Privacy Tipping Points in Smartphones Privacy Preferences

Fuming Shih[*], **Ilaria Liccardi**[*†], **Daniel J. Weitzner**[*]
{fuming, ilaria, djweitzner}@csail.mit.edu

[*]MIT CSAIL
Cambridge, MA, USA

[†]Oxford e-Research Center
University of Oxford, UK

ABSTRACT

The aim of this research was to understand what affects people's privacy preferences in smartphone apps. We ran a four-week study in the wild with 34 participants. Participants were asked to answer questions, which were used to gather information about their personal context and to measure their privacy preferences, by varying app name and the purpose of data collection.

Our results show that participants shared the most when no information about data access or purpose was given, and shared the least when both of these details were specified. When just one of either purpose or the requesting app was shown, participants shared less when just the purpose was specified than when just the app name was given.

We found that the predominant factor affecting users' choices was the purpose for data access. In our study the purpose varied from being not specified, to vague, to being very specific. Participants were more willing to disclose data when no purpose was specified. When a vague purpose was shown, participants became more privacy-aware and were less willing to disclose their information. When specific purposes were shown, participants were more willing to disclose, provided the purpose for requesting the information appeared to be beneficial to them, while participants shared the least when the purpose for data access was solely beneficial to developers.

Author Keywords

Privacy Preferences; Android; Experience Sampling

ACM Classification Keywords

K.4.1 Computers and Society: Privacy

General Terms

Theory, Human Factors

INTRODUCTION

With the pervasive use of smartphones and advances in sensor technology, context-aware services are an integral part of consumers' daily life. However, while extremely useful in serving and guiding us in our daily activities, smartphones can also be used to silently collect data about us, allowing app companies to create digital dossiers for services like targeted advertisements, extrapolate behavioral patterns valuable for market research. Due to these reasons, smartphones might pose great privacy risks to consumers. Users today have to make decisions regarding disclosure of personal information without being fully aware of the privacy implications behind data collection [11]. This is despite the fact that consumers consider their mobile data as private and are strongly opposed to apps collecting their information from their smartphones and tracking their locations without their consent [33].

Different solutions have been implemented to give users more awareness of what information is collected about them [15]. However, individuals have no way to understand how their data is actually used. A recent report released by the White House addresses these concerns, focusing on the idea of disclosing the range of possible uses for peoples' data [36].

Research [7] has shown that the perceived sensitivity of the collected data affects the willingness of people to disclose the data. This sensitivity is often context-specific and changes dynamically according to the individual's context, such as location and activities [5]. Users should not be expected to consent to the collection of their data for every foreseeable purpose, given the incomplete or missing information [22] they receive when making privacy decisions [2]. Users' privacy concerns rise when they find data collection happening in a context or for a purpose that is unexpected. Such an unpleasant experience leads to a sense of "creepiness" and results in a loss of trust [29]. To ease tensions revolving around data collection, app companies are advised to follow the principle of "respect for context" [35] when harvesting users' data.

In order to properly address users' privacy concerns while simultaneously allowing companies to use mobile data, people should have control of their data in terms of how services can use it, rather than simply granting *carte blanche* access [37]. We are interested in investigating users' choices and which circumstances might lead them to disclosure of their personal information in the form of location and activities. We wanted to investigate effects on privacy preferences by varying the application name and purpose of data collection within real contexts. In particular we are interested in these questions:

1. Do different purposes affect users' willingness to share their data?

CHI 2015, April 18–23, 2015, Seoul, Republic of Korea.

http://dx.doi.org/10.1145/2702123.2702404

2. Are people more likely to share their personal data with apps that they frequently use?
3. Does the type of context (location or activity context) affect people's willingness to share their data? If so, what are the effects and what are the specific contexts?

We applied the Experience Sampling Method (ESM) [8] to elicit responses from users to disclosure of their personal information in the form of their location and activities throughout the day. We developed a framework that creates survey questions customized to an individual's app usage and personal context. We conducted a four week in-the-wild experiment that would prompt users each hour during the day, from 9am to 10pm, asking them to describe their personal context (captured location type, activity context, social surroundings and disclosure preferences), followed by questions asking about their willingness to share contextual data (location or activity type) with familiar apps (gathered from their smartphones) for seven different types of purposes. Purpose types varied from being not specified, to vague, to very specific. Specific purposes varied from being beneficial solely for users, to beneficial solely for developers, to beneficial for both. The main findings were:

1. The more information provided to participants, the more reluctant they were to share personal information. Hence, when presented with *no information* about purpose or data access, participants were willing to share the most.
2. The type of purpose affected a large number (15) of participants' sharing behaviors. Participants were more willing to share for purposes that were beneficial solely to them.
3. The type of app, but not the frequency of use, affected participants' sharing behavior. Participants might share more with less frequently-used apps.
4. Participants were affected in their disclosure behavior according to the type of location they were in.

RELATED RESERCH

Research has shown that many smartphone users lack the knowledge needed to perform changes in privacy control settings and mistakenly trust that the app will protect the privacy of their data [27]. Only few users actually read and understood the implications of permissions requested by Android apps [11], [23]. Many people may still hold unrealistic beliefs about how their data should be treated, and consider information on their smartphones to be private and overwhelmingly reject data collection [33]. Users' privacy expectations often do not reflect current practices. Many apps transmit sensitive data to third parties [10] that users intend to only use on-device.

Given the disparity between users' expectations of privacy [28] and the opaque practices of data collectors [22], researchers have sought to address the privacy issues in the context of consumers' experiences when using apps, by making privacy part of the app selection decision [16], by improving their privacy expectations [24] or by exposing data leakage [6].

Researchers have found that people are willing to disclose their personal information for short-term benefits [2], for perceived beneficial gain (such as monetary rewards) [3], and for high relevance and need for the service provided [13].

