Evaluating the Effect of the Apple Privacy Nutrition Label on App Download Decision Making

Amanda Crawford, *CMU* Denise Tang, *CMU*
Ananya Bhat, *CMU* Evan Zeng, *CMU* Jimmy Ray, *CMU* Mira Mookerjee, *CMU*

Abstract

Data privacy is important for any user, including those who use mobile devices. The goal of an app privacy label is to inform users about the privacy policies of a specific app. Apple's recently released Privacy Nutrition Labels, found in the app store, are a step in the right direction for informing users about the privacy practices of many apps. However, little research has been performed into evaluating its transparency and effectiveness. In this paper, we investigate the new nutrition labels and attempt to illuminate points of confusions in these apps. We also investigate the app decision-making process of people and if the new privacy labels affect the privacy awareness of participants. We find that while some participants have issues with understanding certain parts of the labels, the overall consensus is that the information presented in the label is helpful to our participants.

1. Introduction

The main goal of this study is to answer the question: what is the impact of Apple Privacy labels on user understanding and awareness of app privacy practices and on user decisions related to using apps? Some sub-questions that are addressed in this study include: what information do people pay most attention to when downloading an app, how do users use this information when making decisions about apps, and do the privacy labels change privacy awareness? The end goal of this research is to understand how helpful the Apple Privacy Labels are for consumers and to find ways to make Apple Privacy Labels more usable for consumers.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2021. August 8 -- 10, 2021, Vancouver, B.C., Canada.

2. Related Work

2.1. Privacy Policies

Privacy policies are a crucial tool in informing a user's understanding and consent to the various data practices of companies they interact with. Experts believe the top harms of sharing smartphone data for consumers are social problems, financial harms, surveillance & monitoring as well as privacy concerns [3]. These privacy concerns are believed to be mitigated by educating users of privacy and security information, encouraging users to only download trusted apps, and increasing transparency by application and operating system developers, as well as other techniques [3]. While user education is highly encouraged, Balebako et al. [3] believe that comprehension of privacy notices is a key factor for users actually using notice information. They found that if the information is presented too late, consumers will not be affected by the information unless they are browsing multiple places for the same product [3]. However, if privacy information is presented before a website is chosen for sensitive product purchasing, then most consumers will choose a higher privacy rated website even if they must pay more for the same product [3].

Currently, privacy policies presented on the Internet have no required standardized format, and instead usually exist as full-text, natural language policies [9]. However, full-text, natural language formats perform poorly in allowing users to quickly and accurately find information about the policies and are also disliked by many users [9]. The PPChecker, a tool created by researchers to identify problems in privacy policies on the Google Play Store, found that 74% of 2500 analyzed policies were problematic in some way [7]. Layered text policies have become more popular, and are somewhat more enjoyable for users, but even they do not fare much better than the full text policies in terms of comprehension [4,9]. In fact, layered text policies may even reduce transparency and effectively obscure important information because users rarely click through to the full policy to find information that is not immediately available in the layered format [4]. Standardized formats of privacy policies, both in tabular presentation and simple textual presentation, perform much better in both user enjoyability and comprehension of the policies [9]. However, the flexibility of standardized formats allows different policy authors to

present the same information in various ways, which may undermine their expected effectiveness [4]. Still, research from Balasubramanian et. al. [2] supports the idea that highly motivated and less knowledgeable consumers benefit the most from nutrition-style labels, and that simplified labels lead to greater understanding.

2.2. Decision Factors

In 2013, Kelley, Cranor, and Sadeh, designed a privacy facts checklist [10]. By testing the new design in lab studies and MTurk surveys, it is found that users' decisions could be affected by the way privacy information is displayed. Both lab and online participants also reported they are "aware of the (privacy) display but did not look at it". Other factors considered more or as important include: ratings, user reviews, price, branding and design, word of mouth, number of downloads, popularity, permissions, size of the app, developer/company, and advertising. Our work explores what factors affect users' privacy decisions and provides some insight confirming the presence of all of these factors.

2.3. Trust and Misconceptions

There are a lot of myths and misconceptions when it comes to privacy labels. In 2013, Kelley, Cranor, and Sadeh suggested participants "assume that all applications collect the same information" [10]. Our work investigates if the new nutrition label helps people compare apps more easily and how they do it.

Kelley, Cranor, and Sadeh also found that "users overwhelmingly trust the application's permission and privacy facts display" [10]. They assume the information is being verified and when it breaks their expectations, they assume it's their own mistake in understanding. In our study, we ask whether participants trust the information they see presented in the Apple App Store.

In Lin et al. 's experiment [11], participants are paid to read the privacy notice. But even in an optimal situation where people do read the privacy notices, it is hard for the majority of the participants to guess the purpose of why their data is being collected. Having used certain apps before doesn't help with a better understanding either.

2.4. Apple Privacy Label

In December 2020, Apple came out with a new feature which they call a "privacy label" that aims to inform users of the data practice and privacy policies of the apps that they download and interact with. In theory, these labels address the "Catch-22 implication" that comes with the common internet practice of assuming that access implies consent, as they aim to inform a user's consent to the app's privacy practices before they actually download and interact with the app [1]. The labels resemble the prototypes used in research done on standardized and tabular formats of privacy policies [9], sometimes referred to as a privacy "nutrition label". Past research has been conducted on prototypes of standardized formats and privacy nutrition labels. For example, Naeini et. al [5] studied a prototype of a privacy nutrition label for IoT devices and found that users liked how the IoT devices were given ratings from independent research labs. They also found that users would like to see these labels at the time of purchase to help the user make more informed decisions [5]. Kelley et. al [7] went through an iterative process of designing a privacy nutrition label, starting with a P3P Expandable grid and refining based on feedback several times unless they came up with their final design. This final design was discussed in a focus group and compared with the standard privacy policy in a within-groups user study, revealing even more potential for improvement [7]. However, little research has been done on the usability and understandability of Apple's labels.

