystem, we provide both a point of comparison between the two settings and also how privacy labels in iOS, in particular, have the potential to affect consumer behavior.

We conducted an online survey on *Prolific* [52] in January

2023 with $n = 1,505$ participants to measure the effectiveness of privacy labels in conveying privacy risk to users, and the impact labels have on users' willingness to install an application. The survey structure was based on the methodology of Emami-Naeini et al. [23], which looked at how the proposed design for Internet of Things (IoT) labels would influence consumers' purchase decisions of IoT devices. We asked users about their experiences with the privacy labels on the App Store and how these labels impacted their app installation decision-making. These methods allowed us to answer the following research questions:

RQ1 *[App Concern] What experiences and concerns do users have with the apps they have already installed or considered installing?*

When considering social media and note-taking apps, a greater percentage of participants, 62% versus 34%, reported being at least *Somewhat concerned* with how social media apps would use, collect, and store information. Yet more participants reported they had previously installed a social media app than a note-taking app, 88% to 49%. Privacy concerns were more often cited as a reason for not installing a social media app than a note-taking app.

RQ2 *[Understanding of Privacy Labels] How do users understand the data collection information summarized on the privacy label?*

Participants generally understand the meaning of many privacy label data categories and privacy types. However, we found participants had trouble understanding some of the data categories such as *Other data*, *Diagnostics*, and *Identifiers*. There were also issues of understanding with particular jargon such as "track" and "linked", as well as confusion with the terminology such as *Contacts* versus *Contact info*.

RQ3 *[Risk and Willingness to Install] Which app privacy label attributes significantly influence user risk perception and willingness to install and in what ways?*

The *Data not collected* privacy type was the only label attribute that consistently decreased risk perception and increased willingness to install. Most attributes increased risk perception and reduced willingness to install by at least some amount. The attributes that caused the most significant increase in risk perception and decrease in willingness to install were the *Financial info*, *Sensitive info*, and *Browsing history* data categories, and the *Data used to track you* privacy type.

Participants in the study expressed dissatisfaction with the clarity of privacy labels, emphasizing that these labels often needed more detailed information about an app's data collection behavior, making it challenging to gauge its security and privacy risks. This phenomenon is termed the "transparency paradox," [51] where trying to summarize information handling practices, like through privacy labels, might necessarily omit critical details, leading to confusion and mistrust. Striking the right balance between offering summarized information and exhaustive detail is vital for informed user decisions about privacy. The study also indicated that while privacy labels can reassure users about upfront data collection, they might foster a false sense of security, leading to complacency. There is a need for more effective oversight of privacy label accuracy and consumer education on their limitations.

Furthermore, the availability of alternative apps in the marketplace can influence users' willingness to compromise on privacy. Participants in our study associated data collection with the app's purpose, where incongruences led to reduced trust in the app. Overall, while privacy and security labels have the potential to be influential in shaping user perceptions and decisions, their efficacy relies on their accuracy and completeness, necessitating further research to optimize their design and implementation for a transparent app environment.

2 Background and Related Work

Labels have been used as an effective means to communicate information to end users on products like food (Nutrition Facts) [28] and home appliances [3, 18]. Drawing inspiration from these labels, Kelly et al. [35, 36] developed a privacy label that presents how websites collect, use, and share consumers' personal information. This was later extended [38] in the design of a "Privacy Facts" label for mobile apps. The label detailed information that apps collect along with their intended use. Subsequently, Emami-Naeini et al. [22] developed and evaluated similar labels for Internet of Things (IoT) devices. Over the years, multiple researchers have studied and provided recommendations on designing similar privacy notices from a variety of perspectives [9, 19, 22, 23, 35, 36, 38, 55, 57].

To determine which privacy and security label attributes most impact consumers' risk perception and willingness to purchase Internet of Things (IoT) devices, Emami-Naeini et al. [23] designed a study to measure the effectiveness of each privacy and security attribute-value pair in isolation. This allowed the researchers to assess each attribute's impact and identify misconceptions associated with individual attributes. The study found that attribute values intended to communicate increased risk were generally perceived that way by participants. Still, the study also found risk perception did not always align with willingness to purchase the device. Furthermore, they make recommendations for improving the privacy labels, including reducing information uncertainty (purpose, harms, controls), improving information placement between primary and secondary layers, and reducing misconceptions by providing explanations to consumers.

Other work on Apple's privacy labels. Li et al. [45] interviewed 12 developers and reported their difficulty understanding labels. Gardner et al. [30] developed a tool that analyzes code and prompts developers to report data collection practices in their labels. Kollnig et al. [41] evaluated 1,759 apps before and after the introduction of Apple's App Tracking Transparency and privacy labels. They found instances of

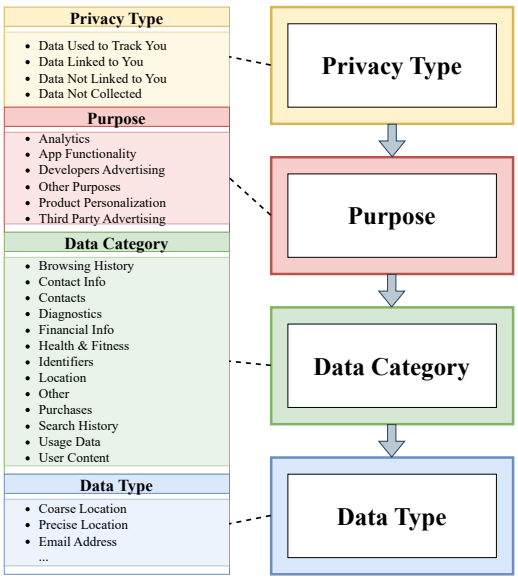


Figure 1: Hierarchical Structure of a Privacy Label.

apps violating Apple's policies and tracking users. In a longitudinal study of privacy label adoption, Balash et al. [8] analyzed weekly snapshots of the App Store for over a year and identified an increase in apps with labels and likely underreporting by developers forced to provide a label on a version update. Xiao et al. [61] analyzed data flows within 5,102 apps and found inconsistencies between app practices and reported privacy labels. Garg and Telang [31] reported a reduction in app demand following privacy label disclosures of the collection of sensitive information. To test the usability of iOS app-based privacy nutrition labels, Zhang et al. [62] conducted an interview study with lay iPhone users. They found dissatisfaction and misunderstandings that reduced the effectiveness of the label, such as confusing structure, unfamiliar terms, and lack of control over permissions settings.

Structure of Apple's Privacy Labels. Apple's privacy labels are similar in structure and content to prior work on privacy nutrition labels [38]. The label follows a hierarchical model (see Figure 1) and describes data collection practices under four levels: **(1) Privacy Type:** Describes how the collected data is handled, i.e., (a) if the data is anonymized, (b) if the data can be used to identify users, and (c) if the data is used to track users (with third parties). An app's privacy label may contain a combination of one, two, or all three of these types. A fourth, mutually exclusive privacy type indicates that the app collects no user data. **(2) Purpose:** Describes the reason for data collection, e.g., for advertising, analytics, etc. **(3) Data Category:** Presents a high-level category for collected data, e.g., *Location*, *Contact Info*, etc. **(4) Data Type:** Granular information under the Data Category, e.g., data types under *Location* can be *Precise Location* and *Coarse Location*.

3 Method

Study Procedure As previously noted, the methods of this study are replicating the work of Emami-Naeini et al. [23] from the IoT privacy nutrition label to Apple's iOS App labels. In Emami-Naeini et al.'s design for IoT, they considered a single label applied to two hypothetical devices: a smart light bulb and a smart speaker. They hypothesized that the light bulb would have low privacy implications with consumers while the smart speaker would have higher privacy implications. This helped them compare participants' associated privacy risk and willingness to purchase in different settings with different privacy expectations.

We replicated their design in the context of iOS apps. Participants viewed two hypothetical apps with different privacy expectations: a note-taking app (less privacy-invasive) and a social media app (more privacy-invasive). Like Emami-Naeini et al., we compared how each privacy label, when individually applied to different settings, affects consumers' willingness to install an app and their associated privacy risks.

Following the design of [23], we considered the hypothetical app as a between-subject factor and the privacy information displayed on the iOS privacy label as a within-subject factor. We randomly assigned each participant to answer questions about 3 of the 43 possible privacy label attributes. Forty-two privacy label attributes combine the three privacy types with the 14 data categories. The additional privacy label is a *Data Not Collected* label with no associated data categories. The *Data Not Collected* label essentially offers a comparison to an app that has no privacy labels.

We completed two pilots with co-workers to refine the questions, and we also performed a final test run ($n = 20$) on *Prolific* [52]. Participants had to be 18 years of age or older and reside in the United States. There was no requirement for participants to be smartphone users. Below, we describe the final procedure in detail, and the complete survey instrument can be found in Appendix A.

1. *Informed Consent:* Participants consented to the study, risks, benefits, and right to withdraw.
2. *App Related Questions (**Q1–Q7**):* We presented each participant with the description (see the **Notebook** app in Appendix A) of a randomly assigned hypothetical iOS application and asked them to imagine they were making an install decision. We asked about participants' concern level and install history for the app type assigned to their study condition.
3. *Privacy Label Related Questions (**Q8–Q12**):* The image of a randomly selected Apple privacy label and questions about understanding, perceived risk, and willingness to install were displayed. Each participant was shown three labels and the same set of questions for each label.
4. *Demographics Questions (**D1–D5**):* Participants were asked (optionally) to provide demographic information, such as age, identified gender, and education.

Between-Subject Factor. We considered *app type* as a between-subject factor and tested two types of iOS apps: a social media app, which we hypothesized that most participants would have privacy concerns about [26, 34, 53]; and a note-taking app, which we expect to have fewer privacy concerns. To test this hypothesis, we asked participants how concerned they were about how the app would collect, store, and use their information and to explain their reasons. If they have this app installed on their device, we asked how long they have had it and why they installed it. If they did not have it installed, we inquired whether they had considered installing it and what deterred them from doing so.

Within-Subject Factor. In our study, we included 43 privacy label attributes. We tested three Apple privacy label privacy types, *Data Used to Track You*, *Data Linked to You*, and *Data Not Linked to You*, paired with one of 14 data categories, such as *Contact info*, *Location*, and *Purchases*. In addition, we included the *Data Not Collected* privacy label, which indicates that the app will not collect user data and does not pair with any of the 14 data categories. Out of the 43 privacy label attributes (shown in Table 3), each participant answered questions about three randomly selected privacy label attributes contextualized with a hypothetical app installation scenario. The implementation precludes a participant from being randomly assigned the same privacy label attribute multiple times. Each privacy label attribute is presented in the survey as an image of the privacy label as it would appear on the Apple app store when making an installation decision for an application. To evaluate how well participants believed they understood the label, we asked them how confident they knew what the presented label meant (**Q8**).

