Exploring Notice and Choice: Design Guidelines for a User-Centered Permission Model for Personalized Services

Maritza Johson[1] **Oriana Riva**[2] **Jaeyeon Jung**[2] **David Wagner**[1]
[1] UC Berkeley
[2] Microsoft Research

ABSTRACT

When users install an application on their mobile Android device, they must decide whether to grant the application access to privacy-sensitive information. Building upon recent research on the limitations of the current installation dialog, we investigate a new permission model better aligned with users' current understanding of data collection and privacy concerns. Using three mobile applications that offer highly personalized services, we conduct a series of online surveys (n = 1,316) varying the information disclosed in the installation screen and exploring the option to limit data collection. First, we identify two factors that significantly affect users' willingness to install applications—the frequency of data collection and the third party sharing policy—and find that in the absence of such information, as in the case of existing dialogs, many survey respondents incorrectly assume that an application's capabilities are more restricted than they actually are. Second, we find that offering a third option beyond existing all-or-nothing choice affects users' willingness to install applications. We further analyze how our participants would like to limit data collection, and compile a set of suggestions for the design of a permission model that can provide better privacy notice and meaningful controls to users.

INTRODUCTION

Mobile applications continue to evolve in order to provide a better user experience through context-aware, personalized services. For example, with GPS-equipped smartphones, users can easily search nearby restaurants, look up bus schedules, and receive coupon offers. It is not hard to imagine that in the near future these simple location-based applications will provide better recommendations by leveraging more data about the user (e.g., a restaurant recommendation application that infers the user's culinary preference from food-related websites that the user has visited); a few examples are already available in application markets [9, 19, 2].

Meaningful personalization requires copious amounts of user data. Some applications may require near continuous access to privacy-sensitive data, such as location, browsing/search history, email, and call history. As such, privacy risks that users perceive can be high and may become a barrier for these applications to be adopted. This study focuses on aspects of applications' collections and sharing of data that affect the perceived privacy risks measured as the users' willingness to install applications.

Increasingly, mobile devices allow users to install third-party applications from marketplaces. The mobile platforms and the marketplaces have introduced new fine-grained permission models for applications. Although this trend is an improvement compared to the desktop security model, many recent studies [13, 8, 7] show that existing permission models are ineffective as users do not understand the information presented by the permission screen.

To address these issues and build a foundation for a user-centered permission model, we need to understand users' current assumptions about granting an application permission to access personal data and their concerns about sharing personal data. Toward this goal, we explore the following research questions:

- **Understanding of approving permissions.** Do smartphone users understand the capabilities granted to an application when a permission request is approved?
- **Understanding of the implications of exposing sensitive personal data.** Do smartphone users understand that an application can infer additional information about a person when it is granted the ability to collect a continuous stream of personal data?
- **Attitudes toward sharing of data with third parties.** How are installation decisions affected by an upfront disclosure of an application's third party sharing policy?
- **Offering choices to limit data collection at install time.** How does offering choices to limit access to sensitive data affect users' installation decisions?
- **Designing options to limit data collection.** What data users want to keep private from applications and why?

We create three fictitious yet well-received mobile applications that rely on continuous access to privacy-sensitive information, i.e., location, email messages and web browsing history. We then design online surveys using installation dialogs similar to that of Android to measure users' willingness to install these applications We iterate the survey designs and collect responses from 1,316 participants using Amazon Mechanical Turk. We use various measures to filter out "mindless" respondents.

We find that study participants are confused about the capabilities of applications, how data is collected, and how collected data might be used. Furthermore, we find two

key points of confusion that affect participants' willingness to install applications: the frequency of data collection and whether the data will be shared with third parties. We also find evidence that suggests that people are more willing to install an application when they are offered choices to limit data collection.

Based on our findings, we contribute a set of design guidelines for improving the existing permission installation model with a focus on increasing notice and providing the user with meaningful choices. After discussing related work, we present our survey methodology for investigating the research questions listed above. We then discuss the quantitative and qualitative analysis results of the responses we collected. We conclude the paper with guidelines for the design of a user-centered permission model for personalized services.

RELATED WORK

We categorize related work by three themes. First, we discuss work demonstrating the limitations of the current privacy framework of notice and consent. Second, we discuss recent user studies that highlight users' misunderstanding of permission requests by smartphone applications. Third, we present user studies that investigate meaningful ways to allow users to limit data disclosure.

Notice and Consent

Online services rely on "notices" such as terms of service (ToS) and end user license agreement (EULA) when communicating privacy risks to users, and users are asked to give "consent" in order to use the services. Jensen and Pott [11] evaluated the usability of privacy policies and concluded that although policies are widely available on web sites, their format, location on the site, and legal context make them ineffective even for privacy-concerned users. Another study shows that people are habituated to accept EULAs and consequently they tend to blindly accept any terms whose form mimics a EULA [3]. Good et al. [10] investigated how notices such as software agreements, ToS, and EULAs influence users' decisions on installation of applications that may contain spyware. Short and concise notices are more likely to be noticed by users, but they do not seem to significantly impact their decisions. However, privacy and security are noticed when a user is deciding between two services. Our study explores how short and concise notices about specific aspects of an application's functionality (e.g., frequency of data collection, sharing with third parties) that seem to be poorly understood by users can influence their installation choices, and possibly their perception of privacy risks.