Our study aims to address this gap and identify the impact of Apple Privacy Nutrition labels on user understanding and awareness of app privacy practices, and on user decisions related to using apps. We then compared the findings from past research on privacy nutrition label prototypes with our own findings to note any parallels or divergences. Because privacy labels make use of a defined set of terminology, our usability and understandability research may also help to determine which industry terms they are familiar with, which was a subject that earlier researchers encouraged more investigation into [4]. This paper also extends the work of Balebako et al. [3] by evaluating whether Apple's privacy labels are an effective notice and comprehensible by smartphone users. Past studies also suggest participants "assumed that all applications collect the same information" [8]. Our work investigates if the new nutrition label helps people compare apps easier.

This research will help designers and developers understand flaws in the current Apple label, identify improvements that could be made to the label to increase user understanding, and devise new formats that can be used as the industry standard for privacy policies.

3. Methodology

3.1. Recruitment

We originally tried recruiting participants on CMU's CBDR but had to turn to a convenience sample of colleagues because of timing constraints. Since the privacy nutrition label is a feature of the newest operating systems of Apple, we recruited 13 participants to be users of either iPhones (iOS 14) or Macs (OS Big Sur). Participants were at least 18, and selected for a balanced demographic with 38% male, 62% female, and the age range of the participants was between 18-59. All participants completed a virtual lab study that

took about an hour, and we paid $15.00 to each participant who completed the study.

3.2. Virtual Lab Study

3.2.1. Pre-Task Survey

We asked the participants to answer a brief questionnaire at the beginning of the lab study. This pre-survey asks the participant to indicate the importance of different factors, all on a Likert scale. Both security and non-security factors are included, so as to not prime the participants' answers towards security. The same survey is administered post-interview to see if the interview had any effect on how they rate the importance of app security, data privacy, and the nutrition label.

3.2.2. Task

We asked the participants to evaluate four apps in total, one at a time. The apps are selected from a list of eight apps that are pre-selected to be unknown so as to not bias responses with brand loyalty. All app evaluations were performed as "Think-Aloud tasks", where the participants vocalized their stream of consciousness as they evaluated the apps, resulting in them speaking aloud about what they noticed and about what was affecting their decision. For the first task, the participants were asked to evaluate two apps from the same category individually, and then compare them and indicate which they would pick and why. The privacy label was not mentioned, so that we could observe whether participants noticed it on their own. The second task was similar to the first, differing in only two ways: the two apps were selected from a different category, and participants were asked to evaluate the privacy label as well. This ensured that we were able to see how participants evaluated apps before being prompted to think of privacy and security.

The apps were selected on several requirements. All of the apps had to have less than 5,000 ratings and reviews, the apps couldn't be well known, and all of them had to have a similar average rating so that a participant wouldn't just choose a specific app because of bad ratings. The apps and categories used can be found in the appendix.

3.2.3. Scenario

We asked participants to pick one app from each pair for their friend after evaluating each pair. We asked this to observe if participants would prioritize an app with a different level of security if it were intended for someone they cared about. Specifically, we hypothesized that participants would want more secure apps for friends they care about than for themselves.

3.2.2. Post-Task Survey

For the final part of the virtual lab study, we asked the participants to answer the same survey as they did at the beginning of the study. No parts of the survey were changed from the original. The objective of the post-task survey was to evaluate if privacy attitudes changed after interacting with the privacy labels.

3.3. Data Analysis

Our data analysis consisted of different methods for the qualitative interview data and the quantitative survey data to see what patterns emerged before, during, and after the interview process.

3.3.1. Quantitative Data

To analyze the Likert-scale questions, we used two different methods. First, we scored each questionnaire and ran a paired-sample t-test on the total score for the pre- and post-study questionnaire score for each participant. The answers were converted to a 1-5 scale with extremely unimportant being equal to 1 and extremely important equaling 5. The question about "App Privacy Label" importance was excluded from this analysis because of too many people answering "I don't know".

Second, we used the Wilcoxon Signed Ranks test to check for differences immediately before and after the lab study for each of five app decision factors (ratings, reviews, app store rank, security, data privacy). This non-parametric test values the null hypothesis that there is no difference before and after the lab study and the alternative hypothesis that the importance rating of the app decision factor changes.

In addition to the main analysis, we described the quantitative data of the participant's rating on "App Privacy Label" because the sample size is too small to run a statistical analysis. The sample size is small because we excluded the number of participants who answered, "I don't know what it is", which cannot be converted to a numerical number. However, we felt that this question deserved its own analysis because of the number of people who originally answered that they didn't know what the label is who then went on to rank it after the interview.

3.3.2. Qualitative Data

Interview transcripts were analyzed using inductive/emergent coding. An initial codebook was created based on early interviews, and all six researchers collaborated iteratively to improve the codebook throughout the coding process. 12 interviews were double coded to ensure inter-rater reliability, and all coding discrepancies in these interviews were discussed and resolved. Due to time constraints, 3 of the 15 interviews were coded independently (each by different researchers); however, the researchers met afterwards to discuss any perceived ambiguities in the coding of particular data points, as well as any necessary changes or additions to the codebook. The final codebook contained 90 codes across 6 categories.

3.3.2. Demographics

With our 13 participants, we interviewed eight users who were female and five users who were male. Of these participants, nine of them were between 18-29 years old, one was 30-39 years old, one was 40-49 years old, and two were 50-59 years old. All of our participants have completed undergraduate degrees. Six participants are current graduate students, and then we had a wide range of professions for the rest: a nurse, analyst, program manager, software consultant, Chief Information Security Officer, and a biology research scientist.

4. Results

4.1. Quantitative Data

The survey where participants ranked what factors are important to them when downloading an app showed that Ratings was the most important factor prior to the survey out of Ratings, Reviews, App Store Rank, Security, Data Privacy, and the Privacy Nutrition Label. Before the interview 10 participants said that ratings were important to them, and the second most important factor reported was data privacy with nine participants ranking it as important or extremely important.

The post-interview showed that eleven participants now ranked Ratings as important or extremely important which was the highest scoring category. Reviews was the second highest category with 10 participants now categorizing it as an important factor. There was no change in the number of people who rated Data Privacy and Security in the important categories, but fewer people rated both categories as extremely important after the interview.