To gauge the participants' risk perception, we asked them to specify how the presented privacy label would change the privacy and security risks they associated with the specific app in question (**Q9**-**Q10**). Afterward, we asked participants to explain the reason behind their choice. We asked similar questions to ascertain the impact of the privacy label on changing participants' willingness to install the app (**Q11**-**Q12**).

Analysis Methods. We also used the same analysis methods as [23], including a large logistic regression with repeated measures to determine the likelihood of installing (or purchasing) an app and the associated risk perception. Notably, we utilized a repeated-measures design for the within-subject factor, in which we presented participants with similar question types in multiple scenarios. Consequently, three observations for each participant were not entirely independent. We accounted for this dependence using a statistical method that included random effects. Following [23] and prior work that modeled ordinal responses [7, 12, 24, 42, 59, 60, 63], we used Cumulative Link Mixed Models (CLMMs) with logit as the link function to assess the significance of our independent variables [15, 16]. The CLMM allowed us to model all five levels of our ordinal Likert scale responses for our dependent variables: risk perception and willingness to install. We used a significance threshold of 0.05. We describe the methods in context for the remaining quantitative analysis.

For qualitative responses (five free-text questions), we utilize open coding to analyze the results of open-text questions. To achieve this, the research team's primary coder developed a codebook and identified descriptive themes for each question. Two secondary coders were responsible for coding a randomly sampled subset of 30% to ensure consistency and provided feedback on the codebook. Primary and secondary coders worked collaboratively to improve the codebook, iterating until inter-coder agreement was achieved (Cohen's $\kappa \geq 0.7$). Inter-rater reliability [50], measured with Cohen's κ ranged from 0.76 to 0.87 per question, with a mean of 0.80 (sd = 0.04). This level of agreement is "substantial" [43] or "excellent" [27].

We divided each qualitative response into two sets based on the app type assigned, note-taking, or social media. Due to the large number of responses, we used randomly sampled subsets of each free-text response. The size of the random subset (the percentage of responses for that particular question) was selected by the coders to reach thematic saturation, 20% for questions **Q2**, **Q10**, and **Q12**, 30% for **Q5**, and 65% for **Q7**.

Table 1: Demographic information of our study participants and the 2020 US Census data [10]. Categories not included in the US Census are denoted by –.

Metric	Levels	Study		Census
		n	%	%
Gender	Man	733	48.7	49.0
	Woman	724	48.1	51.0
	Non-binary	37	2.5	–
	Prefer not to disclose	10	0.7	–
	Prefer to self-describe	1	0.1	–
Age	18–24 years	324	21.5	12.9
	25–34 years	566	37.6	13.9
	35–44 years	341	22.7	12.7
	45–54 years	157	10.4	12.1
	55–64 years	83	5.5	13.0
	65+ years	31	2.1	16.8
	Prefer not to disclose	3	0.2	–
Education	No high school	23	1.5	13.9
	High school	156	10.4	26.6
	Some college	476	31.6	26.3
	Bachelor's degree	558	37.1	21.1
	Advanced degree	238	15.8	12.1
	Other	47	3.1	–
	Prefer not to disclose	7	0.5	–
Tech. Background	Yes	280	18.6	–
	No	1154	76.7	–
	Prefer not to disclose	71	4.7	–

Recruitment and Demographics. We recruited 1,505 participants via *Prolific* [52] for the survey between January 10,

2023 and January 20, 2023. Participants received $3.25 USD for completing the survey, and the median time to complete the survey was 8m, 41s. Participants were generally younger than the general population, with 21.5% between 18–24 years old, 37.6% between 25–34 years old, 22.7% between 35–44 years old, and 18% were 45 years or older. The identified gender distribution was 49% men, 48% women, and 3% non-binary, self-described, or chose not to disclose. Participant characteristics are presented in Table 1.

Limitations. As an online survey, our ability to observe real app installations and ask follow-up questions to understand the full range of responses was limited. To compensate, we used thematic coding across multiple responses to capture opinions and feelings in a hypothetical installation scenario.

We also structured our study to measure the effectiveness of each privacy type and data category pair in isolation, allowing us to evaluate the impact of each label attribute and identify any misconceptions related to individual attributes. Nevertheless, as a complete privacy label would consist of more than one attribute, additional research is required to examine the subtleties in consumers' risk perception and willingness to install when presented with a complete Apple privacy label. It is expected that the impact of each attribute will be less pronounced when viewed in the context of a complete label, and interaction effects between label attributes may arise.

Some of our results may have been affected by social desirability bias, where participants overstated their privacy concerns or intention not to install an app. These results could be viewed as a potential upper bound on true behavior.

Furthermore, we acknowledge that our recruitment sample was younger and had higher educational attainment than the population overall (see Table 1). Still, our results offer valuable insights into willingness to install and risk perception upon viewing applications and associated privacy labels. Tang et al. [58] demonstrated that gender-balanced *Prolific* studies, including questions about user perceptions and experiences, provide reliable approximations of populations' behavior.

Ethical Considerations. Our Institutional Review Board approved the study protocol. All participants were informed and consented to the study, and all collected data is only associated with random identifiers. We also reviewed each row in the dataset for potential personally identifiable information.

4 Results

This section is structured along our research questions. We first present our findings concerning participants' experiences and concerns with the apps they have installed or considered installing. Next, we show how participants understand the data collection information summarized in a privacy label. Finally, we discuss how privacy label attributes influence participants' risk perception and willingness to install an app.

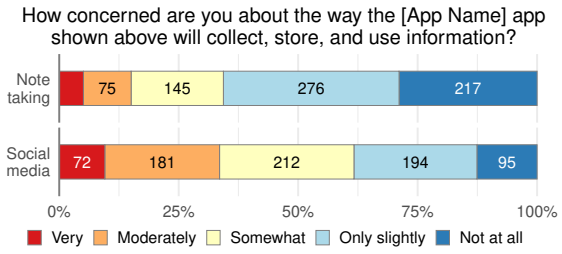


Figure 2: We asked participants to report their concern regarding the collection, storage, and use of information (**Q1**).

4.1 RQ1: App Concern

In **RQ1**, we seek to understand participants' previous experiences installing social media and note-taking apps and their preexisting concerns regarding how those applications collect, store, and use their information.

App Concern Level Participants were presented with a description of a randomly assigned generic note-taking app or a generic social media app. We then asked them to quantify their level of concern regarding the application's data collection and use (**Q1**). We hypothesized that participants assigned the social media app would report a greater level of concern than those assigned the note-taking app.

Quantitative. We found a strong correlation (Pearson's chi-square) between the app type and the level of concern. We considered the level of concern as a binary variable with *Not at all concerned* as 0, and all other concern levels as 1, $X^2(1, N = 1505) = 59.81,\ p < 0.001,\ \phi = 0.20$.

Of the participants who were assigned the note-taking app 34% reported being at least *Somewhat concerned* about how the app will collect, store, and use information. While 62% of participants assigned the social media app reported being at least *Somewhat concerned*. For full details regarding the levels of concern, please refer to Figure 2.

Qualitative. When describing their concern (**Q2**) for the note-taking application, common themes included concerns about their electronic notes being added to cloud storage or automatically synced across devices, the lack of information regarding data collection and use in the app description, and the potential for a data breach or exposure of their notes.

Common themes found when participants described their concern (**Q2**) for the social media app included unknown data collection and use policies, the reputation of social media apps for excessive data collection, sensitive information entered into the app, and data sold to third parties for targeted advertising. For instance, P805 (*Moderately concerned*) reported,

"The fact that it is free, and that social media companies are infamous for selling users' information."

Takeaway. The observed difference in participants' level of concern being greater for the social media app versus the note-taking app is consistent with our hypothesis and with previous research [26, 34, 53]. Emami-Naeini et al. [23] similarly found a strong correlation between the type of IoT device and participants' level of concern, with significant concern about smart speakers due to their always-listening capabilities.

Installation History. To understand participants' previous experiences installing an app of the type they were randomly assigned, we asked participants if they had installed an app of this type on their device (**Q3**), and if so, how long the app had been installed (**Q4**) and for what reason (**Q5**). For those who had not installed such an app, we asked if they had considered installing (**Q6**) and why they ultimately decided not to install (**Q7**). Responses to these questions allowed us to gain insight into participants' prior exposure to the apps and previous privacy concerns.

Quantitative. 88% reported having a social media app installed on their device, while only 49% reported installing a note-taking app (see Figure 5 in Appendix B). Among participants who had installed a social media app, 86% had the app for more than a year. For the note-taking app, 52% of participants indicated the app came preinstalled, and 33% had the app for over a year (see Figure 6 in Appendix B).

Among those who did not have a social media app installed, 76% reported that they had considered installing such an app. Of the participants who did not have a note-taking app installed, 38% reported that they had considered installing such an app on their device (refer to Figure 7 in Appendix B).

Qualitative. Participants most frequently reported (a) connecting with friends and family, (b) following and sharing content, (c) news and entertainment, (d) social pressure, and (e) accessing the social network on a mobile device, as reasons for installing a social media app. While the main reasons cited for installing a note-taking app included writing notes, making lists, keeping organized, setting reminders, syncing notes across devices, and storing important information.

The most common explanations for not installing a social media app included a dislike of social media, privacy concerns, too time-consuming or distracting, preferring to use a web browser to connect to the social media service, data collection concerns, mental health concerns, and concerns about sensitive data. Common reasons for not installing a note-taking app included that it did not meet current needs, that they would not use it often enough, preference for a physical notebook, and privacy concerns.

Takeaway. More participants reported they had previously installed a social media app than a note-taking app (88% to 49%). However, 52% of participants indicated that a note-taking app was preinstalled on their device. Privacy concerns were more often cited as the reason for not installing a social media app than a note-taking app.

4.2 RQ2: Understanding of Privacy Labels

Apple's privacy labels provide considerable insights into the data collection practices of an application. With this research question, we measure participants' understanding of the information presented on the label. We evaluate a Likert question (**Q8**) about participants' confidence in the meaning of the information presented on the privacy label, as well as analyzing open-response questions (**Q10**, **Q12**) for misconceptions regarding the terminology used on the label.

Confidence Level in Understanding Label Information. For all but five privacy label data categories, more than 70% of participants reported (**Q8**) being *Somewhat confident*, *Moderately confident*, or *Very confident* about knowing what the privacy label information meant. However, for the data categories *Other data*, *Diagnostics*, *Identifiers*, *User content*, and *Sensitive info*, participants' level of confidence was significantly lower (p-value < 0.001). See Figure 12 in Appendix B for a full list of the data categories and a visualization of the responses. This result strongly corresponds to Emami-Naeini et al. [23], who found that over 70% of their study participants felt confident in understanding IoT privacy labels. Furthermore, like our study, they found that confidence was significantly lower for specific privacy label attributes, in their case, *security audit* and *data linkage*.