Our study also explores new forms of consent. The implementation of consent for online services is fairly limited. Barocas et al. [1] observe how, in the context of advertising, the opt-in approach to consent is basically non-existent and the opt-out approach is unclear. Our approach to consent is influenced by Nissenbaum's theory of contextual integrity [16]. She argues the privacy norms are not "one-size-fits-all" but rather are distinct to each situation. Inspired by her work, we investigate how users would prefer to limit data disclosure when using different applications and different data types. Based on their expectations and preferences, we derive new ways to help users control data access.

Permission Models for Mobile Applications

Recent studies show that users rarely pay attention to permission requests by mobile applications and find it difficult to understand them, even when making a deliberate effort. Felt et al. show that Android users demonstrate low rates of comprehension [8]. A similar finding was reported by Kelley et al. [13]. These studies motivate our work of designing a better permission model. In an effort to improve the existing permission models, Lin et al. propose a new privacy summary interface showing data uses that are known to violate people's expectations [15]. The authors use crowdsourcing to capture users' expectations about which sensitive resources each mobile application can use and build the summary of expectation-violating data uses. In contrast, rather than looking at expectations about specific applications, we study users' misconceptions about how applications access (e.g, at which frequency) and share (e.g, with advertising companies) personal data. As recently proposed by Kelley et al. [14], we then provide privacy information in the installation process and measure how addressing the major misunderstandings we discovered in the application's description screen affects installation decisions. Our findings complement Kelley's design of privacy display.

Users' Attitudes Toward Sharing Personal Data

Several other studies have investigated people's attitudes toward sharing location data [5, 17, 20, 12]. In particular, Kelley et al. [12] explore a usable permission model for sharing location with advertisers. They find that mechanisms that allow users to selectively share location data are needed rather than the existing "all-or-nothing" control. In contrast, we focus on users' attitudes toward data sharing in exchange for a service—we consider only applications that require personal data for their operation. Our findings confirm the observation that a wider range of permission choices is needed.

Previous work explored different data access modes for location applications. Two of the data access modes that we derived from our participants' suggestions are similar to the "deleting" and "discretizing" modes proposed by Brush et al. [4]. Our recommendations to allow finer-granularity control over data sharing is in line with the findings of Tang et al. [18]. Their study was in the context of sharing location with people, while we look at sharing with applications, but in both cases, users seem more willing to share location information if given the option to control its granularity.

Few studies have looked at users' attitudes toward sharing personal data beyond location. Egelman et al. [6] evaluate how the number of permissions requested by an application can influence users' installation decisions. In particular, they find that users are less concerned about sharing location and more about sharing their address book or audio data; most users see location-based features as desirable rather than privacy-invasive. Our work suggests that users may be concerned with sharing of web browsing history and emails than location, but, contrary to their findings, we find that sharing of location is still a major concern. In our re-

Survey	Condition	Install options	n
Baseline	Control	Yes/No	260
Inferences	Warning	Yes/No	278
Sharing	Control	Yes/No/Limit	247
	No Sharing	Yes/No/Limit	272
	May Share	Yes/No/Limit	259

Table 1. Summary of the three surveys we conducted.

search, we explore ways to communicate the risks associated with sharing personal data and investigate alternative sharing options that users desire. Our approach enables us to collect people's concerns with respect to sharing different data types grounded in the context of installing personalized applications.

APPROACH AND METHOD

Towards the goal of designing a user-centered permission model for personalized applications, we decided to focus our analysis on three types of potentially sensitive personal data: location, email messages, and web browsing history. We created three fictitious mobile applications representing personalized services that need continuous access to one such data type to provide their service: Fill Me Up!, NewsRecommender and BillKeeper, respectively. To generate compelling applications, we solicited ideas from colleagues and then used a short survey to test twenty potential applications using Amazon Mechanical Turk (AMT). The three applications featured in the surveys were the perceived to be the most useful. For the descriptions of these three applications, see Figure 1.

As we wanted to reach a large population of users, we recruited participants using AMT. We restricted participation to U.S. residents over the age of 18. Moreover, when we posted our task, we requested that only smartphone users accept the task, then we asked phone-related questions (e.g., "Which version OS is running on your smartphone?") to eliminate those who did not seem to own a smartphone. We collected demographic data about all participants including age, gender, and occupation. We collected a total of 1,316 valid responses, after removing incomplete and dubious responses. Our respondents included slightly more men than women (65.8% male). Ages ranged from 18–67 ($\mu = 29.3$, $\sigma = 9.0$). The majority of our participants are Android users (53.1%), and iPhone users were well-represented as well (41.9% of 1,316).