As for the Privacy Nutrition Label, seven participants reported that they didn't know what it was before the lab study. Of the other six participants, only two rated the privacy label as important before the study. The follow up survey showed that only two participants reported not knowing what the label was and two participants rated it as unimportant. There were five participants who reported that the label was important to them in their app download decision making process in the post-interview survey.

While there were individual differences in overall answer distribution as seen in figures A and B, analysis with the paired-sample t test showed that there was no significant difference between the participant scores with the scores only differing by .6 of a point ($t = -0.96777$, $df = 12$, p-value $= 0.3523$).

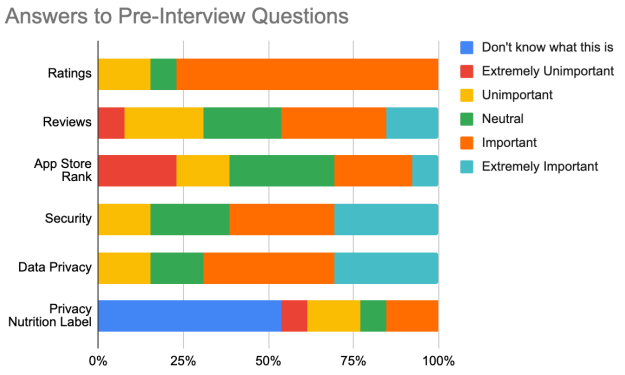


Figure A: Pre-Interview questionnaire answers showing over 50% of participants did not know what the Privacy Nutrition Label is.

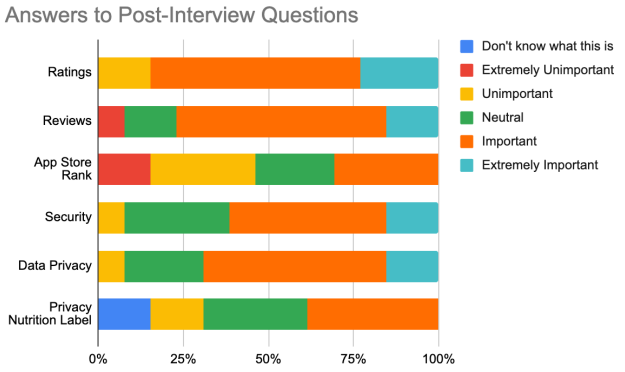


Figure B: Post-Interview questionnaire answers showing most participants who did not know what the nutrition label is now rate it neutral or important.

When using the Wilcoxon Signed Ranks test, we did not find significant differences for any app decision factor before and after the lab study: ratings ($Z=-1.05$, $p = 0.33$), reviews ($Z=-1.42$, $p=0.16$), security ($Z=-4.0$, $p=0.77$), app store ranking ($Z=-0.10$, $p=1.00$), privacy ($Z=-0.37$, $p=0.85$). We therefore conclude that there is not enough evidence to prove that being prompted to use the privacy nutrition label during the lab study changes the importance of any of the app decision making factors.

4.2. Qualitative Data

4.2.1. App Download Process

We asked about what our participants usually do when going into the app store to download an app. We also asked our participants to Think-Aloud while deciding which apps to download from two pairs. Therefore, we collected both self-reported data and study data on what are the decision factors and what the process is like.

Out of all the interviews, the security expert is the only one who mentioned security considerations when we asked, “What's your normal process of downloading an app?” before thinking aloud tasks, mentioning they would do Google searches on the apps’ security.

Functionality

Convenience and functionality seem to dominate all of the other considerations that may factor into the app download process, particularly privacy. As P9 noted, “I just care more about like my user experience and how well I'm being like, served, whether it's like ads or content, like I just want my stuff to be, like, personalized, and if they need my data to do that, I don't mind giving it up.” A majority of participants mentioned that they only initiate the app download process when a specific need arises. Functionality alone is sufficient for some participants to download the app when there aren’t a lot of options.

Ratings and Reviews

All participants investigated the review section, in one way or another. Around one thousand seemed like a threshold for an acceptable number of reviews.

"It has 963 ratings with 4.8 stars, that makes me feel somewhat trustworthy of it"

“It has 4.7 stars with 1.8k ratings. So that leads me to believe that it's pretty safe.”

“As the number (of ratings) hit somewhere above 1k, it just feels the same.”

Others had strategies for selectively reading helpful reviews. P12 mentioned, “I actually like four stars and three stars because they usually give a fair assessment.”

Ranking

The more forward ranking apps receive the more attention and trust. One participant mentioned that they only “limit to the first 10” apps. But exceptions apply, some participants noted that their trust of an app decreases when it is labelled as a promoted advertisement even though it means it would be ranked as the first result. In P5’s words, “If it's like sponsored...I ignore those.”

Legitimacy of App

Brand Recognition

Established brands encourage downloads and also in some cases make participants feel more relaxed about sharing personal information. P10 mentioned, “I went to download Notion then put in my school email. I didn't give a second thought...I feel like it's a more established brand that's out there. "

Professionalism

For instance, the professionalism of the app, which may be reflected by the app’s UI/UX design, the app’s description, or the app’s name, may impact how legitimate the user sees the app, thereby affecting the app decision process. P10 noted, “The icon and the interface…are two things that kind of serve as an indicator for me for how developed and polished this thing is. "

Publisher/developer

Participants consider the identity of the publisher/developer. Participants would look at the other apps that the developers have made in order to determine if the developer is trustworthy. P14 talks about the desire for a more experienced developer, because it infers “quality of the app” and “the end user experience”. P11 learned the only apps developed by the particular developer is the workout app and a Holy Bible app, and surprised us with a vivid speculation, “the range of apps that they have available that they develop makes me think that they are just like one person and their basement making apps for fun. So, this makes me skeptical.” Later in the interview, they mentioned “I am not giving some person in their basement artificial intelligence information”. If the perceived legitimacy of the publisher/developer was very negative, then it would cause the participant to not want to choose that app.

Age of Application

Two participants misinterpreted the age limit listed in the App Store as how old the app is. P2 said, “the older something is the more bugs, it's been run into and the more things that you can kind of fix and update. So, age for me is really important. Just because I know it's been established, I guess.”

Overriding Factors

If the app is recommended by a friend or if it was required for either school or work, the participants don’t go through the decision process. This kind of process is described by some participants as “without thinking”. They treated it as the decision being made for them.