When considering privacy types, 96% of participants reported being at least *Somewhat confident* in their understanding of the label with a *Data Not Collected* privacy type. Participants reported being at least *Somewhat confident* only 73%, 71%, and 72% respectively for the *Data Used to Track You*, *Data Linked to You*, and *Data Not Linked to You* privacy types. Refer to Figure 8 in Appendix B for the full results.

We built a Cumulative Link Mixed Model (CLMM) to understand the impact of the privacy type and data category pairs on participants' confidence levels. We used the *Data Not Collected* privacy type as the baseline privacy label attribute. We found that when the data category *Other data* was paired with privacy types *Data used to track you*, *Data linked to you*, and *Data not linked to you* it was over 49 times, 103 times, and 43 times respectively more likely to cause a participant to reduce their confidence in understanding the label by one level. We also found that when the data category *Diagnostics* was paired with privacy types *Data used to track you*, *Data linked to you* , and *Data not linked to you* it was over 19 times, 27 times, and 20 times respectively more likely to cause a participant to reduce their confidence in understanding the label by one level. See Table 2 for the full CLMM results.

Takeaway. Participants reported confidence in understanding the privacy label, except for the data categories *Other data*, *Diagnostics*, *Identifiers*, *User content*, and *Sensitive info*.

Table 2: We used CLMM and built a model to identify the significance of various factors in changing participants' confidence in the meaning of the privacy label (**Q8**). For the 14 data categories the model captures the three privacy types for each category, i.e., *Data used to track you*, *Data linked to you*, and *Data not linked to you*.

	Row	Factor	Confidence in meaning				
			$OR_{(+)}$	$OR_{(-)}$	Estimate	Std. Error	p-value
		Data category by privacy type (baseline = Data not collected)					
Data used to track you	1	Other data	0.02	48.86	-3.89	0.35	***
	2	Diagnostics	0.05	19.27	-2.96	0.34	***
	3	Identifiers	0.06	16.72	-2.82	0.33	***
	4	User content	0.08	12.69	-2.54	0.32	***
	5	Sensitive info	0.08	12.56	-2.53	0.33	***
	6	Usage data	0.15	6.73	-1.91	0.32	***
	7	Health & fitness	0.16	6.44	-1.86	0.33	***
	8	Financial info	0.18	5.53	-1.71	0.33	***
	9	Purchases	0.23	4.39	-1.48	0.33	***
	10	Contact info	0.23	4.31	-1.46	0.33	***
	11	Contacts	0.30	3.37	-1.22	0.34	***
	12	Search history	0.37	2.69	-0.99	0.33	**
	13	Location	0.67	1.49	-0.40	0.33	0.2
	14	Browsing history	0.87	1.15	-0.14	0.35	0.7
Data linked to you	15	Other data	0.01	103.00	-4.63	0.34	***
	16	Diagnostics	0.04	26.90	-3.29	0.33	***
	17	Identifiers	0.04	22.76	-3.12	0.32	***
	18	User content	0.07	14.88	-2.70	0.31	***
	19	Sensitive info	0.07	13.59	-2.61	0.31	***
	20	Purchases	0.15	6.54	-1.88	0.32	***
	21	Usage data	0.17	5.94	-1.78	0.31	***
	22	Health & fitness	0.19	5.33	-1.67	0.32	***
	23	Financial info	0.25	4.04	-1.40	0.32	***
	24	Contact info	0.27	3.68	-1.30	0.32	***
	25	Search history	0.40	2.51	-0.92	0.32	**
	26	Contacts	0.54	1.84	-0.61	0.33	0.06
	27	Browsing history	0.69	1.44	-0.37	0.33	0.27
	28	Location	0.76	1.31	-0.27	0.32	0.40
Data not linked to you	29	Other data	0.02	43.41	-3.77	0.32	***
	30	Identifiers	0.05	19.89	-2.99	0.31	***
	31	User content	0.08	12.06	-2.49	0.31	***
	32	Sensitive info	0.09	11.58	-2.45	0.30	***
	33	Diagnostics	0.10	9.60	-2.26	0.31	***
	34	Usage data	0.15	6.77	-1.91	0.30	***
	35	Health & fitness	0.17	6.02	-1.79	0.31	***
	36	Contact info	0.18	5.58	-1.72	0.31	***
	37	Financial info	0.18	5.57	-1.72	0.31	***
	38	Purchases	0.19	5.31	-1.67	0.31	***
	39	Contacts	0.21	4.72	-1.55	0.31	***
	40	Browsing history	0.29	3.51	-1.26	0.32	***
	41	Location	0.30	3.30	-1.19	0.31	***
	42	Search history	0.33	3.01	-1.10	0.31	***
		Prior labels (baseline = 0 labels)					
	43	1 label	0.88	1.14	-0.13	0.09	0.16
	44	2 labels	0.89	1.13	-0.12	0.12	0.34
		Threshold coefficients					
	45	Not at all|Slightly	0.01	175.60	-5.17	0.28	***
	46	Slightly|Somewhat	0.03	34.24	-3.53	0.27	***
	47	Somewhat|Moderately	0.13	7.93	-2.07	0.26	***
	48	Moderately|Very	0.90	1.11	-0.10	0.26	0.70
		Random effects					
	49	σ_u^2	–	–	2.86	–	–

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

Common Misconceptions. Qualitative responses revealed common misunderstandings about the terminology used on the privacy label, including the terms track, linked, *Contact info*, and *Identifiers*, among others. Participants sometimes conflated the term tracking, which Apple defines as using data to track users across apps and websites owned by other companies, to mean tracking their interactions with the device. In a response about the *Diagnostics* data category with the *Data used to track you* privacy type, P562 said, "I don't like the idea that they're tracking what sites I visit." Participants found the data collection associated with particular data categories to be unclear. For instance, P27 who was shown the *Identifiers* data category said, "Not sure what data is being collected." P292 confounded *Contact info* with their contacts when reporting, "I don't want my choices to potentially impact my contacts."

4.3 RQ3: Risk and Willingness to Install

To answer **RQ3**, we presented participants with a privacy label describing an app's data collection behavior. We then asked participants to rate the privacy and security risks on a Likert scale (**Q9**) and provide an open-ended explanation (**Q10**). Following this, we asked participants to rate, using a Likert scale (**Q11**) and an open-ended explanation (**Q12**), how the privacy label would impact their willingness to install the app.

CLMM Models. We developed two Cumulative Link Mixed Models (CLMMs) to assess how different factors influenced two dependent variables (DVs): participants' risk perception and willingness to install an iOS application (see Table 3). We included the following factors in each model:

- *Data category by privacy type*: 43 privacy label attributes consisting of three privacy types paired with the 14 data categories, and the *Data not collected* privacy type. Of the 43 attributes, only three were randomly chosen and shown to each participant, while the remaining attributes were not presented. We selected a label with the *Data not collected* privacy type as the baseline attribute as it is the one privacy type that has no associated data categories.
- *Label meaning confidence level*: The participant's confidence in the meaning of the label, with three levels: (a) *Not at all confident* or *Slightly confident*, (b) *Somewhat confident*, and (c) *Moderately confident* or *Very confident*. We used *Somewhat confident* as the baseline confidence level as it is the middle of the Likert values.
- *Application type*: We considered two levels of app type: social media and note-taking. The note-taking app was selected as the baseline because we expected its information use to be less concerning.
- *Concern about information use*: The participant's concern about the way the app will collect, store, and use information, with three levels: (a) *Not at all concerned* or *Slightly concerned*, (b) *Somewhat concerned*, and (c) *Moderately concerned* or *Very concerned*. We used *Somewhat concerned* as the baseline level of concern.
- *Prior labels*: Number of prior labels seen by that participant, with three levels: 0, 1, and 2 labels. We used zero prior labels as the baseline as it is the first level.
- *Participant age*: The age of the participant, with two levels: (a) less than 35 years old, and (b) 35 and older. We used 35 and older as the baseline age range because of its proximity to the median age of our participants.

Table 3: We used CLMM and built two models to identify the significance of various factors in changing participants' risk perception (**Q9**) and willingness to install (**Q11**). For the 14 data categories our models capture the three privacy types for each category, i.e., *Data used to track you*, *Data linked to you*, and *Data not linked to you*.