Table 1 summarizes the surveys we conducted and the number of participants for each. We designed a total of three surveys which will be described in more detail in the following sections. Each survey respondent was randomly assigned one of the three applications to be the focus of the survey.

To mitigate concerns about the quality of the responses, we included extra questions to identify bogus responses. For example, we asked participants to categorize the application they were given to ensure they read and understood the description; we deleted responses that were clearly wrong (e.g., one participant categorized Fill Me Up! as a game). We used the free-form questions to eliminate respondents who entered strange answers. Finally, we took measures to ensure an individual only completed one survey: we cross-checked AMT identifiers, requested that each person only complete one survey, and used cookies to identify duplicate responders.

EXPLORING NOTICE

We begin our effort of designing a user-centered permission model by examining the existing application installation process. This process has several shortcomings which may limit users' understanding of an application's permission requests and thus impact their privacy decisions. Similar to prior work [8, 13], our experiments use the Android installation process as the status quo for mobile systems, but our results have broad applicability to mobile applications regardless of the platform. We consider the following critical pieces of information that could help the user make a trust decision but are currently unavailable to an end user:

1. When can the application use the permission?
2. How frequently can the application use the permission?
3. How could the sensitive data be used to infer or guess other information about the user?
4. Will the application share the user's data with third parties?

We study smartphone users' present knowledge and ability to answer these questions. Our results reveal a gap between users' current understanding of data collection and actual application behavior; these findings led us to experiment with ways to close this gap.

Smartphone Users' Assumptions about Applications

An understanding of how an application will behave after installation is an important aspect of the user's decision whether to install the application. For this reason, we designed a first survey (*Baseline*) exploring smartphone users' assumptions about how applications work in practice. Each survey respondent was assigned one of our three applications. As an example, those assigned the Fill Me Up! application saw the dialog shown in Figure 1(a). In this survey, the dialog displayed only the description of the application and the permission that would be accessed. We refer to this as the *Control* condition.

The survey began with questions about the fictitious application, and continued with a set of true-false questions to evaluate assumptions about the application's capabilities. We asked participants to respond to the knowledge questions as though they had chosen to install the application introduced in the survey. The questions covered aspects of the application's behavior such as when the resource would be used, how frequently it would be used, whether or not the data would be shared with third parties, and whether or not the application could make specific inferences. Of the ten knowledge questions we asked, we found that there were some questions that participants could already answer, given

(a) *Control* (b) *Warning* (c) *No Sharing* (d) *May Share*

Figure 1. The installation dialogs used when the survey asked, "Based on the installation screen above, would you install *App* on your phone?" The *Control* condition (a) included only the application description and the permission required. The *Warning* condition (b) included a notice on the inferences an application could draw. The *No Sharing* condition (c) included a notice that the user's data would not be shared. The *May Share* condition (d) included a notice that the user's data may be shared.

the *Control* dialog (see Table 2[1]).

Correct Assumptions

For all three data types featured in the survey, we found that participants correctly answered the question about the granularity of the data that would be shared (see Table 2, Question 4). For location, most participants correctly answered that a location-aware application could infer the user's commuting habits. For email, most participants correctly answered that an email-based application would be aware of where the user had financial accounts based on emails received. For browsing history, participants correctly answered that an application could guess their hobbies and interests from the websites they visit.

Inaccurate Assumptions

The survey asked a question about when an application can use a permission that has been granted. About half of the survey respondents thought the application can access data only when running (i.e., they did not think it would be possible for the application to run in the background and collect a continuous trace of their data). The survey asked two questions related to data sharing, even though the installation dialog for the *Control* condition did not explicitly mention data sharing. This allows us to measure users' assumptions about how data is shared. If we consider that users rarely, if ever, read privacy policies, these assumptions are quite interesting. For all three data types the proportion of users who selected each answer are quite similar: approximately one-third of participants thought that the application would not share their personal data with advertising companies (see

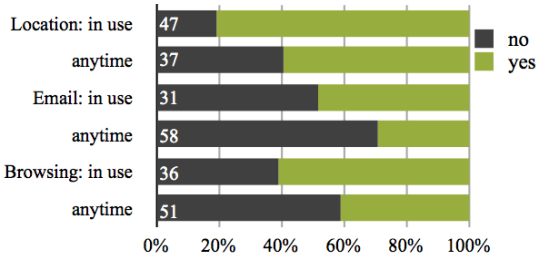

Figure 2. Participants' willingness to install the application in the *Control* conditions of the *Baseline* and *Sharing* survey. Participants from the two surveys were combined, then split by responses to the true-false question on when an application can request the data. 'In use' means the survey respondent answered true to Question 1 in Table 2. 'Anytime' means they answered false. Numbers in the bar report the total number of participants.