4.2.2. When does the Decision Take Place?

Participants also differed in their processes for downloading apps on when they decided to keep or delete an app that a participant were going to or had downloaded. Most participants go to the app store with a specific need and then decide in the app store only which one to download.

Most participants do trial and error, downloading and trying multiple apps and then deciding if they want to keep it or delete it. This category of participants self-report to spend less than a minute in the app store.

A couple of the participants would research the apps on Google to "see what's out there", before opening the app store. This category of participants self-report to spend up to 10 minutes in the app store.

4.3. Privacy Concerns

We found that some participants felt that they had no choice when it came to their personal data. Two participants emphasized that they had no choice when it came to app downloads. Both participants, P9 and P10, tended to say something along the following lines: "... even if I read what I understand ... what they say, is that what they do? You never know. So, what's the point? ... they just want you to consent. And I have to consent because I have to use your products. So, I don't feel like I was given a choice."

A larger number of participants, nine in total, stated that the data practices and labels were in line with what they expected in some way. Expectations differed slightly based on the specific app type. When presented with a pair of social apps, P1 noted in their app review that the information collected was "pretty normal for a social app." P11 was choosing a banking app and looked to make sure that an app was legitimate, and that only one of the apps we presented was in line with their expectations.

More generally, many participants expect that their data is being used for marketing and advertising purposes. P11 noted that while advertising and marketing was undesirable, it's a "normal thing that happens." Their example was that they'll "shop on Nordstrom for sandals and suddenly I have 20 ads on my CNN page about [sandals]."

Finally, some participants were less concerned about data privacy because their Information was already "out there" or being tracked. When it came to data leakage and privacy, P11 explained that "so much of that stuff is already happening like in my computer usage on a day-to-day basis ... it's really not gonna make a difference whether or not [this app has my data]."

Out of the 13 participants, eight people stated that they never looked at the privacy policy. The only participant that always looked at the privacy policy was P14, the only security expert out of our participants.

4.4. Privacy Label

4.4.1. Label Usage

During the study, there was a range of how participants used the privacy label. Six of our participants clicked into the "More details" section, where the participant could see more information on what the terms used in the main label mean. P15 said, "Then under App privacy I go into the details. And I'm trying to understand how they're going to be using this data, just scroll down through this, to understand a little bit better." Five of our participants used the label to compare between apps. P5 said "I don't remember this data used to track you tab on the other app. Maybe it was there. And I just didn't notice. I think this actually says less than the last one, but I'm not sure."

4.4.2. Trust in Information

When participants were asked if they trusted the data presented on the privacy labels, 11 people indicated that they did at some point, even though Apple says on the detail privacy page that they have not verified any of this data. Nobody saw this disclaimer and the general consensus was that Apple was trustworthy and that they had requirements for the data being presented in the labels. In the same vein as many others, P5 said "I was operating under the assumption that this is like a thing that Apple is providing me to like, and then they would… require that of the app or something. So, I assume that it's… truthful."

On the other hand, a couple participants were always skeptical about the information provided in the labels. P2 trusts that the app developers are truthful with the data they're putting on the labels, "but I don't trust that there isn't more information that they aren't disclosing that they're tracking or linking". Skepticism of full disclosure was a common theme seen in the five participants that had any distrust about the information provided in the labels.

We did have three participants who wavered between trusting the data and being skeptical about it such as P11. They said, "My initial reaction was...of course Apple must have some sort of monitor on that situation… [but] I don't know how much oversight is really in the process... [and developers] might be able to pull whatever they want, and we wouldn't really know the difference."

4.4.3. Understanding of Information

Participant understanding of the label was also a very important topic that came up in every interview. There were only four participants that had no trouble with the information on the labels or how it was presented. One such participant, P12, said of the label "It makes sense. I like the way that the sections organized" and had no issues understanding any of the information presented. All of our other participants had some sort of confusion regarding the label; this sentiment, which was echoed by many, best summed up by P8 saying "I don't really know exactly what that means." This confusion was seen both with what the sections titles mean and with what the specific information contained in the label means.

Five participants had difficulty understanding the meaning of specific data types, such as "usage data", "user content", and "other data". P5 also mentioned that while the description for "user content" stated that emails or text messages may be collected, this did not match with the icon used for "user content", which they associated with photos. One par-

ticipant, P4, said that while they didn't find the terminology "other data" to be confusing, it raised their skepticism of the label because "other data could be any data, right? If I'm giving them some input, they could tell me tomorrow that other data could be any data could be time, it could be my inputs to the game, it could be anything else. So other data is a very weird way of saying that we are publishing any kind of data."

A couple of participants mentioned that the label did not sufficiently explain what the data would be used for. One such participant, P14, said "Why does the community sports tracker need to have my information such as my name, my age, my address, my email address, my phone number? I mean, really, I think that they want my data for their own purposes." P10 echoed a similar sentiment, discussing how they usually only realize that their data is being collected when they receive targeted ads, but that "most of the time [data collection] is just very implicit."

Another issue that came up was that the same data label could appear in multiple sections and there is no clear indication of what is different between the sections. When P9 saw "usage data" appear under both the "data linked to you" and "data not linked to you" sections, there was confusion and they said, "either that's like wrong, or there's different types of usage data that's being linked, and they're not at all elaborating on that".

A couple of participants also had difficulty understanding the different section titles of the App Privacy label. One of the participants, P2, said "I'm not really sure what's the difference between data being used to track me versus data being linked to me." Similarly, during the evaluation of the second app in Task 2, P4 confused the section titles and did not notice that in the second app the second section in the App Privacy label was titled "Data Not Linked to You", as opposed to the first app where the second section was titled "Data Linked to You".

5. Discussion

5.1. Privacy Label Implications

While some participants had issues with understanding certain parts of the Privacy Labels, the overall consensus was that having the information presented was helpful to our participants. We found that many people were able to use the information to compare between applications and use it to help make their decision on what app to download if all other features were the same. However, if the apps differed on ratings, reviews, or features then the participants were more likely to weight those factors as more important than any information provided by the Privacy Labels.