Past research has shown that people's privacy preferences for sharing their personal information are *contextual*: sharing current whereabouts [9] [32], [38], updating statuses on social networks [1], [25], configuring context-aware services [17], [19]. In the context of disclosing information to apps, one important factor can be users' trust of data collectors [14]. Knijnenburg and Kobsa found users' willingness to disclose contextual information to recommendation services varied when presented with different justifications for data collection [18]. Tan et. al found developer-specified messages actually affected users' decisions for allowing data access [30]. Their results showed that users were more likely to approve data requests when the purpose for requesting data was displayed.

To model users privacy preferences, Lin [24] used crowdsourced surveys to measure the unexpectedness of certain data accesses by the apps, allowing users to rate apps by comparing their perceived functionality with the actual permissions requested by the app. Toch [31] used a crowdsourcing method to predict each user's privacy preferences for location sharing. While crowdsourcing makes their approach more scalable, crowd opinions only represent users' *a priori* preferences of how an app should work. The results might not reflect users' practical privacy concerns when they actively engage in making "privacy vs. benefit" trade-offs [12].

USER STUDY

We conducted a user study[1] in the wild over a four week period to understand and measure users' privacy preferences toward the sharing of their personal location and activity context with familiar apps for common types of currently-used purposes. We collected their most recent location and enquired about their activities, while also collecting information about apps running on their smartphones. We asked participants to answer a set of *personalized* questions each hour. The questions were personalized based on their location and current activity as well as apps that were familiar to them (already installed). A set of questions contained two segments, and within each segment questions were randomized. The two segments were designed to:

1. **Gather users' personal context** in order to first understand each participant's current state, i.e. where they were (home, work etc.), who they were with (friends, family, colleagues etc.), what were they doing (working, leisure etc.) and who were they willing to share this information with (friends, family, everybody etc.)
2. **Measure users' privacy preferences** in order to understand if the app, the type of purpose and/or the type of location (for example whether they were home or at work) or activity (whether they were working, going somewhere

[1] The study protocol, the rules which participants needed to follow were approved by IRB to ensure subjects' privacy. These details were clearly communicated with participants during information session and documented in the consent form.

etc.) affected participants' willingness to share their personal contexts.

The number of questions varied between sets according to the factors that were present and the way they were shown.

Study Design

Gathering users' personal context

In segment 1, the questions were designed to understand participants' current context in the form of the type of location (Figure 1 (a)), which activity they were currently engaging in (Figure 1 (b)), who was around them (Figure 1 (c)) and who they would have been comfortable sharing the activity they were currently engaging in (Figure 1 (d)). Since we used the answers provided by participants from the activity type question (Figure 1 (c)) within the social sharing attitude questions (Figure 1 (d)), the former would always precede the latter.

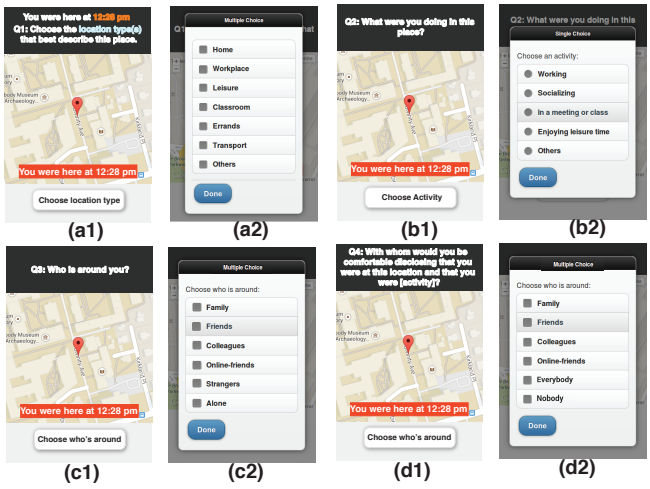

Figure 1. User study questions designed to gather users' personal context in the form of location type (a); activity type (b); social surrounding (c); and social sharing attitudes (d)

Measuring users' privacy preferences

Questions in segment 2 were designed to measure users' privacy preferences. In particular we were interested in testing familiar apps as well as common purposes that have been given as reasons to collect users' data. In order to measure their preferences, we asked questions in which three factors (personal context, app name and purpose) were randomly and evenly selected. Figure 2 shows how the question was presented to participants and where each value condition would appear.

Personal context was set either to the user's current location or the type of activity that the user was currently performing at this location.

App name condition was gathered from the apps used by each user. Usage of each app was measured in the first week of the experiment. According to the amount of usage (low, medium, and high), three apps were randomly selected to be used throughout the study. App usage was calculated individually for each user according to their own level of usage. Questions were also posed with the app name omitted, for a total of four questions.

Purpose condition was randomly selected from six different types of purpose commonly cited for accessing users' personal data. A question was also asked when no purpose was inserted, hence this field was blank (for a total of seven conditions). The types of purposes used were:

- **Vague Purposes**:
  - Without purpose: nothing is displayed;
  - Captures information: "so that the app has your information";
- **User-focused purposes**:
  - Testing needs: "for testing new features";
  - Improving experience: "to improve the app experience for you";
- **User- and developer-focused purposes**:
  - Advertising: "to be used to display personalized ads relevant to you";
- **Developer-focused purposes**:
  - Profiles: "so that the app can learn your daily patterns, to profile you for market research";
  - Revenue needs: "so that it can sell this information and make money".