	Row	Factor	Risk perception					Willingness to install				
			$OR_{(+)}$	$OR_{(-)}$	Estimate	Std. Error	p-value	$OR_{(+)}$	$OR_{(-)}$	Estimate	Std. Error	p-value
		Data category by privacy type (baseline = Data not collected)										
Data used to track you	1	Financial info	11.43	0.09	2.44	0.33	***	0.00	641.94	-6.46	0.36	***
	2	Sensitive info	10.27	0.10	2.33	0.32	***	0.00	303.76	-5.72	0.33	***
	3	Other data	9.00	0.11	2.20	0.32	***	0.00	202.50	-5.31	0.33	***
	4	Purchases	5.96	0.17	1.79	0.31	***	0.01	122.71	-4.81	0.32	***
	5	Browsing history	5.95	0.17	1.78	0.32	***	0.01	151.97	-5.02	0.33	***
	6	Contacts	5.91	0.17	1.78	0.32	***	0.01	161.22	-5.08	0.32	***
	7	Search history	5.75	0.17	1.75	0.31	***	0.01	145.14	-4.98	0.32	***
	8	Identifiers	5.70	0.18	1.74	0.32	***	0.01	113.27	-4.73	0.32	***
	9	User content	4.41	0.23	1.48	0.31	***	0.01	94.38	-4.55	0.31	***
	10	Contact info	4.19	0.24	1.43	0.32	***	0.01	108.36	-4.69	0.32	***
	11	Location	4.18	0.24	1.43	0.31	***	0.01	90.83	-4.51	0.31	***
	12	Health & fitness	3.98	0.25	1.38	0.31	***	0.01	79.93	-4.38	0.32	***
	13	Usage data	3.55	0.28	1.27	0.30	***	0.02	54.33	-4.00	0.31	***
	14	Diagnostics	3.05	0.33	1.11	0.31	***	0.02	41.82	-3.73	0.31	***
Data linked to you	15	Financial info	14.40	0.07	2.67	0.32	***	0.00	363.33	-5.90	0.33	***
	16	Sensitive info	9.80	0.10	2.28	0.31	***	0.00	406.68	-6.01	0.33	***
	17	Browsing history	7.29	0.14	1.99	0.31	***	0.01	157.43	-5.06	0.31	***
	18	Location	6.23	0.16	1.83	0.30	***	0.01	120.71	-4.79	0.30	***
	19	Identifiers	5.51	0.18	1.71	0.31	***	0.01	102.92	-4.63	0.31	***
	20	Search history	5.08	0.20	1.62	0.30	***	0.01	193.09	-5.26	0.31	***
	21	Other data	4.66	0.21	1.54	0.31	***	0.01	106.21	-4.67	0.31	***
	22	Contacts	4.11	0.24	1.41	0.31	***	0.01	144.23	-4.97	0.32	***
	23	User content	4.00	0.25	1.39	0.30	***	0.01	98.56	-4.59	0.30	***
	24	Contact info	3.95	0.25	1.37	0.30	***	0.02	65.25	-4.18	0.31	***
	25	Usage data	3.51	0.28	1.26	0.30	***	0.03	34.65	-3.55	0.30	***
	26	Diagnostics	3.30	0.30	1.19	0.30	***	0.04	23.17	-3.14	0.30	***
	27	Health & fitness	3.19	0.31	1.16	0.30	***	0.02	50.73	-3.93	0.30	***
	28	Purchases	3.15	0.32	1.15	0.30	***	0.01	75.80	-4.33	0.30	***
Data not linked to you	29	Financial info	5.71	0.17	1.74	0.30	***	0.01	95.54	-4.56	0.31	***
	30	Sensitive info	3.64	0.27	1.29	0.29	***	0.03	37.47	-3.62	0.30	***
	31	Identifiers	3.50	0.29	1.25	0.30	***	0.05	19.66	-2.98	0.30	***
	32	Browsing history	3.22	0.31	1.17	0.30	***	0.03	38.80	-3.66	0.30	***
	33	Location	3.20	0.31	1.16	0.30	***	0.04	27.09	-3.30	0.30	***
	34	Search history	2.96	0.34	1.08	0.29	***	0.04	27.67	-3.32	0.30	***
	35	Contact info	2.82	0.35	1.04	0.30	***	0.05	21.08	-3.05	0.30	***
	36	Health & fitness	2.82	0.36	1.04	0.29	***	0.05	19.03	-2.95	0.29	***
	37	Contacts	2.77	0.36	1.02	0.29	***	0.03	36.34	-3.59	0.30	***
	38	Purchases	2.75	0.36	1.01	0.29	***	0.04	28.06	-3.33	0.29	***
	39	Other data	2.59	0.39	0.95	0.30	**	0.05	19.52	-2.97	0.30	***
	40	Usage data	2.51	0.40	0.92	0.28	**	0.09	11.30	-2.42	0.28	***
	41	User content	2.41	0.41	0.88	0.29	**	0.05	22.00	-3.09	0.29	***
	42	Diagnostics	2.09	0.48	0.74	0.29	*	0.09	11.49	-2.44	0.29	***
		Label meaning confidence (baseline = {*Somewhat*} confident)										
	43	{*Very, Moderately*} confident	1.33	0.75	0.29	0.08	***	0.79	1.27	-0.24	0.08	**
	44	{*Slightly, Not at all*} confident	1.04	0.96	0.04	0.09	0.67	0.57	1.77	-0.57	0.09	***
		App type (baseline = Note taking app)										
	45	Social media app	0.79	1.27	-0.24	0.08	**	1.77	0.56	0.57	0.079	***
		Concern about app information use (baseline = {*Somewhat*} concerned)										
	46	{*Very, Moderately*} concerned	1.05	0.95	0.05	0.11	0.64	0.70	1.43	-0.36	0.11	***
	47	{*Slightly, Not at all*} concerned	0.81	1.23	-0.21	0.10	*	1.43	0.70	0.36	0.09	***
		Prior labels (baseline = 0 labels)										
	48	1 label	0.86	1.16	-0.15	0.08	0.08	1.07	0.94	0.06	0.09	0.46
	49	2 labels	0.79	1.27	-0.24	0.11	*	1.29	0.78	0.25	0.11	*
		Participant age (baseline = {*35 - 44, 45 - 54, 55 - 64, 65 or older*} years old)										
	50	{*18-24, 25-34*} years old	1.03	0.97	0.03	0.08	0.71	1.70	0.59	0.53	0.08	***
		Threshold coefficients										
	51	Strongly decreases | Slightly decreases	0.28	3.52	-1.26	0.27	***	0.01	105.60	-4.66	0.28	***
	52	Slightly decreases | No impact	1.06	0.94	0.06	0.27	0.82	0.05	21.10	-3.05	0.27	***
	53	No impact | Slightly increases	3.53	0.28	1.26	0.28	***	0.44	2.28	-0.83	0.27	**
	54	Slightly increases | Strongly increases	17.27	0.06	2.85	0.28	***	3.42	0.29	1.23	0.26	***
		Random effects										
	55	σ_u^2	–	–	1.12	–	–	–	–	0.82	–	–

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

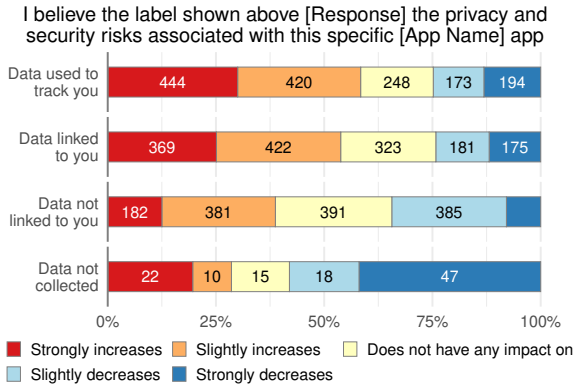


Figure 3: Participant risk perception by privacy type (**Q9**).

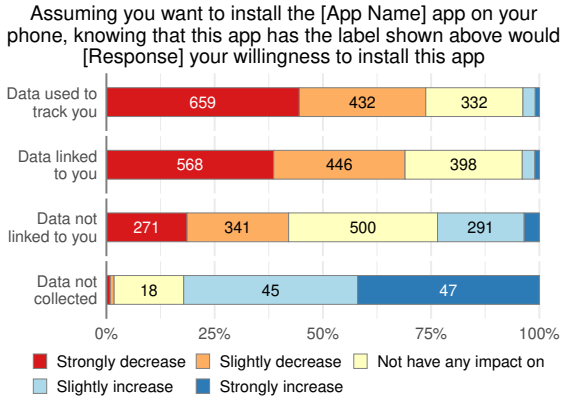


Figure 4: Participant willingness to install by label privacy type (**Q11**). *Data used to track you* caused the greatest decrease in willingness to install. While *Data not collected* increased participants' willingness to install.

We initially included additional factors such as whether the participant had installed the app, how long it had been installed, and if they considered installing the app. We also considered other demographic information, including gender, level of education, and technical background. However, the analysis revealed that these factors had little impact on the models, and so we removed these factors from our final models to improve goodness of fit, evaluated using the Akaike Information Criterion (AIC) [11]. Conversely, excluding other factors decreased model fit, so we retained all remaining factors in the models. Each model included a random intercept per participant to account for individual differences.

The CLMMs are trained on a dataset that includes three privacy label scenarios from each of the 1,505 participants, a total of 4,515 observations. Using a Likert scale, we asked participants to indicate the impact of the presented attribute, which comprised a privacy type and data category pair, on their risk perception and willingness to install the app (**Q9**, **Q11**).

In the risk perception model, a factor with a positive estimate suggests that risk perception has increased compared to the baseline for that factor. In the willingness to install model, a positive estimate indicates that participants are more inclined to install the iOS application. In contrast, a negative estimate suggests a reluctance to install compared to the baseline. In both models, all privacy type data category pairs on the privacy labels significantly affected participants' risk perception and willingness to install. Across all pairs, the impact was consistently in the direction of increased risk perception and decreased willingness to install. See Table 3 rows 1–42.

Risk Perception. According to the CLMM results, a label with the combination of *Financial info* with *Data linked to you* (Table 3, row 15) or *Financial info* with *Data used to track you* (Table 3, row 1) were the top two most significant impacts on increasing participants' risk perception. Additionally, the results indicate that a privacy label with the combination of *Sensitive info* with *Data used to track you* (Table 3, row 2) or the combination of *Sensitive info* with *Data linked to you* (Table 3, row 16) had the next two most significant impacts on increasing participants' risk perception.

The model results also show that the app type has a significant effect on the risk perception of the privacy label (Table 3, row 45). The participants assigned the social media app had a lower odds ratio, i.e., a *reduction* in privacy and security risk perception, than the participants assigned the note-taking app. This suggests that participants considered the label within the context of the app type and had a greater tolerance for or expected more data collection from a social media app.

Furthermore, the CLMM results show a decrease in the odds ratio for participants who reported being only *Slightly* or *Not at all* concerned about the way the app will collect, store, and use information (Table 3, row 47). It suggests that those participants who had less concern about the app's information use also had less concern about the data collection policy information displayed on the privacy label.

We observed a slight decrease in the odds ratios when the number of prior labels increased (Table 3, row 49), which suggests that viewing multiple privacy labels in a row for a single app causes a modest reduction in risk perception. This could be explained by participants being privacy resigned [21] or experiencing warning fatigue [14]. Participants could also be feeling lower risk compared to the previous label [23].

The CLMM results also found that the *Data not linked to you* privacy type (Table 3, rows 29–42) had lower odds ratios overall than the *Data used to track you* and *Data linked to you* privacy types (see Figure 3). Similarly, [23] found that attributes such as *data being sold to third parties* and *lack of access control* notably increased risk perception, while *no cloud retention* and *not sharing data with third parties* significantly reduced perceived risk.

Takeaway. All privacy label data categories increase risk perception. The data categories *Financial info*, *Sensitive info*,

and *Browsing History* were consistently the most likely to increase risk perception across privacy types. Participants with the social media app perceived lower risk than those with the note-taking app, contextualizing the risk by considering the app type. Participants who reported lower concern about app data collection also reported lower risk perception.

Willingness to Install. The CLMM results showed that participants' willingness to install an app was significantly negatively impacted by privacy labels combining one of *Financial info* or *Sensitive info* with one of *Data linked to you* or *Data used to track you* (Table 3, rows 1, 2, 15, 16). This reduction in willingness to install aligns with the corresponding increase in risk perception for these same data categories.

The CLMM shows that when participants reported (**Q8**) being only *Slightly* or *Not at all* confident in the meaning of the privacy label, it had significant impacts on decreasing participants' willingness to install the app (Table 3, row 44). This shows that some participants would be reluctant to install an app whose privacy label they had difficulty understanding.

The model shows that the app type makes a significant impact (Table 3, row 45). The social media app *increased* participants' willingness to install the app, i.e., positive odds ratio. This suggests that participants consider the app type together with the data collection behavior disclosed on the privacy label when making an installation decision.

We found that participants who are under 35 played a significant positive factor in the willingness to install an app (Table 3, row 50), suggesting that younger users are more willing to install the applications regardless of privacy labels.

We also found that the *Data not linked to you* privacy type (Table 3, rows 29–42) had higher odds ratios (increase in willingness to install) overall than both the *Data used to track you* and *Data linked to you* privacy types. Figure 4 shows the full results of the willingness to install by privacy type.

Comparing the odds ratios of risk perception and willingness to install (presented in Table 3), we observe that for all of the label attributes, the odds ratios of decreasing willingness to install are higher than their corresponding odds ratios of increasing risk perception. This finding suggests that the tested label attributes had a greater impact on participants' willingness to install an app than on their risk perception.