Most participants felt that the privacy information is important, but we found that many people do not think there is much they can do to protect their data privacy which led to them not weighting the Privacy Label information as important as Ratings or Reviews. People self-identified as low concern with privacy issues would not change their mind with the presence of Privacy labels. People who are pessimistic about the issue understand the importance of privacy and security, but explicitly say they will not change their behavior. The privacy label does provide users with clarity on privacy information, but the huge gap of user education and behavioral change is not addressed by the current design of the privacy label.

5.2. Design Recommendations

Based on our findings, we recommend that the wording and presentation of the label's section titles in the label be changed in order to better clarify the difference between sections and the purpose of each section. We found that the terminology "linked" was too vague, and that the "Data Linked to You" section was easily confused with both the "Data Used to Track You" and "Data Not Linked to You" sections. At a minimum, users should have a way to learn more about what "linked" means.

Similarly, participants found terminology such as "usage data", "other data", and "user content" to be too vague and did not understand the meaning or purpose of the data; some participants' confusions were clarified after looking at the more detailed view of the label, but others felt that they still did not fully understand what data was being collected about them and what it was being used for. We recommend that "other data" be replaced with more specific terminology, or link to a place where users can view more detailed information about what "other data" entails.

Participants also reported that they would not normally scroll down to where the privacy label is located, and that the label did not stand out to them. We thus recommend placing the label higher up on the app page, closer to the pictures, descriptions, or ratings as these are some of the most prominent components users looked at when evaluating the apps.

One suggestion made by a participant that we found particularly interesting was to include information on how a particular app's data collection practices compared to other apps in that category. For example, if most banking apps do not collect health information but one particular banking app does, the label could note that "Health information is not typically collected by apps in the Finance category." The participant felt that this would help better inform users' consent to the app's data collection practices.

5.3. Privacy Concerns

We found that many users were disillusioned by current data collection and privacy practices. Users who expect applications to take their data are inclined to continue sharing that

data because their data is already "out there." Others feel they have no choice about what data is collected. These attitudes of indifference and helplessness may cause users to care less about the privacy label and their privacy in general, and they hinder the effectiveness of the privacy label. Additionally, users aware of the privacy label may not find the information actionable. In the short-term, privacy nutrition labels might not be helpful for those who are used to releasing their data independent of their desire. More work needs to be done to push users to be more conscious about their data.

Most users don't read privacy policies, a fact confirmed by our study. Privacy policies tend to be very long, and even one of our participants, who is a security expert, was unwilling to spend the time to read these long documents. However, participants rated the importance of data security and privacy in the pre- and post-surveys very highly, which demonstrates a gap between user needs and what is available. This supports a form of information that is summarized, such as the privacy label.

6. Limitations & Lessons Learned

6.1. Limitations

One of the major limitations of these findings is that participants were chosen from a convenience sample. Namely, participants were chosen from among our personal circles. During the interviews, we made sure that the interviewer was not a friend of the participant. In addition, our sample skewed young, and a larger sample size would have been ideal. Furthermore, some of the app categories are more security-focused than others; for instance, banking is more security-focused than games. This may result in variation in participants' attention to security depending on the app group they were assigned. Moreover, adding privacy and security to the pre-survey may have primed participants to recognize this as a criterion they should consider, especially if it is something they usually do not consider. Finally, while most of our interviews were coded by two coders, three of our interviews were coded by only one researcher due to time constraints.

6.2. Lessons Learned

Our team used the process of this study as a growth opportunity to improve our skills in study design and implementation, and there are a few places we would make improvements should this study be repeated on a larger scale. First, we recommend clarifying the survey question labels. We had some confusion later on in the study about what we meant by "privacy" and "security" as well as the difference between rating and rank. We recommend that these be changed in case the participants used the terms interchangeably.

Next, we recommend using a more randomized sample of more participants. We had issues with timing when trying to recruit participants, ended up relying on a convenience sample to get the data needed on time for our study, and were only able to interview 13 people with our time and budget limitations. Going forward, we recommend interviewing more people and having a more balanced age demographic represented.

Finally, our team did not consistently use the words "Privacy Nutrition Label" during the interview which may have led to the two people reporting they didn't know what it was during the post-interview survey. We recommend instead fixing the interview protocol and survey so that matching names are used for the Privacy Label to ensure that the name isn't the source of confusion.

7. Conclusion

Building off of previous research on app privacy labels, we evaluated Apple's newly released Privacy Labels through a virtual lab study of 13 participants. We conducted a pre-study and post-study survey to see which factors participants ranked as most important to their app download process. We also used two Think-Aloud tasks and semi-structured interviewing to explore how and why people download apps as well as how the privacy label affects their decisions. Results suggest that the labels are somewhat influential to individuals but other factors like app Ratings, Reviews, and Features are more important to the download decision. While most participants found the labels helpful, there was also confusion about what labels and data types meant on the label. We suggest that changes be made to better clarify what information is presented and possibly increase the usefulness of the Apple Privacy Labels.

8. Acknowledgements

Our team would like to acknowledge Dr. Lorrie Faith Cranor for her help as our team advisor during the term as well as for funding the project. We are grateful to Dr. Cranor and Sarah Pearman for helping our team successfully complete the study.

9. References

[1] Aleecia M. McDonald, Robert W. Reeder, Patrick Gage Kelley, and Lorrie Faith Cranor. 2009. A Comparative Study of Online Privacy Policies and Formats. Privacy Enhancing Technologies (2009), 37–55. DOI:http://dx.doi.org/10.1007/978-3-642-03168-7_3 .

[2] Siva K. Balasubramanian and Catherine Cole. 2002. Consumers' Search and Use of Nutrition Information: The Challenge and Promise of the Nutrition Labeling and Edu-

cation Act. Journal of Marketing 66, 3 (2002), 112–127. DOI:http://dx.doi.org/10.1509/jmkg.66.3.112.18502

[3] Rebecca Balebako, Cristian Bravo-Lillo, and Lorrie Cranor. 2015. Is Notice Enough: Mitigating the Risks of Smartphone Data Sharing. I/S: A JOURNAL OF LAW AND POLICY FOR THE INFORMATION SOCIETY 11, 2 (2015).