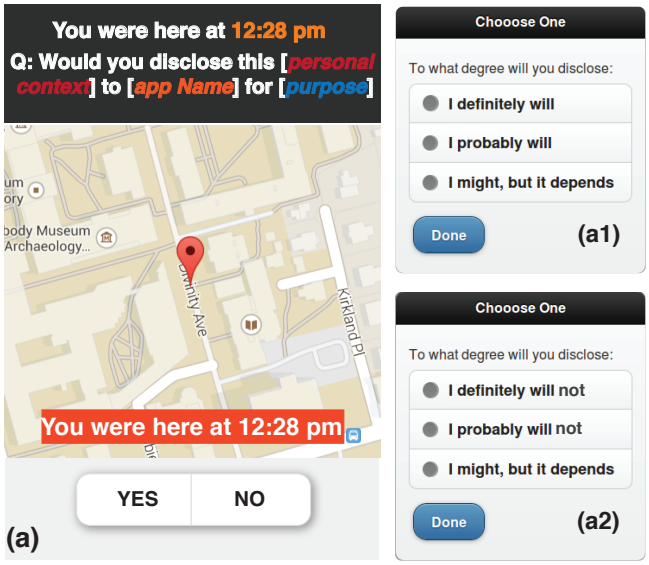

Figure 2. User study questions designed to gather users' willingness (or unwillingness) to share their personal context. Each participant had to choose YES or NO. Afterwards they had to give a rating of positive (a1) or negative (a2) feelings

Randomization of privacy preference questions

To ensure that all factors were collected in an even manner during the *measure users' privacy preferences* part of the study, we introduced a scheme to randomize the appearance of different values for each factor. The privacy preference section of the questions was randomized between two conditions: the *app-specific* and the *purpose-specific* condition.

When the *app-specific* condition was chosen, all seven predefined purposes were shown (including purpose string omitted). First, the study app randomly chose a personal context

- either a location or activity to be used. Then one app name was randomly selected from a predefined list of app names based on each participant's individual frequency of use[2], This app name was shown for each of the seven pre-defined purposes, for a total of seven questions. The purposes appeared in a random order from question to question.

When the *purpose-specific* condition was chosen, the study app randomly selected one purpose from the list of seven purposes, then showed four choices for app names (including the no app option) for both location and activity context (appearance was randomized between questions) for a total of eight questions. The app names appeared in random order from question to question.

The reason for randomly selecting location and activity within the *app-specific* question condition was to avoid overwhelming the participant with an additional seven questions within this set. Similarly, we decided to show both location and activity within the *purpose-specific* condition to present a similar number of questions.

During week 1, information about app usage was collected. Hence, in week 1 when the *purpose-specific* condition was chosen, the study app would show two questions for disclosing activity and location context, with the app name omitted. Table 1 shows that the factors were evenly distributed (please note that *no app* name is larger in to week 1).

Apparatus

We developed a framework to conduct the study. The framework contained two parts: an Android app (*ContextProbe*) used to collect sensor data and to prompt participants with survey questions, and an application server that communicated with the app, received and computed data, then sent new survey content to the app.

ContextProbe app

The *ContextProbe* app has four main modules: 1) data collector, 2) survey, 3) uploader, and 4) message receiver. The data collector module was implemented using a mobile sensing library called *Funf* [4] that allowed the app to schedule data collection tasks periodically. The message receiver used Google Cloud Messaging (GCM) technology to receive pushed notifications from the application server. This allowed for the content of questions to be changed in real-time and tailored according to participants' collected data.

Application server

The application server was developed using Google Apps Script[3], which provides integration across multiple web services on the Google Cloud platform. The application server can run and operate directly on the experimenter's Google Drive space without a complicated setup process. The data handler module on the application stored data uploaded from participants' smartphones in a Google Spreadsheet for easy access and data analysis. The context processor read these data files to compute app usage patterns for each user and used this information as the new content for customizing each participant's survey questions. The cloud messaging module sent GCM contents directly to each participant's phone to update their survey questions.

Framework interactions to create personalized surveys

Figure 3 gives an overview of the architecture of the user study framework and shows step-by-step how different elements interacted with each other to create a customized survey, tailored to each participant's app usage pattern and up-to-date contextual information about their location and activity:

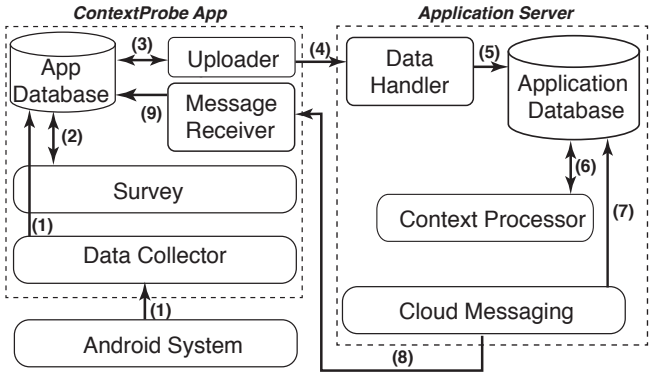

Figure 3. Architecture of the user study framework showing the ContextProbe App and the supporting application server, and how different elements interact in creating a customized survey.

The Data Collector module runs as a background service within the ContextProbe app. Every 15 minutes it collects information about users' locations and continuously listens to system events related to running apps. Every time an app is used, sent to the background or switched between, the event is collected (step 1). Every hour the Survey Module queries the database to gather users' most recent location and captured time. This information is used by the survey module to create survey questions on a map (step 2). The Data Uploader module exports the collected data from the database each hour, including location data, information on running apps, and users' answers, and then uploads data to the remote server (step 3 and 4).

On the application server, the Data Handler module aggregates data collected from different users' smartphones (step 5). The Context Processor module extracts information related to each user's app usage patterns from the collected data on running apps on the users' phones, and identifies candidate apps based on the frequency that each app is used by the user (step 6). The Cloud Messaging module fetches information about the candidate app names for each user and pushes the app names to users' phone (step 7 and 8). The Message Receiver module saves the pushed content, which is used by the Survey Module later to compile new survey questions (step 9).

Procedure

One hour information session

Each participant had to attend a one-hour information session prior to the start of the four week study. This session was used

[2]The app name was chosen from a list of apps divided according their usage frequency; this was based on the captured and collected apps during the first week of the study as aforementioned

[3]https://developers.google.com/apps-script/

to explain how the app worked, to install the app on their smartphone, to allow them to sign the consent form and to familiarize them with the study requirements and respective remuneration.

User Study Requirements

Each participant had to install the *ContextProbe* app. Each user participating in the study had to answer a minimum of 10 sets of questions per day. Each set contained two parts: 1) four initial questions, and 2) 7-8 questions depending on how the questions were randomized. The appearance of the questions was randomized between each part for each set. Users needed 2-5 minutes to read/answer each set of questions i.e. 28-70 minutes a day.