Takeaway. The CLMM showed that participants were less willing to install apps when the labels showed that *Financial info* or *Sensitive info* was tracked or linked to them. Those uncertain about the meaning of privacy labels were less inclined to install the app. Younger participants (under 35) were more lenient regarding installation. Overall, privacy labels significantly influenced installation decisions, *more* than they impacted risk perception. This contradicts [23], which found that labels were *less* influential in altering willingness to purchase an *IoT device* than in altering risk perception.

Response Category Analysis. Based on the CLMM estimates, we computed the probabilities of the five response categories for risk perception and willingness to install (Appendix B Figure 11). Participants were more likely to express increased risk perception for all label attributes except *Data not collected*. For most data categories, *Data used to track you* correlated with the highest probability for increased risk perception. However, *Data linked to you* had the highest correlation for *Browser history*, *Financial info*, and *Location*.

For most data categories, when combined with *Data used to track you* or *Data linked to you*, the highest probable response was a *Strong decrease* in their willingness to install the app. The *Diagnostics* and *Usage data* were exceptions, where the responses with the highest probability were *Slightly decrease your willingness to install* or *Not have any impact on your willingness to install* was the highest. This suggests that participants were more accepting of data collection if it was associated with improving the application or if they found the information collected less sensitive.

The *Data linked to you* privacy type had higher probabilities in the *Strongly decrease your willingness to install* response on 9 of the 14 data categories. The *Data linked to you* privacy type played a more prominent role in the reduction of participants' willingness to install an app than in risk perception, where the *Data used to track you* privacy type was the leader in increasing a participants' privacy and security risks. This suggests that tracking collected data is more highly associated with privacy and security risks to participants, whereas linking data is more of a deterrent to installing an application.

Reasons for Concern. Participants' replies to the open-ended questions provide a deeper understanding of the reasoning behind risk perceptions and willingness to install an app. Participants reported concern that the collected data was personally identifiable. They were also concerned about the collection of private or sensitive information, tracking, and unauthorized access (e.g., in a breach or through misuse). When presented with the *Identifiers* data category combined with the *Data linked to you* privacy type, P246 expressed concern that the collection was personally identifiable: "This sounds like it would be information that could be specifically linked to me and me alone." P77 responded (social media app, *Health & fitness*, *Data linked to you*), "I don't want them to know my health info." P1481 shared concerns regarding unauthorized access: "Storing personal and sensitive information in a place the user is unaware of and in a system that could be hacked could mean that information could get into the wrong hands and it's completely outside of the user's control." P1030 (social media app, *Sensitive info*, *Data used to track you*) stated, "I would not want any sensitive information shared, so I would not install an app with this label."

Participants also had common reasons for reduced concern, such as data not being linked to their identity, data not collected, limited data collection, and data categories they did not

consider sensitive. P26 shared that the *Location* data category combined with *Data not linked to you* privacy type *does not have any impact on the privacy and security risks* "because the location is not linked to my identity." P1142 said that the *Data not collected* privacy type, "means that the app developer does not collect any data, so therefore there is no risk to privacy and security because they have no information about you." P1244 (*User content*, *Data linked to you*) found the data collection to be limited, "It shows a level of transparency and also that they are taking labels seriously by doing the minimum." And P956, who was not concerned about the sensitivity of the *Contact info* data category, said, "My contact information is less of a concern than more personal data such as financial information the app may need from me."

Privacy Resigned. Some participants are resigned to data collection as the new standard for applications. P8 (social media app, *Contacts*, *Data used to track you*, *No impact*) shared this example, "Most apps already track all my information and location, so I would not be worried about one more having it." And P1091 (note-taking app, *Health & fitness*, *Data not linked to you*, *No impact*) replied, "I assume all apps and Apple products talk to one another and spy on me." While P77 (social media app, *Identifiers*, *Data used to track you*, *No impact*) added, "Other apps and companies track me on websites already." Participant P808 (social media app, *Health & fitness*, *Data used to track you*, *No impact*) simply said, "There are always privacy risks when using any kind of app."

Lack of Transparency. Some participants complained that the first level privacy label, i.e., privacy types and data categories, did not provide enough transparency about data collection and application practices to make an informed decision regarding their willingness to install the app. Many still wanted to know how their collected information would be used, why the data is collected, who would have access, precisely what data is collected, and how it is protected once obtained. Participant P515 (social media app, *Contacts*, *Data not linked to you*) said, for example, "It doesn't necessarily reassure me about how my data will be used." While P854 (note-taking app, *Other data*, *Data used to track you*) had concerns about how the data is used and shared, "It allows data to be shared to other companies and does not specify exact what it would be used for or how safe it will be." And P1254 (note-taking app, *Identifiers*, *Data linked to you*), who was concerned that the label, while reporting the data collected, did not give any indication about the risks that are incumbent part of that data collection added, "It is merely showing what is stored as data not the risks associated with it."

Lack of Trust. Some participants did not trust the app developers to adhere to the practices reported in the privacy label. For instance, P1076 (note-taking app, *Other data*) said of the *Data not linked to you* privacy type, "There is less worry about information going in the cloud because it supposedly cannot be linked back to me, but I'm still not 100% convinced that any online data can be completely un-linkable to you." P535 (social media app, *User content*, *Data not linked to you* privacy type) replied, "I feel that the wording is not that trustworthy and I feel that some data will still be linked to me in some way." P597 (note-taking app, *Usage data*, *Data not linked to you*) shared, "Software companies routinely claim privacy but have proven to be false." P1014 did not trust the *Data not collected*: "I don't believe I would trust that this claim is true." P179 who also viewed the *Data not collected* label simply added, "I assume it is lying to some extent." A lack of trust can undermine the usefulness of privacy labels.

Privacy Tradeoffs. Some participants expressed the need to trade their privacy through the data collected for the app's utility or the fact that it might be free. For example, when P956 (social media app, *Search history*, *Data used to track you*) explained, "It may be that my desire to have the app outweighs the risk until something actually happens." Furthermore, P1212, assigned the social media app, said, "Most social media apps collect data. I've come to expect it from free ones because I know they need to make money." And P662 (social media app, *Search history*, *Data used to track you*) added, "It is a small price to pay for free apps."

5 Discussion and Conclusion

Privacy Labels Across Contexts As our study replicated the methodology of an IoT privacy label study [23], we can compare the results of both studies to understand how privacy labels function in different contexts. In both studies, over 70% of participants felt confident in their understanding of the labels presented, which is a promising result for the usability of privacy labels. However, both studies found that label attributes that included technical jargon caused participant understanding to be significantly lowered, such as *security audit* and *data linkage* for IoT labels and *Other data* and *Diagnostics* for iOS labels. This result suggests that privacy label attributes should be free of technical jargon and use terminology comprehensible to a broad audience.

Both studies observed differences in participants' level of concern regarding the type of app or product under consideration. In the IoT study, there was more concern about smart speakers due to their always-listening capabilities. In our iOS study, there was more concern regarding social media apps due to their reputation for excessive data collection.

Moreover, both studies found that privacy label attributes involving tracking, linking, or selling consumer data to third parties significantly increased participants' risk perception. Furthermore, all privacy label attributes that reduced consumer data protection in the IoT study and all privacy la-

bel attributes except the *Data not collected* attribute in our iOS study increased participants' risk perception. The results demonstrate that labels can effectively be used as a privacy disclosure mechanism in a variety of contexts to communicate the risks of personal data collection, storage, and use.

Transparency Paradox Our qualitative responses show participants complaining about the vagueness of privacy labels and the lack of transparency. Participants found that the label did not provide the level of detail necessary to determine whether an app's data collection increases the security and privacy risks associated with its use. Thus, they found it difficult to decide whether or not to install the application. This leads us to believe that privacy labels suffer from the transparency paradox, the inherent conflict between transparency of textual meaning and the transparency of privacy practices [51]. Summarizing information handling practices in the form of privacy nutrition labels removes relevant details needed for people to make meaningful choices regarding their privacy. This loss of informational complexity, in turn, leads to a loss in specificity. Reducing informational complexity is a laudable goal; however, it is important to recognize that excessive summarizing of privacy information may lead to confusion and mistrust, especially among users who want to fully understand the implications of the data collection. Participants felt that the first level privacy label did not specify how their data would be used, why it was collected, who would have access, and how it would be protected. However, providing too much detail, such as through a privacy policy, can overwhelm users and deter them from reading privacy information. Prior work has suggested the use of hover text [62] or providing an info link to offer another layer of explanation. Further research can help us find a balance between granularity and effectiveness.

Balancing Comfort with Complacency Our qualitative responses revealed that people are more willing to trust an application when the privacy label provides information about what data will be collected. Providing this information upfront reassures users that the developer is not trying to collect data without their consent or knowledge. However, this can lead to complacency if users do not additionally determine whether the collected data is necessary. This suggests that privacy labels might give consumers a false sense of security, leading them to believe that data collection cannot be harmful if informed about it. This raises the critical question: Do these labels create comfort for consumers but fail to provide actual privacy? Trust in this context could be harmful if it leads to complacency or disinformation. Prior work has shown that privacy labels are often inaccurate due to a lack of oversight and developer confusion when creating labels [45]. One possible solution is establishing more effective oversight mechanisms to ensure that privacy labels are accurate and truthful. Additionally, more education is needed to help consumers understand the limitations of privacy labels and encourage them to take a more active role in protecting their data. It is crucial to balance transparency and accountability to promote informed decision-making and protect privacy.

Impact of Alternatives on Willingness to Install While consumers can find alternatives for certain apps (e.g., note-taking), others (e.g., social media) are harder to replace. Consumers' willingness to install an app that collects data they are uncomfortable sharing depends on the app's necessity and their willingness to make privacy tradeoffs. Labels provide consumers with an easy way to comparison shop for apps that align with their preferences, assuming the app fulfills their requirements. With over 1.5 million apps available, consumers can choose from multiple alternatives. However, with limited choices, users may feel forced to make a privacy tradeoff. Emami-Naeini et al. [23] found that, in a marketplace with few alternatives, labels were more influential in changing risk perception than in altering willingness to purchase, suggesting that while privacy and security are important factors, they are among several factors, including price, features, and quality, that are considered by consumers when deciding to purchase an IoT device. Our study suggests that the availability of suitable alternatives can impact users' willingness to install an app due to privacy concerns.

Data Collection in Context We found that participants were savvy when matching the category of the collected data within the context of the app. For instance, they reported being wary of the note-taking app collecting *Financial info* or *Location* data since the data seemed out of alignment with the app type. The label made them question the motives for collecting data not needed for the application's functionality, and when it does not make sense in context, the practice reduces trust in the app and the developer. Further research can study additional app contexts with label information.