[4] Carlos Jensen and Colin Potts. 2004. Privacy policies as decision-making tools. Proceedings of the 2004 conference on Human factors in computing systems - CHI '04 (2004). DOI:http://dx.doi.org/10.1145/985692.985752

[5] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. 2020. Ask the Experts: What Should Be on an IoT Privacy and Security Label? 2020 IEEE Symposium on Security and Privacy (SP) (2020). DOI:http://dx.doi.org/10.1109/sp40000.2020.00043

[6] Le Yu, Xiapu Luo, Jiachi Chen, Hao Zhou, Tao Zhang, Henry Chang, and Hareton K. N. Leung. 2021. PPChecker: Towards Accessing the Trustworthiness of Android Apps' Privacy Policies. IEEE Transactions on Software Engineering, 47(2), pp.221-242.

[7] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A "nutrition label" for privacy. In Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09). Association for Computing Machinery, New York, NY, USA, Article 4, 1–12. DOI:https://doi.org/10.1145/1572532.1572538

[8] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. 2013. Privacy as part of the app decision-making process. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13). Association for Computing Machinery, New York, NY, USA, 3393–3402. DOI:https://doi.org/10.1145/2470654.2466466

[9] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. 2010. 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). Association for Computing Machinery, New York, NY, USA, 1573–1582. DOI:https://doi.org/10.1145/1753326.1753561

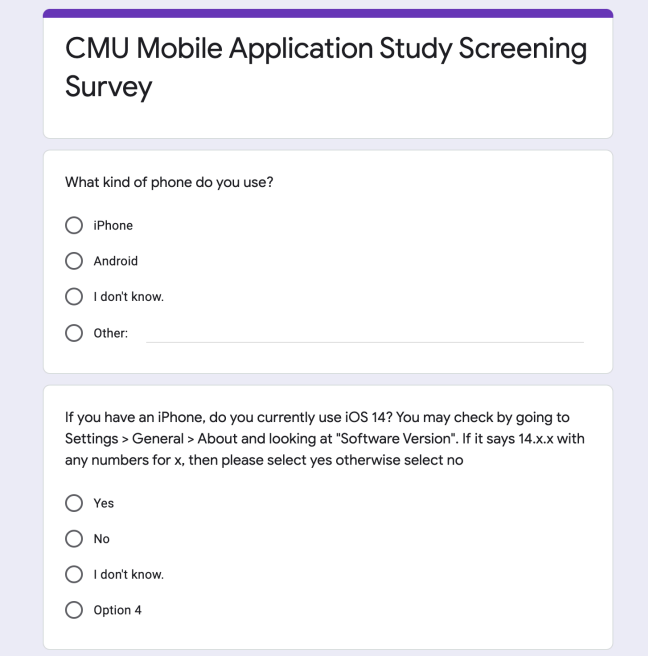
[10] P.G. Kelley, L.F. Cranor, and N. Sadeh. Privacy as Part of the App Decision-Making Process. CHI 2013.

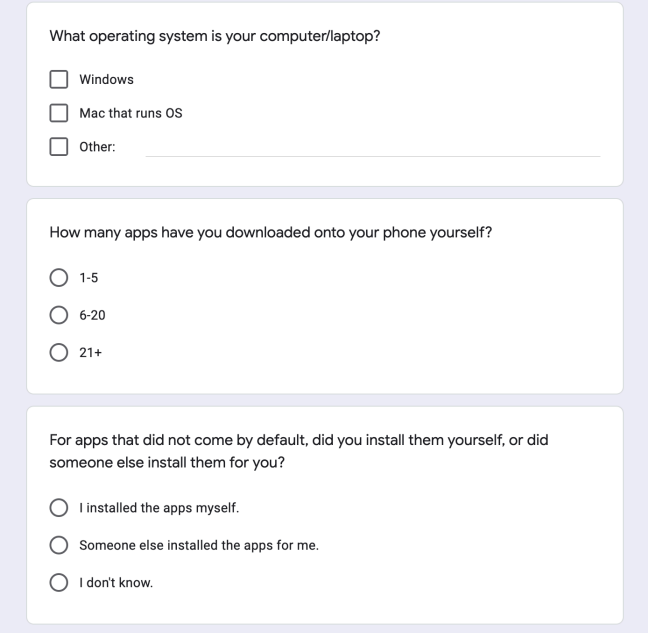
[11] Lin, J., Sadeh, N., Amini, S., Lindqvist, J., Hong, J. I., and Zhang, J. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. UbiComp '12, ACM (2012), 501--510.

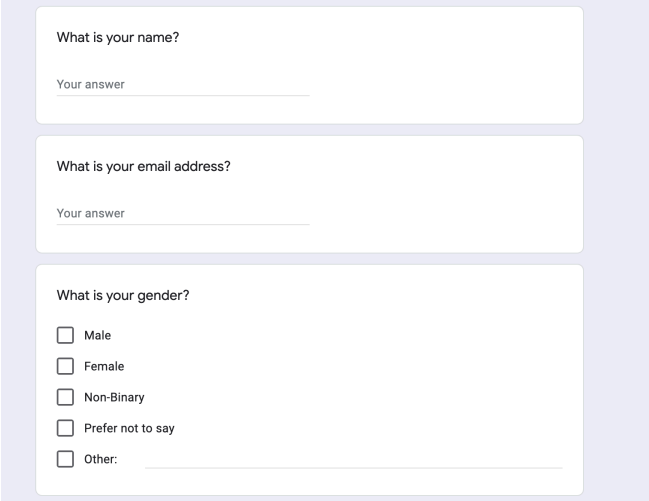
10. Appendix

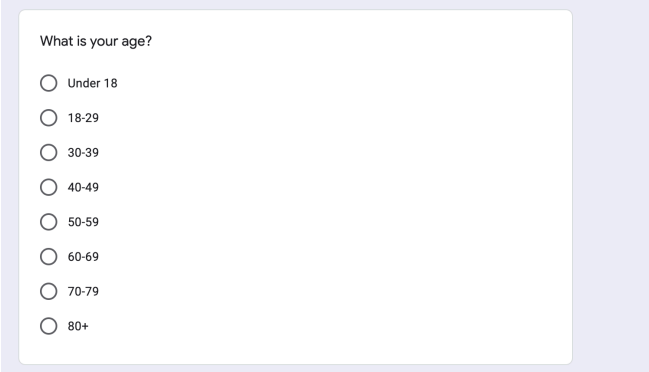
10.1. Screener Survey

Here we provide the survey used to screen participants. We are prioritizing iPhone users who have many apps downloaded. We also collect demographic information, so we can hopefully have a balanced participant group.