The first set was shown at 9:00 am, the last set at 10:00 pm, for a total of 14 sets a day. A notification was used to alert participants that a new set of question needed to be answered. Participants had up to three hours after the set was first shown to answer all of the questions. We chose to show a set of questions for up to three hours, because we wanted to make sure participants clearly recollected their context. Participants could view the ContextProbe home page to see sets that needed to be answered (marked in green), and sets that had expired, i.e. were not answered within the allotted 3-hours time frame (marked in red). The home page also showed greyed-out buttons indicating the next sets that were not yet available to be answered. If participants did not respond to a minimum of ten sets per day, an automatic warning email was sent to them and the researchers, to inform them that they were not complying with the requirements and that they could be dismissed. If this happened one additional time, the participants and the researcher would both receive an email dismissing the participant from the study. Participants were asked to answer questions as honestly as possible. They were notified in the info session that if random clicking was detected either during or after the study, they would not be paid.

Trustworthiness of the data

Because participants were asked to tag their locations, we stored latitudes and longitudes of these locations and checked whether the tag provided by the participant was the same for the same location, accounting for GPS inaccuracy with a bounding box. We also checked for consistency in context tags (e.g *errands* for the activity label together with *in a class* for the location label. Consistency was checked when the same privacy question was given in the same location. We had a script that automatically checked participants' answers. Participants whose responses appeared to be inconsistent (with a lack of reasonable motivation) were removed.

Semi-Structured Personalized Survey

After analyzing the data for each participant, we enquired about qualitative motivations and reasons for sharing or not sharing context information. We conducted these conversations over email. In the emails we included screenshots of the location, location types, app names and if necessary, type of purpose. Semi-structured surveys were conducted after analyzing the quantitative data. Five different surveys were created to accommodate the five different privacy profile groups. We asked 10-25 tailored questions based on users' answers throughout the four week period.

- Privacy-conscious: Why did you not disclose data at all? From the apps collected from your phone, [app-name] has access to [personal data], did you know?
- Purpose-driven: Why did you not disclose your data to [app-name] for [purpose]? Why did you disclose your data to [app-name] for [purpose]?
- Trust-based: Why did you not disclose [personal data] to [app-name] when you did disclose it to [app-name]?
- Privacy-indifferent: Why did you disclose your data to all apps?
- Location-sensitive: Why do you not disclose your data at [location] rather than [location]? Do apps or purpose not affect you when you are at this [location]?

We looked the answers both between each group and individually. The quotes we present in the paper are examples that reflect common and interesting responses.

RESULTS

Participants

We solicited participation in our study using internal mailing lists and CraigsList announcements. Participation in the study was voluntary and each participant received $120 after completing the whole study. 61 participants attended the one hour information session. Out of these people, three participants did not have an Android device capable of running the app and seven participants never started the study. A total of fifty-one people started the study. Fourteen participants were removed from the study; of these, twelve participants were asked to leave (5 participants in week one; 4 participants in week two and 3 participants in week 3.). Two participants were removed because their responses fell below the allotted threshold of 10 sets per day. Ten participants were removed because their answers were not consistent and they could not provide reasonable explanations and motivation for the inconsistencies. The remaining two participants were asked to stop the study in week one because we could not receive any new GPS location data, even though they reported being in different locations. Thirty-seven participants participated in the entire study. Three participants' responses were removed because due to their type of Android device, information about the apps running on their smartphone could not be collected for the study (they received payment); in addition, one had problems receiving questions due to a poor Internet connection around his neighborhood. Thirty-four people successfully completed the study. Of these, 18 were male (av. age = 25) and 16 female (avg. age = 29). Level of educations varied from having completed high school (4), two-year college degree (7), being an undergraduate student (2), completed four-year college degree (10), completed master degree (4), being a graduate student (3) to advanced graduate work or completed Ph.D (4).

Responses

Over the period of four weeks we gathered 74,713 total responses from the 34 participants (Table 1) for four different conditions (Table 2). We collected 4,212 responses for conditions where *no app name* and *without purpose* were displayed; 25,452 responses for conditions where app names were not displayed, but purposes were displayed (excluding the *without purpose* condition); 6,352 responses where all app condition (excluding *no app name*) were displayed, but *without purposes*; and 38,697 responses where all app names (excluding the *no app name* condition) and all purposes (excluding the *without purpose*) condition were displayed. Questions were evenly distributed for each factor (Table 1).

Table 1. Number of participants' responses grouped by app frequency types (no app name condition is included) for each purpose condition (no purpose condition is included). The total number of responses for each and across conditions is also shown.

TYPES OF PURPOSES	No App	Least Used	Medium Used	Most Used	Total responses
Without Purpose	4,212	2,138	2,069	2,145	**10,564**
Captures Information	4,234	2,196	2,132	2,198	**10,760**
Testing Needs	4,181	2,131	2,059	2,135	**10,506**
Improving Experience	4,278	2,184	2,114	2,186	**10,762**
Ads Needs	4,198	2,110	2,048	2,121	**10,477**
Profiles	4,298	2,224	2,154	2,223	**10,899**
Revenue Needs	4,263	2,182	2,118	2,182	**10,745**
Total responses	**29,664**	**15,165**	**14,694**	**15,190**	**74,713**

Users' privacy profiles

Participants in our study were found to share their personal context according to different factors: usage purpose, trust in the app itself and location type. From Figure 4 (a) & (b) we can see that six participants (P4[4], P15, P23, P27[4], P29[4], P30) shared mostly according to the app that they were using (with the exception of never sharing with any app only when specific purposes were specified). However, the majority of participants (fifteen participants: P3, P5, P6, P10, P11, P13, P14, P16, P17, P18, P19, P24, P25, P26, P33) shared according to the type of purpose (Figure 4 (b)) displayed in each question.