Impact of Privacy Labels Our study found that labels significantly impact users' risk perception and willingness to install an app. Accurate labels have the potential to communicate risk and help consumers align their privacy expectations with real-world privacy outcomes even more than other disclosure mechanisms (e.g., privacy policies). Even when participants reported limited understanding of certain attributes (e.g., *Other data*, the labels made them question associated risks. Discomfort with unknowns, such as what data is collected, emerged as a common theme in our qualitative responses. These findings underscore the importance of labels to empower consumers to make informed decisions. However, as our study also shows, the effectiveness of these labels is contingent on their accuracy and comprehensiveness. Further research is necessary to understand how to optimize the design and implementation of privacy labels to better serve consumers and promote a more transparent app ecosystem.

References

[1] How to Use Apple's Privacy Labels for Apps, 2023. URL: https://www.consumerreports.org/privacy/how-to-use-apples-privacy-labels-for-apps-a1059836329/.

[2] Paarijaat Aditya, Bobby Bhattacharjee, Peter Druschel, Viktor Erdélyi, and Matthew Lentz. Brave new world: Privacy risks for mobile users. In *Proceedings of the ACM MobiCom workshop on Security and privacy in mobile environments*, pages 7–12, 2014.

[3] Federal Trade Commission: Consumer Advice. How To Use the EnergyGuide Label To Shop for Home Appliances. http://consumer.ftc.gov/articles/how-use-energyguide-label-shop-home-appliances, May 2021. URL: http://consumer.ftc.gov/articles/how-use-energyguide-label-shop-home-appliances.

[4] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. Your Location has been Shared 5,398 Times! A Field Study on Mobile App Privacy Nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 787–796, New York, NY, USA, April 2015. Association for Computing Machinery. doi:10.1145/2702123.2702210.

[5] Apple. Apple AppStore, May 2022. publisher: Apple. URL: https://apps.apple.com/.

[6] Apple. App Store Review Guidelines - Apple Developer. https://developer.apple.com/app-store/review/guidelines, 2023. Last Accessed: April 26, 2023.

[7] Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A. Hearst, Andrew Head, and Kyle Lo. Paper plain: Making medical research papers approachable to healthcare consumers with natural language processing. *ACM Trans. Comput.-Hum. Interact.*, 30(5), sep 2023. doi:10.1145/3589955.

[8] David G Balash, Mir Masood Ali, Xiaoyuan Wu, Chris Kanich, and Adam J Aviv. Longitudinal analysis of privacy labels in the apple app store. *arXiv preprint arXiv:2206.02658*, 2022.

[9] Rebecca Balebako, Florian Schaub, Idris Adjerid, Alessandro Acquisti, and Lorrie Cranor. The Impact of Timing on the Salience of Smartphone App Privacy Notices. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*, SPSM '15, page 63–74, New York, NY, USA, 2015. Association for Computing Machinery. doi:10.1145/2808117.2808119.

[10] US Census Bureau. Age and Sex Composition in the United States: 2020. https://www.census.gov/data/tables/2020/demo/age-and-sex/2020-age-sex-composition.html, 2020.

[11] Kenneth P Burnham and David R Anderson. Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304, 2004.

[12] Zekun Cai and Aiping Xiong. Understand users' privacy perception and decision of V2X communication in connected autonomous vehicles. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 2975–2992, Anaheim, CA, August 2023. USENIX Association. URL: https://www.usenix.org/conference/usenixsecurity23/presentation/cai-zekun.

[13] Brian X. Chen. What We Learned From Apple's New Privacy Labels. 2021. URL: https://www.nytimes.com/2021/01/27/technology/personaltech/apple-privacy-labels.html.

[14] Hanbyul Choi, Jonghwa Park, and Yoonhyuk Jung. The role of privacy fatigue in online privacy behavior. *Computers in Human Behavior*, 81:42–51, 2018.

[15] Rune Haubo B Christensen. Cumulative link models for ordinal regression with the r package ordinal. *Submitted in J. Stat. Software*, 35, 2018.

[16] Rune Haubo B Christensen. A Tutorial on fitting Cumulative Link Mixed Models with clmm2 from the ordinal Package. https://cran.ms.unimelb.edu.au/web/packages/ordinal/vignettes/clmm2_tutorial.pdf, 2019.

[17] Karen Church, Denzil Ferreira, Nikola Banovic, and Kent Lyons. Understanding the challenges of mobile phone usage data. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 504–514, 2015.

[18] European Commission. Energy-efficient products. https://ec.europa.eu/info/energy-climate-change-environment/standards-tools-and-labels/products-labelling-rules-and-requirements/energy-label-and-ecodesign/energy-efficient-products_en, 2022. Last Accessed: October 28, 2022.

[19] Lorrie Faith Cranor. Necessary But Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice. *J. on Telecomm. & High Tech. L.*, 10:273, 2012. URL: http://jthtl.org/content/articles/V10I2/JTHTLv10i2_Cranor.PDF.

[20] Lorrie Faith Cranor, Candice Hoke, Pedro Leon, and Alyssa Au. Are They Worth Reading? An In-Depth Analysis of Online Advertising Companies' Privacy Policies. SSRN Scholarly Paper ID 2418590, Social Science Research Network, Rochester, NY, March 2014. URL: https://papers.ssrn.com/abstract=2418590.

[21] Nora A. Draper. From privacy pragmatist to privacy resigned: Challenging narratives of rational choice in digital privacy debates. *Policy & Internet*, 9(2):232–251, 2017.

[22] 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, San Jose, CA, USA, May 2020. IEEE. ISSN: 2375-1207. doi:10.1109/SP40000.2020.00043.

[23] 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 519–536, San Francisco, CA, USA, May 2021. IEEE. URL: https://ieeexplore.ieee.org/document/9519463/, doi:10.1109/SP40001.2021.00112.

[24] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. Are consumers willing to pay for security and privacy of IoT devices? In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 1505–1522, Anaheim, CA, August 2023. USENIX Association. URL: https://www.usenix.org/conference/usenixsecurity23/presentation/emami-naeini.

[25] Adrienne Porter Felt, Serge Egelman, and David Wagner. I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. In *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices*, pages 33–44, 2012.

[26] Elizabeth Fife and Juan Orjuela. The privacy calculus: Mobile apps and user perceptions of privacy and security. *International Journal of Engineering Business Management*, 4(Godište 2012):4–11, 2012.

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

[28] FDA Center for Devices and Radiological Health. Device Labeling. https://www.fda.gov/medical-devices/overview-device-regulation/device-labeling, October 2020. URL: https://www.fda.gov/medical-devices/overview-device-regulation/device-labeling.

[29] Marco Furini, Silvia Mirri, Manuela Montangero, and Catia Prandi. Privacy perception when using smartphone applications. *Mobile Networks and Applications*, 25:1055–1061, 2020.

[30] Jack Gardner, Yuanyuan Feng, Kayla Reiman, Zhi Lin, Akshath Jain, and Norman Sadeh. Helping mobile application developers create accurate privacy labels. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 212–230, 2022. doi:10.1109/EuroSPW55150.2022.00028.

[31] Rajiv Garg and Rahul Telang. Impact of app privacy label disclosure on demand: An empirical analysis. *Workshop on the Economics of Information Security (WEIS)*, 2022.

[32] Apple Inc. App Privacy Details - App Store, 2020. URL: https://developer.apple.com/app-store/app-privacy-details/.

[33] Carlos Jensen and Colin Potts. Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 471–478, 2004.

[34] Mohsen Jozani, Emmanuel Ayaburi, Myung Ko, and Kim-Kwang Raymond Choo. Privacy concerns and benefits of engagement with social media-enabled apps: A privacy calculus perspective. *Computers in Human Behavior*, 107:106260, 2020.

[35] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A "Nutrition Label" for Privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, pages 1–12, New York, NY, USA, July 2009. Association for Computing Machinery. doi:10.1145/1572532.1572538.

[36] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1573–1582, New York, NY, USA, April 2010. Association for Computing Machinery. doi:10.1145/1753326.1753561.

[37] Patrick Gage Kelley, Sunny Consolvo, Lorrie Faith Cranor, Jaeyeon Jung, Norman Sadeh, and David Wetherall. A conundrum of permissions: installing applications on an android smartphone. In *Financial Cryptography and Data Security: FC 2012 Workshops, USEC and WECSR 2012, Kralendijk, Bonaire, March 2, 2012, Revised Selected Papers 16*, pages 68–79. Springer, 2012.

[38] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3393–3402, Paris France, April 2013. ACM. URL: https://dl.acm.org/doi/10.1145/2470654.2466466, doi:10.1145/2470654.2466466.

[39] Hammad Khalid, Emad Shihab, Meiyappan Nagappan, and Ahmed E Hassan. What do mobile app users complain about? *IEEE software*, 32(3):70–77, 2014.

[40] Jennifer King. How come i'm allowing strangers to go through my phone? smartphones and privacy expectations. *Smartphones and Privacy Expectations.(March 15, 2012)*, 2012.

[41] Konrad Kollnig, Anastasia Shuba, Max Van Kleek, Reuben Binns, and Nigel Shadbolt. Goodbye Tracking? Impact of iOS App Tracking Transparency and Privacy Labels. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, Virtual Conference, April 2022. Association for Computing Machinery. arXiv: 2204.03556. URL: http://arxiv.org/abs/2204.03556, doi:10.1145/3531146.3533116.

[42] Marvin Kowalewski, Christine Utz, Martin Degeling, Theodor Schnitzler, Franziska Herbert, Leonie Schaewitz, Florian M. Farke, Steffen Becker, and Markus Dürmuth. 52 weeks later: Attitudes towards covid-19 apps for different purposes over time. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW2), oct 2023. doi:10.1145/3610042.

[43] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, 33(1):159–174, 1977.

[44] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. A survey of mobile phone sensing. *IEEE Communications magazine*, 48(9):140–150, 2010.

[45] Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. Understanding Challenges for Developers to Create Accurate Privacy Nutrition Labels. In *CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3491102.3502012.

[46] Tong Li, Yong Li, Tong Xia, and Pan Hui. Finding spatiotemporal patterns of mobile application usage. *IEEE Transactions on Network Science and Engineering*, 2021.

[47] Tong Li, Tong Xia, Huandong Wang, Zhen Tu, Sasu Tarkoma, Zhu Han, and Pan Hui. Smartphone app usage analysis: datasets, methods, and applications. *IEEE Communications Surveys & Tutorials*, 2022.

[48] Jialiu Lin, Shahriyar Amini, Jason I Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, pages 501–510, 2012.

[49] Aleecia M. McDonald and Lorrie Faith Cranor. The Cost of Reading Privacy Policies. *HeinOnline*, 4(3):543–568, 2009. URL: https://heinonline.org/HOL/P?h=hein.journals/isjlpsoc4&i=563.

[50] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), nov 2019. doi:10.1145/3359174.

[51] Helen Nissenbaum. A Contextual Approach to Privacy Online. *Daedalus*, 140(4):32–48, 2011.

[52] Prolific, Academic Ltd. A Higher Standard of Online Research, December 2022. https://www.prolific.co, as of June 10, 2024.