Table 2, Question 5). Furthermore, about the same proportion of participants believed that the application would not share data even at the request of law enforcement officials (see Table 2, Question 6).

We hypothesized that people who assume an application can only access the requested resource when they are using the application will be more willing to install it. To test this, we combined the responses from the *Control* condition for all three applications from both the *Baseline* and *Sharing* survey[2], and then divided the responses based on the answer to Question 1 in Table 2. Finally, we compared the two groups by willingness to install the application. We found that participants who believe an application can only request the resource when they are using the application were more

[1]Question 2: on iPhone and Android phones, a little icon appears in the status bar when an application uses location.

[2]The *Inferences* survey will be introduced later, but the structure is similar to the *Baseline*

App	True-false Statement	A	Control
1.	*App* can access my *Resource* only when I use the app (when the app is open on my phone's screen).	F	58.4%
2. L	A notification/icon will appear on my phone's screen when Fill Me Up! accesses my location.	T	62.6%
2. E	A notification/icon will appear on my phone's screen when *App* accesses my email messages.	F	64.5%
2. B	A notification/icon will appear on my phone's screen when *App* accesses my browsing history.	F	67.1%
3. L	Fill Me Up! can access my phone's location at any time, even when my phone is off.	F	76.6%
3. E	BillKeeper can access my phone's email messages at any time, even when my phone is off.	T	58.1%
3. B	NewsRec. can access my phone's browsing history at any time, even when my phone is off.	T	67.7%
4. L	Fill Me Up! can access the GPS coordinates of my phone's current location (e.g., latitude: 47.505 , longitude: -127.2045).	T	94.7%
4. E	When BillKeeper accesses my email messages, it receives the subject, body, sender, and recipients of each message.	T	84.9%
4. B	When NewsRecommender accesses my browsing history, it receives the list of websites I've visited and how many times I visited each website.	T	92.7%
5.	*App* could share my *Resource* with advertising companies that I had no intent of sharing information with.	T	63.3%
6.	*App* will never share my *Resource* with anyone, not even at the request of law enforcement officials.	F	73.2%
7. L	Fill Me Up! can infer my commute patterns (e.g., what time I leave for work on weekdays or the routes I commonly drive).	T	84.8%
7. E	BillKeeper can see who I email and determine how frequently I email them.	T	72.7%
7. B	NewsRecommender can infer my interests and hobbies based on the websites I visit most often.	T	93.3%
8. L	Fill Me Up! can guess my income range based on where I live, work, and shop.	T	35.7%
8. E	BillKeeper can learn about my online shopping habits from the order confirmation emails I receive.	T	76.7%
8. B	NewsRecommender can guess my political and religious affiliations.	T	71.3%
9. L	Fill Me Up! can guess whether I have children if I visit a daycare center or primary school.	T	44.4%
9. E	BillKeeper can guess my nationality and spoken languages based on my emails.	T	44.8%
9. B	NewsRecommender can guess my nationality and spoken languages based on the websites I visit.	T	82.9%
10. L	Fill Me Up! can guess whether I have a medical condition if I visit hospitals or specialty clinics.	T	39.2%
10. E	BillKeeper can learn where I have financial accounts based on the emails I receive from banks.	T	87.8%
10. B	NewsRecommender can guess whether I have a medical condition if I visit websites about specific health conditions.	T	68.9%

Table 2. We asked true-false knowledge questions to investigate users' assumptions about applications' capabilities. The *App* column indicates the question number and the data type that the user saw: location (L), email (E), or browsing history (B). The absence of a letter indicates a summary of the responses aggregrated across all three data types. The *Ans.* column indicates the correct answer based on the information shown in the survey. The remaining column reports the percentage of participants who answered the question correctly. Questions that summarize all *Control* participants represent 507 people. Questions that represent one application represent: location = 171 people, email = 172 people, and browsing history = 164 people.

likely to be willing to install it (location p = 0.02808; email p = 0.06067; browsing p= 0.053; one-tailed Fisher's exact test) See Figure 2. Furthermore, the difference was consistent across all three applications: an increase in willingness to install by about 20 percentage points.

For each application, we presented four true-false questions related to personal information that an application could potentially infer with access to the requested data type (Questions 7–10 in Table 2). Each question represents information that an application could potentially infer with varying degrees of accuracy depending on individual circumstances, however, it is interesting to learn which inferences participants believe are possible. For example, nearly all of the participants realized that sharing their web browsing history reveals their hobbies and interests (93.3%, Question 7B). However, very few participants realized that sharing their location could reveal demographic information such as their income range (35.7%, Question 8L).

Warning Users about Potential Inferences

We tested a simple design fix in an attempt to adjust smartphone users' assumptions about the inferences that an application could make with access to personal data. First, since our previous survey showed that many smartphone users are confused about how often an application can access their data, we decided to make this information explicit in the installation dialog. Second, we added a one-time notice that warns users about the potential for an application to infer additional information about the user from the disclosed data.