These latter participants did not share their personal information when the purpose specified did not present any conceivable gain to them. Of the remaining participants, eight demonstrated behavior of either always being willing to share their data (P1, P2, P21) or being extremely privacy-sensitive and never sharing (P8, P9, P12, P22, P28). The remaining five participants (P7, P20, P31, P32, P34) showed no effect according to the type of purpose or the app; a closer look at their data showed that they tended not to share their personal information when at particular locations. Among these participants, four participants tended not to share when they were at *home* and one participant tended not to share when at work. Our results highlight five distinct patterns of behavior with respect to privacy preferences:

- **Privacy-conscious users (3)**: These users did not tend to share their personal data. They were very conservative about when to share, and they do not share.
- **Purpose-driven users (15)**: These users shared their personal data according to different purposes. They tended to share data when there was a benefit to them.
- **Trust-based users (6)**: These users trusted certain apps and, no matter the activity or the purpose, would share the data.
- **Privacy-indifferent users (5)**: These users shared their personal data consciously because they saw no problem or harm in doing so.
- **Location-sensitive users (5)**: These users shared their personal data when they were not in a particular location. When they were at a specific location, they tended to never share their location information or the activity they were currently undertaking.

Usage and Collection effects on sharing preferences

Participants in the study tended to share more ($\mu = 3.12$; $\mu_{rank} = 3.04$) when there was no information on the purpose for data collection or which app was collecting it (combination 1) (Table 2). When details were given, such as name of the app requesting the information and the purpose for which it would be used, users tended to share the least ($\mu = 2.71$; $\mu_{rank} = 1.60$; combination 4). When just one of either the purpose or requesting app was shown (data purpose versus data collection), participants tended to share less than when data purpose was specified ($\mu = 2.86$; $\mu_{rank} = 2.50$; combination 3) rather than when the app requesting the data was specified ($\mu = 3.02$; $\mu_{rank} = 2.85$; combination 2) (Table 2).

Table 2. Descriptive statistics and Friedman test for each factor: combination of absence and/or presence of purpose and app name. The mean is calculated using a 5-point Likert-scale (ranging from 5-Will Definitely share to 1-Definitely NOT share). The Friedman test is significant Chi-Square (3, N of Participants = 34) = 25.295, $p < .0001$. Kendall's W is 0.248, indicating fairly strong differences among the four combinations.

	COMBINATIONS APP NAME	COMBINATIONS PURPOSE	Num. of Responses	Mean (μ)	Std. Dev (σ)	Mean Rank (μ_{rank})
G1	✗	✗	4,212	3.12	0.61	3.04
G2	✓	✗	6,352	3.02	0.65	2.85
G3	✗	✓	25,452	2.86	0.69	2.50
G4	✓	✓	38,697	2.71	0.69	1.60

The Friedman test shows a statistically significant difference in all four groups (Table 2). Post hoc analysis with Wilcoxon conducted with a Bonferroni correction with $p < 0.0083$ ($p = 0.6/6$) shows statistical significance to users' sharing preferences in G1 vs. G3 (Z=-3.198; $p < .001$), G1 vs. G4 (Z= 3.676; $p < .000$), G3 vs. G4 (Z = -3.258; $p < .001$), and G2 vs. G4 (Z = 3.771; $p < .000$).

[4]These participants did not share their personal information when the purposes displayed within the question did not show a conceivable gain to them. Otherwise they based their decision on the app. Even when the purposes displayed within the questions were tailored to have a clear and conceivable gain for them, they still would not share their personal information with certain apps.

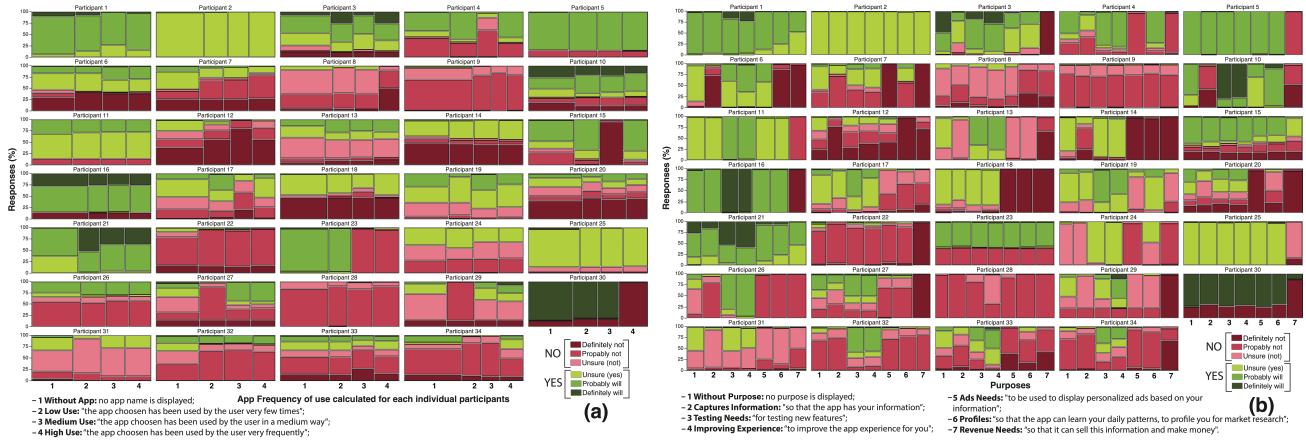

Figure 4. Overview of participants' responses regarding willingness or unwillingness to share their personal information with apps, in the form of either their location or their activities, for six different types of purpose, as well as no purpose.

App effect: Does the frequency or the type of app matter?
Six participants based their decisions to share or not share their personal information on the app they were using.

Frequency of use does not affect willingness to share data, with some participants choosing to share more with less frequently-used apps, rather than the most frequently-used apps. For these participants, the reasons for not wanting to share depended on the type of app. P4 explained that he was afraid that his sensitive information could be used for other purposes by the app (web browser app):

> P4: *"I tend to type in a lot of personal stuff via [browser's name] that I don't want used. That is why I denied its access to my information."*

While P15 was found to be reluctant in sharing personal information with an app (dating app) that had already collected and stored a lot of personal information:

> P15: *"I was just getting a little aggravated with the site [Name], it already had a ton of my data, so no more!"*

P4 in addition tended not to share his personal information (with any type of app, including when the app name was not displayed) when purpose 6 (targeted advertising) and purpose 8 (revenue needs for company) were shown.