10.2. Pre- and Post-Interview Survey

Here we provide the pre-survey and post-survey. Note that the survey is the same for both; the only difference is whether it is administered before or after the interview.

Indicate how important each factor is when you are downloading an app.

	Extremely Important	Important	Neutral	Unimportant	Extremely Unimportant	Don't know what this is
App store rating	○	○	○	○	○	○
App store reviews	○	○	○	○	○	○
App security	○	○	○	○	○	○
App Store Ranking	○	○	○	○	○	○
Data Privacy	○	○	○	○	○	○
Privacy Nutrition Label	○	○	○	○	○	○

10.3. Apps used for Tasks

The apps and categories we used in our study are listed below.

Fitness:

- Sports Tracker for All Sports (https://apps.apple.com/us/app/sports-tracker-for-all-sports/id426684873) (1.8k ratings)
- SmartGym: Gym & Home Workouts (https://apps.apple.com/us/app/smartgym-gym-home-workouts/id922744883) (4.6k ratings)

Social Media:

- VK – social network (https://apps.apple.com/us/app/cobra-kai-card-fighter/id1505336575) (2.9k ratings)
- Clapper: Video Community (https://apps.apple.com/us/app/solitaire-mystery/id1474804524) (3.9k ratings)

Games:

- Cobra Kai: Card fighter (https://apps.apple.com/us/app/cobra-kai-card-fighter/id1505336575) (813 ratings)
- Solitaire Mystery (https://apps.apple.com/us/app/solitaire-mystery/id1474804524) (2.4k ratings)

Online Banking:

- Porte: Mobile Banking (https://apps.apple.com/us/app/porte-mobile-banking/id1494730613) (722 ratings)

- OneUnited Bank Mobile Banking (https://apps.apple.com/us/app/oneunited-bank-mobile-banking/id520589589) (4.6k ratings)

Below are the app categories we assigned to each participant for each respective task. The categories "Fitness" and "Social Media" were used in 7 of the interviews, while "Games" and "Banking" were used in 6 of the interviews.

Participant ID	Task 1	Task 2
P1	Fitness	Social Media
P2	Games	Banking
P3	Banking	Social Media
P4	Fitness	Games
P5	Social Media	Banking
P8	Social Media	Fitness
P9	Fitness	Banking
P10	Games	Fitness
P11	Banking	Fitness
P12	Social Media	Games
P13	Banking	Games
P14	Fitness	Social Media
P15	Games	Social Media

10.4. Interview Discussion Guide

Here we provide the complete discussion guide used during interviews. Note that we are following a semi-structured interview format, so the guide may not be rigidly followed.

UPS Class: Evaluating the Effect of Apple Privacy Nutrition Label on App Download Decision Making

Ananya Bhat, Amanda Crawford,

Evan Zeng, Jimmy Ray, Mira Mookerjee, Denise Tang

Introduction

Hello! My name is _____ and I'll be leading the study today, and this is _____ who will be taking notes. Thank you again for agreeing to participate in this research study!

Consent and Description

Today we will be doing a virtual lab study that consists of two quick surveys at the beginning and end of the session, a few scenarios, a think aloud task which will be explained before we do it, and some semi-structured interview questions to learn more about your experiences with the Apple App Store.

During certain parts of the study, we will be asking you to share your phone or computer screen in order to see what you are doing in regard to the specific tasks we ask you to do. Please be aware that anything on your screen will be seen by the researchers at that point so please put anything confidential off screen and silence notifications if possible.

We will be asking to record this interview through video for further analysis, though your name and any other potentially personally identifiable information will be kept confidential and will not appear in any report or document outside of the team. All study data will be stored for a minimum of 3 years at CMU. The entire study process should take no more than one hour to complete and you will receive $15 upon completion of the study. If you can, please ensure that you are in a private space for the interview so that passersby are not inadvertently recorded.

Please be aware that at any point in time during the interview, you may stop the interview, choose to not answer any questions, or take a break if you wish. Do you have any questions so far about what we will be doing?

Introduction [~5 minutes]

[Goal: Ensure compliance with the requirement for participation and then Introduce the idea of the Apple App store and downloading applications as well as learn what parts of the download process matter to this participant.]

To start off, we are interested in learning about the Apple App store and requirements for this study are that you either have an iPhone or a Mac Computer.

What type of mobile device do you have?

[If an iPhone] What version of iOS are you running? You can check this by going to Settings>General>About and reading the number for "Software Version".

Do you have a mac computer instead?

[if yes] What version of macOS are you running? You can check this by clicking on the apple icon in the top left corner, going to About this Mac and reading the line that starts "macOS"

[If the participant doesn't have an Apple iPhone that is currently running iOS 14 or a Mac computer running Big Sur then end the interview because the participant does not qualify.]

One last question before we begin, have you downloaded at least 5-10 applications from the Apple App store yourself in the entire time you have had your device?

[if the answer is no, end the interview because the participant does not qualify.]

Now we will move on to the lab study portion, we will be asking you to fill out this survey so we can learn more about what is important to you when downloading something from the Apple App Store. Please take a few minutes to complete this form and let us know if you have any questions or when you are done.

Your number is [participant ID number for survey identification], please write this number down for the first question.

https://forms.gle/ubWBZu4Ft6qMTYXRA

Initial semi-structured questions [~10 minutes]

[Goal: Get a basic understanding of the participant's specific experiences with the Apple App Store and technology comfort levels].

To start off, [researcher then starts a conversation around the Apple App store using questions like the ones below as they fit into the conversation] ...

Would you consider yourself technically savvy? Why or why not?

Would you consider yourself an early technology adopter?