We designed a second survey, which we call the *Inferences* survey. It consists of one condition (*Warning*), where we used the installation dialog shown in Figure 1(b). The dialog mentioned how frequently data would be collected and included a warning about the potential inferences that the application could draw. For example, in the case of Fill Me Up! requesting access to location, the installation screen notified the participant that:

> An app can learn or infer many things about your habits based on the places you go (e.g., where you live and work, whether you have children, how frequently you visit a hospital, etc.).

As in the previous survey, each respondent was randomly assigned one of the three applications. The survey included questions about the fictitious application, and true-false questions to evaluate assumptions about the application's capabilities.

We hypothesized that people who are notified of the possible inferences an application could make are less willing to install the application. We tested this hypothesis by comparing participants' willingness to install the application in the *Control* and *Warning* conditions, and found that participants in the *Warning* condition were less willing to install the application for each of the three data types The difference between the two conditions was significant (location p = 0.05132; email p = 0.05311; browsing p= 0.005384; one-

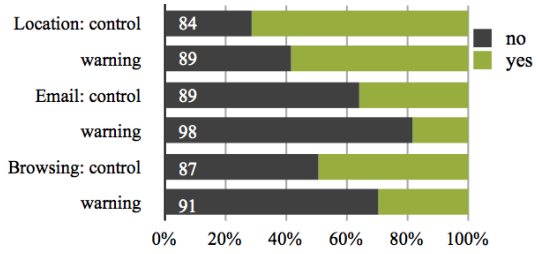

Figure 3. Participants' willingness to install the application in the *Control* and the *Warning* condition, respectively, from the *Baseline* and *Inferences* surveys. Numbers on the bar report the total number of survey respondents.

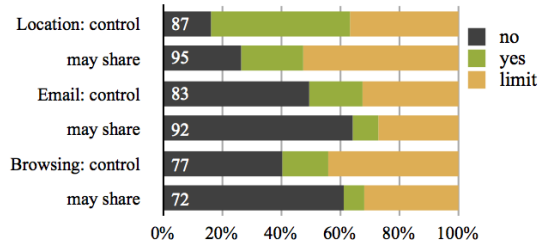

Figure 4. Participants' willingness to install the application in the *Control* condition and the *May Share* condition from the *Sharing* survey. Numbers on the bar report the total number of survey respondents.

tailed Fisher's exact test). See Figure 3.

In general, willingness to install the location-based application (Fill Me Up!) was higher than for the applications of the other two data types. This may be because participants considered emails and browsing history more privacy sensitive, but it could also be that they simply found the Fill Me Up! application more appealing.

Prominently Stating the Data Sharing Policy

After exploring the frequency and inference aspects, we designed a third survey (the *Sharing* survey) to investigate the effect of including a notice about whether or not the data might be shared with third parties. Also, in contrast to the two previous surveys, in the installation question, we included another choice in addition to "yes" or "no." We added the option "yes, if I could limit data collection." We will evaluate in the next section the effect of this alternative installation option.

In the *Sharing* survey, we tested two experimental conditions (installation dialogs are shown in Figure 1(c) and 1(d)). In the *No Sharing* condition, the installation dialog displayed a notice stating that the application will not share the collected data with third parties. In the *May Share* condition, the dialog displayed a notice stating that the application may share personal data with third parties. In the *Control* condition, we used the same installation dialog used in the *Control* condition of the previous two surveys (see Figure 1(a)). Survey respondents were randomly assigned to a condition and to an application.

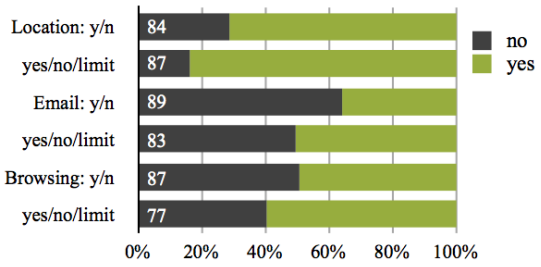

Figure 5. Participants' willingness to install the application in the *Control* condition from the *Inferences* survey and the *Control* condition from the *Sharing* survey. Numbers on the bar report the total number of survey respondents.

We hypothesized that people would be less willing to install an application when notified that an application may share their data. We compared participants' willingness to install the application between the *Control* and *May Share* conditions. We found that in the case of location and browsing participants were significantly less willing to install the application (location $p = 0.00094$; browsing $p = 0.02917$; Fisher's exact test). For email, the difference was not quite significant ($p = 0.08808$; Fisher's exact test). See Figure 4.

EXPLORING CHOICE

The results of our investigation of users' assumptions about applications' capabilities suggest that many smartphone users have misconceptions about how their data is collected and used. We also saw that once we made our survey respondents more aware of an application's actual behavior, their willingness to install the application decreased. We investigate whether having more control on data collection can help address their concerns and therefore make them more interested in using personalized services.