P27 and P29 tended not to share their personal information with their least-used apps, however the reasons for these choices are different. For P27, frequency of use and hence familiarity with the app (diet app) was an important factor. P27 explained that once he has established a relationship with an app, he becomes more willing to disclose his personal data, for certain types of purposes that are beneficial to him:

> P(27) *"[...] I didn't use the app, I was not able to establish a relationship with it"*

While P29 explained not wanting to share based on a distrust in the app (banking app) itself:

> P29 *"I feel it's an extreme invasion of privacy letting my bank know where I am during the day. My phone is for leisure, for fun [...]. The less the bank knows about me, the better."*

Similarly to P4, P27 and P29 also tended not to share with any type of app (including when no app name was shown), when purposes did not show any conceivable gain to them (purpose 2,5,6,7).

P23 and P30 instead were unwilling to share their personal data with their mostly frequently used apps. P30's most frequently used app was a social network app and similarly to P15, the reason behind not wanting to share was to avoid giving additional personal data on top of the data already collected and stored by the app. P23 explained concerns that his former employer (bank app developer) would be able to use his personal data:

> P23 *"I used to work for [Company Name] and I would just not want them to have any of my information."*

Participants did not base their decision to share or not share with specific apps on the frequency of (and possible need and desire for) use, but rather on the app itself.

We found that participants denied access to their location to apps which actually requested the location permission on installation. This means that these apps can already collect this data on the user's location. These participants were not aware that this could be happening, underlining the need for clearer and more specific regulation of data purpose rather than data access. Such conflicts are, however, not being addressed enough today by mobile platforms.

Purpose effect: Does the type of purpose matter?
Data usage and collection matters to users when deciding to share their information (Table 2). As we can see from Table 3, Figure 5 and Figure 4(b), the type of purpose affects users' decisions to share or not their personal context.

Participants showed willingness to share their personal context when there was a conceivable gain to them. Participants were prompted to share their personal data for purposes of *testing needs* and *improving app experiences*, their sharing preferences had a $\mu = 3.324$ and $\mu = 3.36$ respectively (Table 3). When the purposes specified were more beneficial for developers (ie. companies), the willingness of participants to

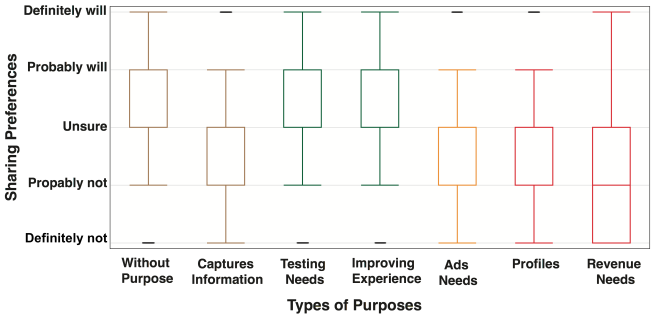

Figure 5. Box plot of participants' responses to sharing of their personal context (location and activity) grouped by different types of purpose (including no purpose).

share dropped, with $\mu = 2.60$ (*ads needs*), $\mu = 2.66$ (*gathering profiles*) and $\mu = 1.96$ (*revenue need*).

Table 3. Descriptive statistics and Friedman test regarding willingness or unwillingness to share, for different types of purposes. Purposes vary according to seven predefined purposes. The Friedman test is significant Chi-Square (6, N = 34) = 126.8, $p < .0001$. Kendall's W is 0.622, indicating fairly strong differences among the seven purposes)

	TYPES OF PURPOSES	N	Mean (μ)	Std. Dev (σ)	Mean Rank (μ_{rank})
1	Without Purpose	34	3.06	0.59	4.94
	Captures Information	34	2.74	0.71	3.28
2	Testing Needs	34	3.34	0.72	5.9
	Improving Experience	34	3.37	0.71	6.06
3	Ads Needs	34	2.61	0.88	3.29
4	Profiles	34	2.66	0.88	3.09
	Revenue Needs	34	1.97	0.91	1.44

When the purpose specified was non-explanatory or non-existent, participants' willingness to share their data differed (Figure 5), even though these two purposes described the same circumstance [20]. When the purpose was non-explanatory and vague, $\mu = 2.74$, while when the purpose was missing, $\mu = 3.05$. This underlines the fact that participants are more willing to share data when they are not aware of what their data is used for.

The appearance of a vague and non-explanatory purpose caused participants to disclose less compared with the results when the purpose was missing. Participants might have been alerted by vague purpose strings, indicating an intention to keep participants' data for other unknown uses (*so that it (the app) can have your information*). P10 and P6 underlined the importance of purpose information:

> P10 *"The purpose of the data collection is very important to me. If it is just collecting it to store, I would not be comfortable because I wouldn't know what it is doing with the data"*.

> P6 *"[...] whether or not to release my location depended on the other factors. [...] it's easy to say "I'm enjoying leisure time" sure, so are 1 billion other people, but having my location is a *lot* more specific, and so I'm less inclined to share that data [...] Purpose of use and nature of app are both extremely important factors in deciding whether or not to disclose the information.*

Similarly, P7 and P29 both had concerns about unknown uses of their personal data.

> P7 *"I am not about volunteering information to unknown sources, [...] Just because an app is particularly useful doesn't mean I would grant it a blank check to record and sell my personal data."*

> P29 *"I am more sensitive to disclosing data that may have personal information that can be intercepted or used without my permission."*

Effect of context on sharing preferences

While participants were highly affected by the type of purpose and the type of apps shown to them, we found five participants who displayed differences in their sharing behavior based on their location. In Table 4 we reported individual odds ratios representing each participant's willingness to disclose their location in each of the four popular reported places (*home*, *work*, *leisure* and *transport*). Three of these participants (P20, P31, P34) showed that when people were in locations other than home, the odds ratio of disclosing at the location is 2 times greater than when at home.