How do you feel about the App store?

When was the last time you downloaded new apps?

What did you download?

How often do you download new apps?

Why do you download new apps?

What activities do you do on the App Store? [find out if they compare apps, and search eg.]

How long do you spend in the App Store before deciding which app to purchase? Why?

First Think-Aloud Task [~15 minutes]

[Goal: understand specifically how a participant decides whether to download a specific application].

Now we will be entering the "Think Aloud" portion of the study. Here, we will ask you to talk out loud as you are doing a task so we can understand what you are thinking and we can have a conversation about what you are doing and why. For example, if I was asked to find a submit button on a web page, I would say "I am looking at the whole web page and scanning for the button. I am looking for the word submit and am looking in the bottom right because that is normally where I would expect to find it."

Do you have any questions about what a think aloud is? (wait for and answer any question the participant has before moving forward)

This section also requires you to share your screen with us through zoom. Are you using your phone or computer for this?

[Guide through instructions for sharing phone screens if needed https://docs.google.com/document/d/10TGTP0H7SUsCQ3Y-KaLulxbgBWmUfffudEZTI2TG0Wc/edit?usp=sharing]

Great, now that you are sharing your screen, your task is to "Think Aloud" as you evaluate two fitness apps. Please evaluate each app individually, and then compare them and decide which one you would download. You do not have to actually download the app, just let me know which app you would download from the two provided and why.

[Give app names from interviewer list https://docs.google.com/document/d/18dRChneXie-rSA3OJF6SFlFe2uDOHW7Z7d00dBybd6Y/edit?usp=sharing]

[researcher then starts a conversation around what factors the participant is considering]...

Is there anything you like about this app? [Why?]

Is there anything you dislike about it? [Why?]

[If the participant is quiet] what are you thinking about right now?

Why did you decide to look at the [privacy nutrition label, rating, reviews, photos of game, etc.]?

When they make a decision about what app to download:

Why did you choose that app to download?

For the not chosen one, what could be changed so that you choose it?

What information does this app have access to when you download it?

Is there anything else you would like to say about these apps?

Second Think-Aloud Task [~15 minutes]

[Goal: understand specifically how a participant decides whether to download a specific application when prompted to look at the Privacy label. Understand what information is understandable, relevant, and important to the participant].

Alright, now we will be doing the second task. For this task, we will once again have you "Think Aloud" as you evaluate two apps. Please evaluate each app individually, and then compare them and decide which one you would download. We ask that you specifically look at the information under the "App Privacy Section" during your evaluation. For this round we will be comparing two social media applications.

[Give app names from interviewer list https://docs.google.com/document/d/18dRChneXie-rSA3OJF6SFlFe2uDOHW7Z7d00dBybd6Y/edit?usp=sharing]

[If they have trouble finding the privacy section show directions https://docs.google.com/document/d/1xe_9MX_4dAciIyCQ_QQubiWthWNG1ga594K6iXec1qE/edit?usp=sharing]

Again, for the think aloud portion, please describe anything you are thinking about or that you are looking for. And please continue sharing the screen of your device. Any questions? (questions…) Let's get started.

[researcher then starts a conversation around what factors the participant is considering]...

Is there anything you like about this app? [Why?]

Is there anything you dislike about it? [Why?]

[If the participant is quiet] what are you thinking about right now?

Is there information here that is important to you?

[if yes] What? Why? What is the impact of this information?

[if no] Why not?

Is there any information here that you do not understand?

[if yes] What? Why? What is the impact of you not understanding this information?

Why did you decide to look at the [privacy nutrition label, rating, reviews, photos of game, etc.]?

When they make a decision about what app they would download:

Why did you choose that app to download?

For the not chosen one, what could be changed so that you choose it?

Is there anything else you would like to say about these apps?

What information does this app have access to when you download it?

Thank you for completing the think-aloud tasks, you can stop sharing your screen now.

Post task Interview questions [10 minutes]

[Goal: understand how privacy concerned the participant is, whether they have seen the labels before, and any final comments about downloading applications].

We have just a few more questions for you. To start, have you ever seen or used the Apple Privacy label section before?

[wait for answer, either way ask] What do you think about the labels?

[Researcher continues the conversation with questions like the following] :

Did the information provided on the privacy data label affect your decision? [Why?] [Why not?] [How?]

Prior to this study, have you read the privacy policies (partly or fully) of any apps that you own?

Why or why not?

Do you think the information presented in the labels is reliable?

Why or why not?

Do you have any reservations when deciding to download an app that were not mentioned earlier?

[If yes]

Have you ever regretted downloading an app because of data privacy issues?

Why or why not?

Did you take any further actions because of these regrets, such as changing your privacy settings, or uninstalling the app? Why or why not?

[Always ask as time is ending] Is there anything else you would like to tell us?

Wrap-up [~5 minutes]

Alright, the interview portion is over and now we will be asking you to fill out the same survey as at the start of the interview, so please take a few minutes to complete this form and let us know if you have any questions or when you are done.

As a reminder, your number is [participant ID number for survey identification], please write this number down for the first question.

https://forms.gle/BtvYEM5jn5dYiuJf8

Well, that is all the questions we have for you, do you have any questions for us? Any other thoughts, feedback, advice you would like to share?

Thank you so much for your time!

10.5. Survey Data

Here are the links to the spreadsheets with the pre- and post-interview survey data.

Pre-Interview:
https://docs.google.com/spreadsheets/d/1MTvYgn6xAzjSbI3Npe0f8lUxtTRI1XJzNNg-RF1B0Ec/edit?usp=sharing

Post-Interview: https://docs.google.com/spreadsheets/d/1-N29osrgN_5Yn8oVk_JSpFgpUulpCMpo3HmrRRrjSYI/edit?usp=sharing

10.6. Survey Data

Here is the link to our spreadsheet with the code definition, summary of which codes apply, and all coded interviews. All identifiable information about participants is removed from this spreadsheet and kept only in Otter.ai and the cloud recordings.

Code:
https://docs.google.com/spreadsheets/d/13cQ7dBAj6E9eZ57ghxZD-yFdOeHJjxZgHIW1uff9Bxk/edit?usp=sharing