Specifically, we offered our participants a third installation option. The *Sharing* survey asked whether participants would like to install the application and offered three options: "Yes", "No", and "Yes, if I could limit data collection". If the latter option was selected, the survey asked "How would you want to limit data collection?" and collected free-form responses. We measured how many users chose this option. We also analyzed users' free-form responses that describe how and why they would use this option.

Beyond "All or Nothing" for Installation

We hypothesized that people who are presented with three options for installing an application, rather than a simple yes-no choice, will be more willing to install the application.

We compared the participants' willingness to install the applications between the *Control* conditions of the *Inferences* and *Sharing* surveys. These two surveys presented exactly the same installation screen in both *Control* conditions, but the installation options were different. For this analysis, we combined the respondents who reported they would install the application and those who reported they would install the application if they could limit data collection. The results are

shown in Figure 5. We found that the participants who were given three choices were more likely to report willingness to install the application for Fill Me Up! and BillKeeper (location, $p = 0.03735$; email, $p = 0.03704$; one-tailed Fisher's exact test). For NewsRecommender, the difference was not quite significant (browsing history, $p = 0.1217$; one-tailed Fisher's exact test).

Exploring how to limit data collection

Having a third choice besides "all" or "nothing" increased participants' willingness to install the application across all three applications we tested. This is promising, but how would users like to limit data collections? We were particularly interested in seeing whether their suggestions could lead to concrete guidelines for application developers. In the following, we report on a qualitative analysis of the answers we collected in the *Sharing* survey to the question "How would you want to limit data collection? Please be as specific as possible. You could give examples of the *Resource* you would not want to share and why, or you could describe the general circumstances that you would not want to share."

Location

For the location-aware application, 110 out of 278 participants in the *Sharing* survey chose to install Fill Me Up! with the option to limit data collected. About half of these participants (47% of 110) said they would like to *manually* control data collection, either by limiting the application to access location only when running ("I would only want the app to track my location when I am running it") or by manually turning location tracking on and off ("I would just like to have the option to turn off the data collection when I didn't want it on").

The second most popular answer (11% of 110) was a *white (or black) list* option such that specific locations or routes could be included (or excluded) consistently. One participant said, "I would only want (to share) my route to and from work and that is all." A few respondents described wanting the option to share location at a reduced granularity (5% of 110, e.g., "I typically don't allow any app to record my location, so a zip code/address feature would be nice."), only during specific time intervals (4%, e.g., "provide an option to turn on/off data collection at time interval the user chooses"), at a limited frequency (5%), or only when driving (4%).

Many participants not only specified how they would like to limit data collection but also said why. Out of the 110 participants who said they would install the application if they could limit data collection, 20% explicitly mentioned third-party tracking. For example, a user who was presented with the notice that the application "may share" the data said, "Not collecting personal info that could be sent to advertisers. That would be a deal breaker instantly. Also, the ability to turn on and off the location tracker". Some participants explicitly mentioned privacy or a concern about being tracked all the time (10% of 110), while others (8%) mentioned battery consumption. Finally, 6% emphasized that they would be willing to share their location with an application, but no other personal information (age, gender, contact lists and browsing history were named).

Email

For the email-based application, about one-third of the participants specified they would install BillKeeper if they could limit data collection (34% of 267 participants). The main reasons for doing so were limiting access to emails related to bills, having means to "[...] choose what goes in and out" or preventing personal emails from being accessed ("I have e-mails from friends that could be private and I don't want anyone else to find out"). About half of the participants suggested specific ways to restrict data sharing (46.7% of 90 participants): 23 said they would like to share with BillKeeper only emails from specific senders ("It would be great to be able to set what emails the app could or couldn't access. For example, you could have it set to collect all emails from Bank of America or ATT. Essentially, I would want to blacklist every other email received that wasn't from an address on the whitelist which the user could establish.") and 15 said they would manually filter emails (one participant said he would manage this by maintaining a folder for bills).

Interestingly, seven participants expressed the desire to redact private information from the emails shared with BillKeeper: specifically username, passwords, contacts' information, account numbers, payment amounts, phone numbers, addresses and bank statements. Unlike location, where the whole object (typically in GPS format) is shared with the app, emails are complex objects where the ability to redact specific content is important. In general, participants explicitly mentioned concerns about privacy (17% of 90), third-party tracking (12%), and battery or data plan consumption (3%).

Web Browsing history

For the web browsing-based application, about one-third of the participants specified that they would install NewsRecommender if they could limit data collection (37% of 252 participants). Many of these participants explicitly mentioned privacy and third-party tracking concerns (29% and 9% of 94, respectively). They specifically mentioned personal information that they would like to protect, such as username/passwords, emails viewed in a web browser, social media accounts (e.g., Facebook, Twitter, or Pinterest), banking, name, home address, phone number, adult content, and credit card information.