The remaining two participants (P7, P32) showed the opposite, ie. when their current location was among (*work*, *transport*, or *leisure* the odds of them not disclosing was 1.2 to 3 times larger than when at *home*.

Table 4. Repeated measured logistic regression with *location* context as the predictor variable and participants' disclosure preferences as a binary variable (0 = NOT DISCLOSE; 1 = DISCLOSE). The odds ratios for each context is displayed. The computation is performed per individual participant. The intercept value represents the bias of each participant to respond positively or negatively to disclose their information. The higher the value, the more likely they are to respond (irrespective of context). (***) $p < 0.001$; (**) $p < 0.01$; (*) $p < 0.05$; (+) $p < 0.1$.

PART. ID	INTERCEPT	*Home*	*Work*	*Leisure*	*Transport*
20	0.46***	0.54***	1.40*	1.34	1.05
31	0.51***	0.49***	2.78***	0.66*	1.11
34	0.22***	0.43***	0.94	0.96	0.98
7	0.66***	0.81+	N/A	0.22***	0.53***
32	0.35***	1.04	0.82	0.76	0.62*

However even among these five participants, purpose was a dominant factor affecting their disclosure preferences. When presented with beneficial purposes or no purpose, participants became more willing to disclose, even when the location was labeled as more private to them (e.g. less willing to disclose home for three participants and other locations for the two remaining ones).

Participants' qualitative remarks revealed how they considered these factors together. P20 noted that the sensitivity of disclosed information related to personal context, but also emphasized that purpose was the main factor for disclosing his information:

> P(20) *"I don't like having my home location or home activities available to any app, [...] while I don't mind anyone knowing if I'm running errands or at work. [...] Purpose of use was the most important thing - if it was just for the app developers benefit [...] I didn't want to do it. But if I could get something out of it [...] I was more willing to do it."*.

P7 thought it was fine to disclose his location if it is used appropriately:

P7 *"[...], I would be fine with an app knowing that I am relaxing in [coffee shop] or at home if it actually put that information to use"*.

This observation shows that while users' locations and activities might be helpful in governing the social norms of sharing information with other people [26], these factors are outweighed by purpose when considering disclosing information to apps.

In Android apps, users grant permission prior to installation, allowing apps to use personal information for a variety of unknown purposes. In this study, we have shown that participants based their decisions to share their personal information on the different types of information or a combination of factors: ie. purpose, app and location context. The all-or-nothing decision that the Android OS currently uses could therefore be improved to allow finer-grained control for popular visited locations.

DISCUSSION

Privacy awareness: vague vs. explicit

Our study highlighted the importance of showing *specific information* regarding information about data collection (*app name*) and data usage (*purpose*). As seen in Table 2 participants are more willing to share their personal context when none of this information is displayed. When the appearance of this information is alternated, participants shared less than when information about data usage (*purpose*) was displayed.

Presenting a vague and non-explanatory purpose caused participants to disclose less, compared to when a purpose was missing. Tan et. al [30] reported that users were more willing to disclose their personal information when a purpose was shown in the permission request. This disparity in results might be due to different approaches we took in probing privacy preferences. Tan et. al [30] used an online survey, showing screenshots of permission requests from real apps with hypothetical question about personal data, whereas our study was conducted in the wild with privacy preference questions about participants' real, personal and current contextual information tailored to specific apps that were really used by each participant. Subjects in Tan et. al's study [30] might not have been familiar with the apps shown in the survey, and each subject was given the survey questions once.

In our study, answers from each participant were collected using repeated measures designed to cover all conditions for different purposes. The main goal of conducting the study in the wild with repeated measures was to compare subjects' responses under different conditions, including their physical context (their locations and activities) and different purpose strings presented to them. The aim was to trigger privacy concerns that are more subjective and sample responses from real life situations.

One plausible explanation for the impact of showing a non-explanatory purpose is that privacy awareness was increased. Participants were reminded of the trade-offs [12] between the unpredictable costs (privacy risks) and benefits (functionalities) brought by apps, therefore becoming less willing to disclose personal information. This finding highlights the importance of specificity when describing purpose of data access, since a vague purpose can alert users to possible privacy risks and discourage them from sharing.

Developers are able to collect personal data about users, because to date, mobile platforms lack support for fine-grained control over data collection with specified purposes. Our study suggests that when any explanation, even if vague, is provided for the purpose of data collection, people do get alerted about privacy concerns and make different decisions compared to when no information is displayed.

Purpose matters: give me a reason to share

Differently from previous research [34][17][32] which found that users' preferences for disclosing their locations are contextual, our study demonstrates that users' decisions about whether to disclose context information are affected not only by the sensitivity of the disclosed information itself, but also by the purpose for collecting the data.

In our study, users' contextual information (current location, activity and social surrounding) did not have a significant impact (like purpose) on their privacy preferences for disclosing personal information to apps.

Our results have shown that when control is given to users, they tend to make more specific choices regarding their perceived benefits. For some users, preferences for sharing can be strictly app-specific or location-specific. However, even in these cases participants were unwilling to share their information for some purposes (beneficial to developers).

Our findings confirm that participants used the purpose string to separate data access from data collection and find clues to justify whether such data collection is reasonable or not. As shown in Table 3, the purpose string either increases acceptance of data collection or alerts the user to unfavorable privacy risks that can decrease the willingness of users to disclose personal information to apps.

We analyzed all 34 participants' apps and found that many (85%) of the apps that these users denied access to their location actually request the location permission on installation, meaning that these apps can collect users' locations at any time. This underlines the drawbacks and failings of the current practice of allowing any access to users' personal data without being able to specify and restrict the usage. Such conflicts, however, are not being addressed enough by mobile platforms and they should be addressed by regulations and legislation aimed at helping users to safeguard and protect their right to privacy.