[53] Li Qin, Yongbeom Kim, and Xin Tan. Understanding the intention of using mobile social networking apps across cultures. *International Journal of Human–Computer Interaction*, 34(12):1183–1193, 2018.

[54] Joel R Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T Graves, Fei Liu, Aleecia McDonald, Thomas B Norton, and Rohan Ramanath. Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding. *Berkeley Tech. LJ*, 30:39, 2015.

[55] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A Design Space for Effective Privacy Notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 1–17, Ottawa, Canada, July 2015. USENIX Association. URL: https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub.

[56] Irina Shklovski, Scott D Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2347–2356, 2014.

[57] FTC Staff. Protecting consumer privacy in an era of rapid change–a proposed framework for businesses and policymakers. *Journal of Privacy and Confidentiality*, 3(1), 2011. URL: https://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers.

[58] Jenny Tang, Eleanor Birrell, and Ada Lerner. Replication: How Well Do My Results Generalize Now? The External Validity of Online Privacy and Security Surveys. In *Proc. SOUPS '22*, pages 367–385, Boston, Massachusetts, USA, August 2022. USENIX Association.

[59] Christine Utz, Steffen Becker, Theodor Schnitzler, Florian M. Farke, Franziska Herbert, Leonie Schaewitz, Martin Degeling, and Markus Dürmuth. Apps against the spread: Privacy implications and user acceptance of covid-19-related smartphone apps on three continents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3411764.3445517.

[60] Zixin Wang, Danny Yuxing Huang, and Yaxing Yao. Exploring tenants' preferences of privacy negotiation in airbnb. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 535–551, Anaheim, CA, August 2023. USENIX Association. URL: https://www.usenix.org/conference/usenixsecurity23/presentation/wang-zixin.

[61] Yue Xiao, Zhengyi Li, Yue Qin, Xiaolong Bai, Jiale Guan, Xiaojing Liao, and Luyi Xing. Lalaine: Measuring and characterizing Non-Compliance of apple privacy labels. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 1091–1108, Anaheim, CA, August 2023. USENIX Association. URL: https://www.usenix.org/conference/usenixsecurity23/presentation/xiao-yue.

[62] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. How Usable Are iOS App Privacy Labels? *Proceedings on Privacy Enhancing Technologies*, 4:204–228, 2022.

[63] Yuhang Zhao, Yaxing Yao, Jiaru Fu, and Nihan Zhou. "If sighted people know, i should be able to know:" privacy perceptions of bystanders with visual impairments around camera-based technology. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 4661–4678, Anaheim, CA, August 2023. USENIX Association. URL: https://www.usenix.org/conference/usenixsecurity23/presentation/zhao-yuhang.

A Survey Instrument

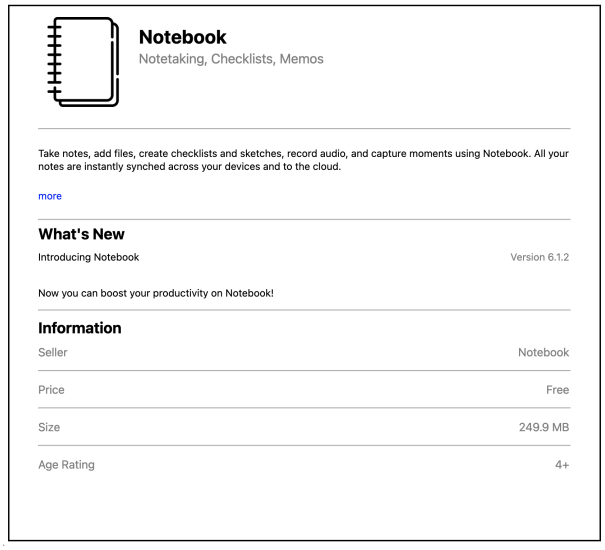
Thank you for your interest in our survey.
Please read the following instructions carefully: (i) Take your time in reading and answering the questions. (ii) Answer the questions as accurately as possible.
Definitions: (i) App: In this survey the word "app" refers to an application found on the Apple App Store that can be installed on your Apple device. (ii) Privacy Label: a short summary of an app's data collection behavior displayed on the application pages of the Apple App Store.
On the next page we will provide an introduction to this survey.
[A horizontal rule, like below, indicates a new page in the questionnaire.]

Survey Introduction

This survey is designed to investigate your awareness of app privacy labels displayed on the application pages of the Apple App Store. You will answer questions regarding potential app installation decisions and how an app privacy label may impact your thoughts about the app.
On the following pages you will be presented with an application and asked questions about this application and its privacy labels. For each of the labels we will ask a set of similar questions, so please pay close attention.

App Related Questions

[Apps are randomly assigned.]
Imagine you are making a decision to install a *[App Name]* app on your phone that was recommended by a friend. The price of the app is within your budget (or it is free) and the features are what you would expect from a *[App Name]* app.
Assume you do not have a *[App Name]* app installed on your phone.
Please review the app description before answering the questions.



[An example image of a note taking app displayed to participants.]

Q1 How concerned are you about the way the *[App Name]* app shown above will collect, store, and use information?
- ◯ Not at all concerned
- ◯ Slightly concerned
- ◯ Somewhat concerned
- ◯ Moderately concerned
- ◯ Very concerned

Q2 What about data collection, storage, and use by the *[App Name]* app makes you feel concerned?
- • Please type in your answer.

Q3 Do you currently have a *[App Name]* app installed on your phone?

◯ Yes ◯ No

*[Included only if Yes selected in **Q3**.]*

Q4 How long have you had this *[App Name]* app? If you have more than one device, answer the question for the one that you have had for the longest time.

◯ Less than a month ◯ More than a year
◯ Between a month and a year ◯ I don't remember

Q5 What were your reasons to install the *[App Name]* app?

- Please type in your answer.

*[Included only if No selected in **Q3**.]*

Q6 Have you ever considered installing a *[App Name]* app on your phone?

◯ Yes ◯ No

*[Included only if Yes selected in **Q6**.]*

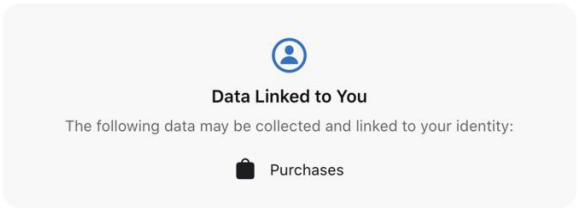
Q7 What made you decide not to install the *[App Name]* app?

- Please type in your answer.

Privacy Label Related Questions

[Q8 - Q12 will be asked once per privacy label. The privacy labels are chosen randomly. Participants are shown and asked to respond to three privacy labels.]

Please imagine the following privacy label (a short summary of the app's data collection behavior) was shown on the App Store page of the app when answering the questions below.



[An example image of a privacy label displayed to participants.]

Q8 How confident are you that you know what the label shown above means?

◯ Not at all confident ◯ Moderately confident
◯ Slightly confident ◯ Very confident
◯ Somewhat confident

Q9 I believe the label shown above

◯ Strongly decreases the privacy and security risks associated with this specific *[App Name]* app
◯ Slightly decreases the privacy and security risks associated with this specific *[App Name]* app
◯ Does not have any impact on the privacy and security risks associated with this specific *[App Name]* app
◯ Slightly increases the privacy and security risks associated with this specific *[App Name]* app
◯ Strongly increases the privacy and security risks associated with this specific *[App Name]* app

Q10 Please explain why you believe the label (decreases/increases/does not have any impact on) the privacy and security risks associated with this specific app

- Please type in your answer.

Q11 Assuming you want to install the *[App Name]* app on your phone, knowing that this app has the label shown above would

◯ Strongly decrease your willingness to install this app.
◯ Slightly decrease your willingness to install this app.
◯ Not have any impact on your willingness to install this app.
◯ Slightly increase your willingness to install this app.
◯ Strongly increase your willingness to install this app.

Q12 Please explain why knowing that this app has the label (decreases/increases/does not have any impact on) your willingness to install *[App Name]*

- Please type in your answer.

Demographic Questions

D1 What is your gender?

◯ Woman ◯ Prefer not to disclose
◯ Man ◯ Prefer to self-describe
◯ Non-binary

D2 What is your age?

◯ 18 – 24 ◯ 45 – 54 ◯ Prefer not to disclose
◯ 25 – 34 ◯ 55 – 64
◯ 35 – 44 ◯ 65 or older

D3 Are you a student?

◯ Yes ◯ Prefer not to disclose
◯ No

D4 What is the highest degree or level of school you have completed?

◯ No schooling completed
◯ Some high school, no diploma
◯ High school graduate, diploma, or equivalent
◯ Some college credit, no degree
◯ Trade / technical / vocational training
◯ Associate degree
◯ Bachelor's degree
◯ Master's degree
◯ Professional degree (e. g., J.D., M.D.)
◯ Doctorate degree
◯ Prefer not to disclose
◯ Other (please specify)

D5 Which of the following best describes your educational background or job field?

◯ I have an education in, or work in, the field of computer science, computer engineering or IT.
◯ I do not have an education in, nor do I work in, the field of computer science, computer engineering or IT.
◯ Prefer not to disclose

B Additional Figures and Tables

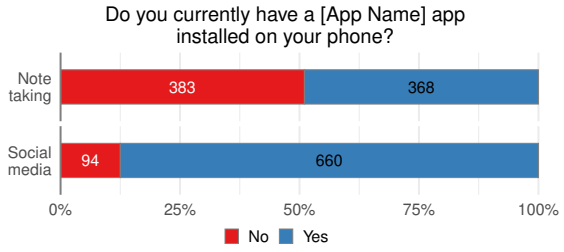


Figure 5: We asked participants if they had an app of this type already installed on their mobile device (**Q3**).

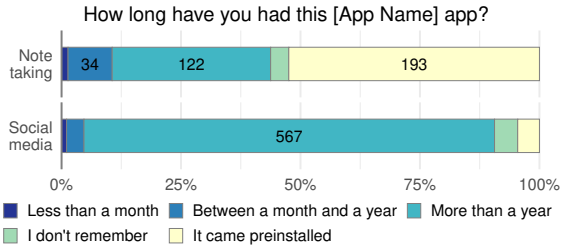


Figure 6: We asked participants how long the app was installed on their mobile device (**Q4**).

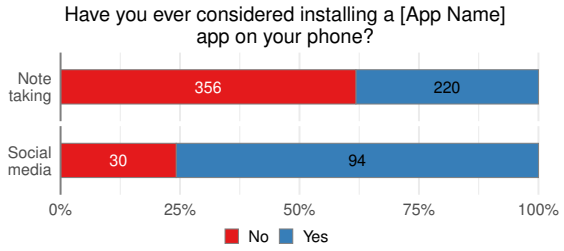


Figure 7: We asked participants if they had ever considered installing an app of this type on their mobile device (**Q6**).