A few participants were less privacy concerned, but more interested in sharing with the application only relevant sites and avoiding spam (13% of 94, "I don't want to be spammed for something I looked up once and was never particularly interested in"). Finally, a few participants (14% of 94) provided specific suggestions on how they would like to control sharing of web browsing data. Their suggestions included a white (or black) list of sites, manually selecting websites to share with the application, a private mode (e.g., "a 'privacy mode' or something like that would be ideal; an option I could flick on where it would stop tracking what I'm reading and aggregating it. Sometimes I want to just read a trash article and not have it think that's what I want all the time")

or a sharing mode that depended on the browser used.

DESIGN GUIDELINES

We now discuss how our findings contribute to the design of a permission model for personalized mobile applications. To enable users to make an informed installation decision, the permission model ought to meet the following two requirements.

- Provide proper notice of the aspects of data collection and sharing that influence users' privacy risk assessments.
- Provide options that allow users to limit the collection of certain data that is deemed sensitive by individuals.

Although these requirements seem obvious, existing permission models fall short in both, and our study demonstrates how permission models lacking these requirements impact users' willingness to install applications. Especially, we show that the absence of certain information in the notice screen can misguide users: many users seem to have inaccurate presumptions about third-party data sharing and how often and when their data may be accessed by applications, and the status quo does not correct this misconception. These misconceptions can hurt not only (optimistic) users who would not have installed some applications if notified about their privacy implications, but also (pessimistic) users who would have installed applications that in fact properly protect their information. Furthermore, we show that applications that require blanket access would turn some users away from installing them.

The following sections present specific design guidelines to address these two shortcomings. Although our suggestions focus on a permission model for personalized mobile applications, they could be applicable to other personalized services such as ones running on social networking sites such as Facebook.

Proper Privacy Notice

Although it remains challenging to get users' attention to read privacy notices (or the permission request screen on Android) [13, 8], it is still necessary to provide proper notice when users decide to install an application. To make such notice effective, we recommend providing access to information about factors that matter to users' installation decisions (such as factors identified in previous work [15]). Our study identifies two additional factors that are relevant to users:

- *Frequency of data collection*. We found that many users have misconceptions about when and how often applications can collect data, and that these misconceptions influence user decisions. The frequency at which applications may take location samples affects how accurately applications can infer the users' location trail and, ultimately, the privacy risks. We recommend that designers look for ways to help users better understand the frequency of data collection, especially for highly personalized applications that rely on continuous access to users' location data, as these may pose the greatest privacy risk. Even for other data types that sampling rates do not apply (e.g., email and browsing history), it may still be useful to remind users that their data is continuously collected by applications.
- *Third-party data sharing policy*. The privacy risks increase as users' data is spread to more entities. People rightly have concerns about their data being shared with third parties. Hence, we recommend that privacy notices should highlight whether applications share data with third parties or not. However, we have not explored whether the level of risk or user concern depends on who the data may be shared with or how it may be used (e.g., advertising companies vs. analytics services) and this may be an interesting direction to study in the future.

Meaningful Choices to Limit Data Collection

Existing permission models allow an application to have unlimited access to a particular resource once the permission request is approved by users. Our studies suggest that this policy appears to violate user expectations.

One alternative approach would be to limit data collection so that, by default, data can only be collected when the application is actively being used by the user (i.e., the application is in the foreground). This is the behavior expected by the majority of the respondents in the *Control* condition, in the absence of an explicit notice about the frequency of data collection (58.4% of 296; 95%CI: [54%, 62.6%]). However, this approach will limit the capabilities of personalized applications that require continuous access to data. For such applications, it may be possible to provide users with several options for limiting data collection:

- *Content-based access control*. Some mobile devices (e.g., the iPhone) offer settings to turn on or off location sharing per application. Similarly, web browsers offer incognito or private browsing mode to allow users to limit tracking when activated. However, these settings require manual intervention, which might be burdensome. An alternative might be to allow participants to specify sharing rules based on the content of information (e.g., blocking websites that contain home address, phone number, or adult content or only allowing my route to and from work). Supporting this type of access control might require content analysis and accurate classification, though some control rules would be easier to implement such as only allowing email from specific senders to applications.
- *Label-based access control*. Some participants suggested access control rules based upon "labeling" activity by the users. For instance, people are used to "checking-in" their location to explicitly share it and to tagging or moving email to folders to organize it. Rather than requiring users to specify access control rules up front, another alternative might be to leverage these user-provided labels and allow users to limit disclosure to only data that has been explicitly categorized in some way.

In addition to providing different data control modes, there are many challenges in allowing users to limit data collection. For instance, the controls must be intuitive, so they can

be used with no prior knowledge. Also, overly restrictive settings may degrade the application experience, so users may need guidance on how to configure these settings so the app remains useful, while not compromising their privacy. We leave these challenges as an interesting avenue for future work to explore.

LIMITATIONS OF OUR STUDY

Although our results suggest that users need more choices, one limitation of our method is that we collected self-reported data. Our participants did not install the applications on their smartphones, rather they reported what their responses would be, which may have affected their answers. Another drawback of our method is that our participants did not experience the personalized services described in the study. Users might be more willing to share personal data with a service they find valuable. It would be interesting to investigate the long term effects of disclosing the frequency of data collection and the third party sharing policy.

We designed the survey such that participants were asked to place a greater importance on privacy concerns than they normally might have. We also asked users to make installation decisions in a situation where they did not have access to many of the features that would typically influence the decision: reviews, ratings, price, or application developer. Finally, in our study, we purposefully focused on applications that requested only a single data type, in the scope of personalized services, it's more realistic to consider an application that combines more than one source of personal data.

CONCLUSION

Our study reveals shortcomings of existing permission models. Our exploration with highly personalized applications that rely on continuous access to privacy-sensitive information such as location and email shows that a new permission model is even more critical for future applications. While the limitations of our study prevent us from knowing definitively whether our participants' preferences would extend to real-world situations, we believe our findings would move us closer to a permission model that allows users to mitigate privacy concerns while enjoying highly personalized services.

REFERENCES

1. S. Barocas and H. Nissenbaum. On Notice: The Trouble with Notice and Consent. In *Proc. of the Engaging Data Forum*, 2009.
2. P. Bernier. Qualcomm Gimbal to power personalized Star Trek smartphone experiences. `http://www.tmcnet.com/tmcnet/ces/articles/322428-qualcomm-gimbal-power-personalized-star-trek-smartphone-experiences.htm`, January 2013. Accessed: Mar 17, 2013.
3. R. Böhme and S. Köpsell. Trained to accept?: a field experiment on consent dialogs. In *Proc. of CHI '10*, pages 2403–2406. ACM, 2010.
4. A. J. B. Brush, J. Krumm, and J. Scott. Exploring end user preferences for location obfuscation, location-based services, and the value of location. In *Proc. of UbiComp '10*, pages 95–104. ACM, 2010.
5. S. Consolvo, I. E. Smith, T. Matthews, A. LaMarca, J. Tabert, and P. Powledge. Location disclosure to social relations: why, when, & what people want to share. In *Proc. of CHI '05*, pages 81–90. ACM, 2005.
6. S. Egelman, A. P. Felt, and D. Wagner. Choice Architecture and Smartphone Privacy: There's a Price for That. In *WEIS '12*, 2012.
7. A. P. Felt, S. Egelman, and D. Wagner. I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. In *Proceedings of the 2nd ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*, SPSM '12, pages 33–44, New York, NY, USA, 2012. ACM.
8. A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner. Android permissions: user attention, comprehension, and behavior. In *Proc. of SOUPS '12*, pages 3:1–3:14. ACM, 2012.
9. Forbes. Move over Siri, Alohar wants to learn everything about you. `http://www.forbes.com/sites/ryanmac/2012/03/23/move-over-siri-alohar-wants-to-learn-everything-about-you/`, 2012.
10. N. Good, R. Dhamija, J. Grossklags, D. Thaw, S. Aronowitz, D. Mulligan, and J. Konstan. Stopping spyware at the gate: a user study of privacy, notice and spyware. In *SOUPS '05*, pages 43–52, 2005.
11. C. Jensen and C. Potts. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *CHI '04*, pages 471–478, 2004.
12. P. G. Kelley, M. Benisch, L. F. Cranor, and N. Sadeh. When are users comfortable sharing locations with advertisers? In *CHI '11*, pages 2449–2452, 2011.
13. P. G. Kelley, S. Consolvo, L. F. Cranor, J. Jung, N. Sadeh, and D. Wetherall. A conundrum of permission: Installing applications on an android smartphone. In *Proc. of USEC '12*, 2012.
14. P. G. Kelley, L. F. Cranor, and N. Sadeh. Privacy as part of the app decision-making process. In *CHI '13*, 2013.
15. J. Lin, S. Amini, J. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. In *Ubicomp '12*, 2012.
16. H. Nissenbaum. A Contextual Approach to Privacy Online. *Daedalus 140*, (4):32–48, Fall 2011.
17. N. M. Sadeh, J. I. Hong, L. F. Cranor, I. Fette, P. G. Kelley, M. K. Prabaker, and J. Rao. Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing*, 13(6):401–412, 2009.

18. K. Tang, J. Hong, and D. Siewiorek. The implications of offering more disclosure choices for social location sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 391–394. ACM, 2012.

19. TechCrunch. Intelligent, context-aware personal assistant app Friday makes its public debut. `http://techcrunch.com/2012/07/20/intelligent-context-aware-personal-assistant-app-friday-makes-its-public-debut/`, 2012.

20. J. Wiese, P. G. Kelley, L. F. Cranor, L. Dabbish, J. I. Hong, and J. Zimmerman. Are you close with me? are you nearby?: investigating social groups, closeness, and willingness to share. In *Proc. of UbiComp '11*, pages 197–206. ACM, 2011.