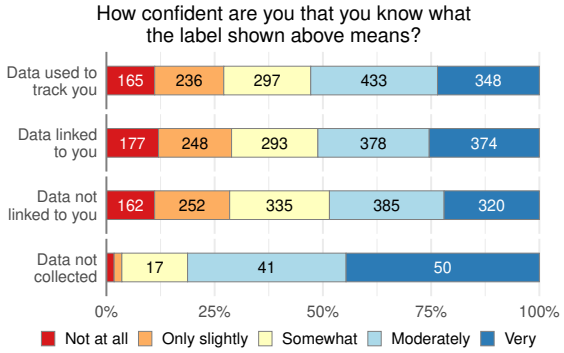


Figure 8: Confidence in label meaning by label privacy type (**Q8**).

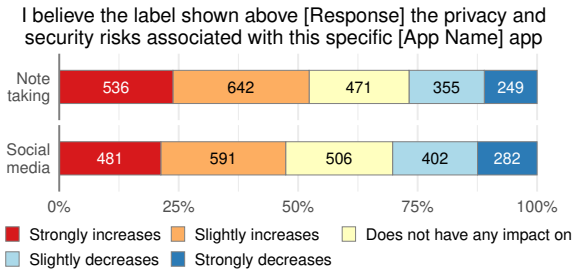


Figure 9: Participant risk perception by application type (**Q9**).

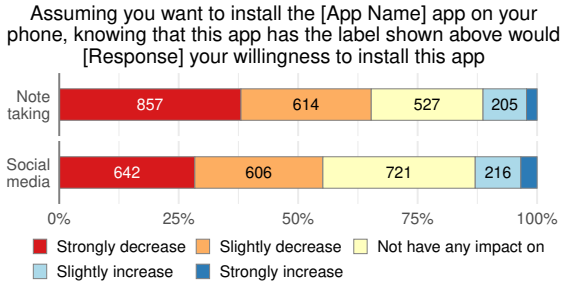


Figure 10: Participant willingness to install by application type (**Q11**).

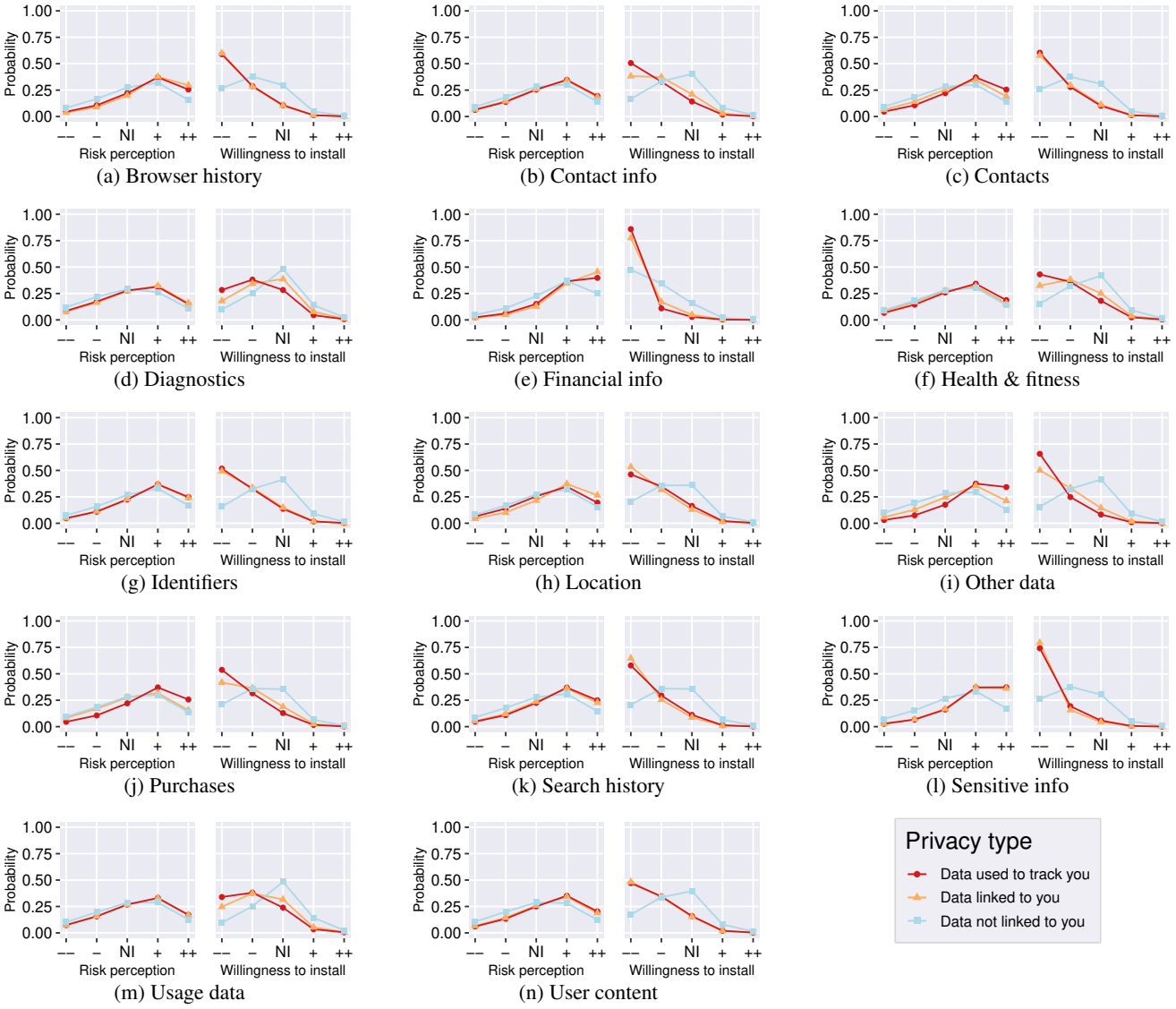


Figure 11: Based on the CLMM parameters, we computed and plotted the probabilities of each data category increasing, decreasing, or having no impact on risk perception (left plot) and willingness to install (right plot). We use the following notation to label the x axes: $--$ is *strongly decrease*, $-$ is *slightly decrease*, NI is *no impact*, $+$ is *slightly increase*, and $++$ is *strongly increase*. For most data categories, the *Slightly increases the privacy and security risks* was the highest probability of the five response categories for risk perception. The exception was the *Financial info* and *Sensitive info* data categories when combined with the *Data used to track you* or *Data linked to you* privacy types, in which case *Strongly increases the privacy and security risks* was the highest probability.

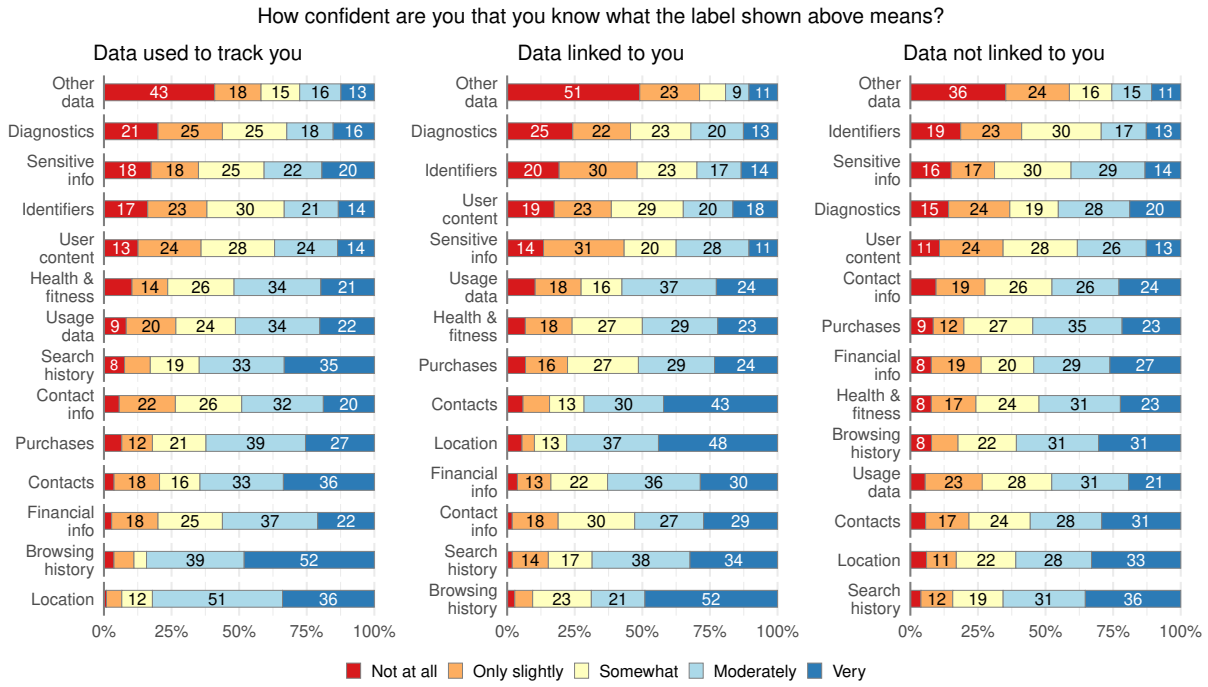


Figure 12: We asked participants how confident they were that they knew what the label shown means (**Q8**).

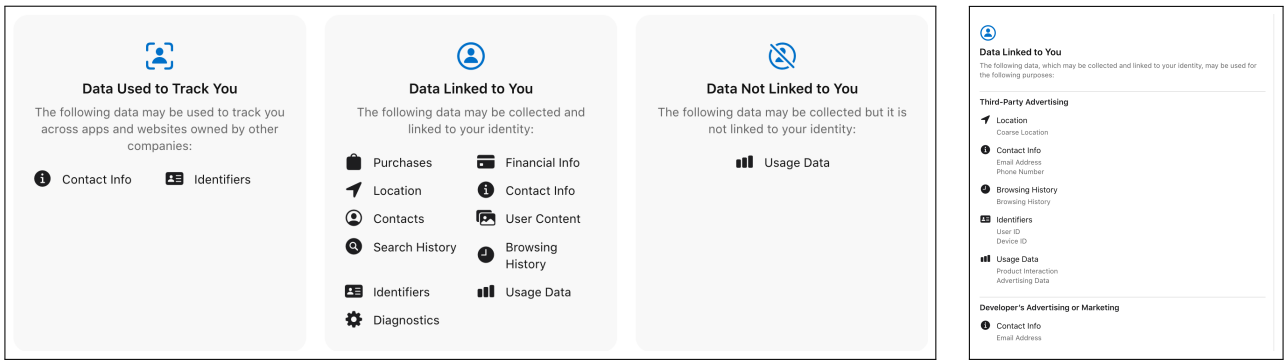


Figure 13: *(left)* An illustrative example of a privacy label from the Apple App Store, and *(right)* an illustrative example of the privacy label details from the Apple App Store. The details display the *Purpose* for the data collection and the detailed information about the *Data Types* collected.