Experiment Limitations

When using experience sampling to probe participants' privacy preferences, we were aware of the natural bias introduced by the time-based triggers as discussed in [21]. For example, *home* and *work* were the predominant places for location context. Therefore we carefully reported the results with appropriate statistical indicators such as odds ratios to show the effects of location context. Our understanding for user location and activity context depends on self-reported

data from participants. It is possible that some errors were introduced when annotating the locations.

Our study required the participants to respond to questions fairly frequently (once an hour). It is possible there could be a fatigue effect that caused decays in response rate or lesser quality of data as a result. Due to the monetary incentive and weekly removal of disqualified subjects, we found only a slight decrease in response rate in the final week (5%). They study was conducted with Android smartphone users and might not be completely representative for other smartphone operating systems.

Acknowledgments

Ilaria Liccardi was supported by the European Commission Marie Curie International Outgoing Fellowship grant 2011-301567 *Social Privacy*. Our thanks to Evan W. Patton and our anonymous reviewers for their help, suggestions and insights.

REFERENCES

1. Abdesslem, F. B., Parris, I., and Henderson, T. Mobile experience sampling: Reaching the parts of facebook other methods cannot reach. In *In Privacy and Usability Methods Powwow* (2010).
2. Acquisti, A., and Grossklags, J. Privacy and rationality in individual decision making. *Security Privacy, IEEE* (2005), 26–33.
3. Acquisti, A., and Grossklags, J. When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. In *WEIS* (2007), 12.
4. Aharony, N., Pan, W., Ip, C., Khayal, I., and Pentland, A. Social fmri: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing 7*, 6 (2011), 643–659.
5. Ahern, S., Eckles, D., Good, N. S., King, S., Naaman, M., and Nair, R. Over-exposed?: privacy patterns and considerations in online and mobile photo sharing. In *Proc. ACM CHI* (2007), 357–366.
6. Balebako, R., Jung, J., Lu, W., Cranor, L. F., and Nguyen, C. "Little Brothers Watching You": Raising Awareness of Data Leaks on Smartphones. In *Proc. of ACM SOUPS* (2013), 12.
7. Bansal, G., Zahedi, F., Gefen, D., et al. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems 49*, 2 (2010), 138–150.
8. Cherubini, M., and Oliver, N. A refined experience sampling method to capture mobile user experience. *Proc. of ACM CHI Workshop of Mobile User Experience Research* (2009), 1–6.
9. Consolvo, S., Smith, I. E., Matthews, T., LaMarca, A., Tabert, J., and Powledge, P. Location disclosure to social relations: why, when, & what people want to share. In *Proc. ACM CHI* (2005), 81–90.
10. Enck, W., Gilbert, P., Chun, B.-G., Cox, L. P., Jung, J., McDaniel, P., and Sheth, A. N. Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones. In *Proc. OSDI'10* (2010), 99–106.
11. Felt, A. P., Ha, E., Egelman, S., Haney, A., Chin, E., and Wagner, D. Android permissions: User attention, comprehension, and behavior. In *Proc. ACM SOUP* (2012), 3.
12. Hann, I.-H., Hui, K.-L., Lee, T., and Png, I. Online information privacy: Measuring the cost-benefit trade-off. *Proc. of ICIS* (2002), 1.
13. Hurwitz, J. User choice, privacy sensitivity, and acceptance of personal information collection. In *European Data Protection: Coming of Age*, S. Gutwirth, R. Leenes, P. de Hert, and Y. Poullet, Eds. Springer Netherlands, 2013, 295–312.
14. Joinson, A. N., Reips, U.-D., Buchanan, T., and Schofield, C. B. P. Privacy, trust, and self-disclosure online. *Human–Computer Interaction* (2010), 1–24.
15. Kelley, P. G., Cesca, L., Bresee, J., and Cranor, L. F. Standardizing privacy notices: an online study of the nutrition label approach. In *Proc. of ACM CHI* (2010), 1573–1582.
16. Kelley, P. G., Cranor, L. F., and Sadeh, N. Privacy as part of the app decision-making process. In *Proc. ACM CHI* (2013), 3393–3402.
17. Khalil, A., and Connelly, K. Context-aware telephony: privacy preferences and sharing patterns. In *Proc. ACM CSCW* (2006), 469–478.
18. Knijnenburg, B. P., and Kobsa, A. Helping users with information disclosure decisions: Potential for adaptation. In *Proc. ACM IUI* (2013), 407–416.
19. Knijnenburg, B. P., and Kobsa, A. Making decisions about privacy: information disclosure in context-aware recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS) 3*, 3 (2013), 20.
20. Langer, E. Minding matters. *In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 22). New York, Academic Press.* (1989).
21. Lathia, N., Rachuri, K. K., Mascolo, C., and Rentfrow, P. J. Contextual dissonance: Design bias in sensor-based experience sampling methods. In *Proc. of ACM UBICOMP* (2013), 183–192.
22. Liccardi, I., Pato, J., and Weitzner, D. J. Improving Mobile App selection through Transparency and Better Permission Analysis. *Journal of Privacy and Confidentiality: Vol. 5: Iss. 2, Article 1.* (2014), 1–55.
23. Liccardi, I., Pato, J., Weitzner, D. J., Abelson, H., and De Roure, D. No technical understanding required: Helping users make informed choices about access to their personal data. In *Proc. ACM Mobiquitous* (2014), 140–150.
24. Lin, J., Amini, S., Hong, J. I., Sadeh, N., Lindqvist, J., and Zhang, J. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proc. of ACM UBICOMP* (2012), 501–510.
25. Mancini, C., Thomas, K., Rogers, Y., Price, B. A., Jedrzejczyk, L., Bandara, A. K., Joinson, A. N., and Nuseibeh, B. From spaces to places: emerging contexts in mobile privacy. In *Proc. ACM UBICOMP* (2009), 1–10.
26. Nissenbaum, H. Privacy as Contextual Integrity. *Washington Law Review 79* (2004), 119.
27. Park, Y. J., and Jang, S. M. Understanding privacy knowledge and skill in mobile communication. *Computers in Human Behavior 38* (2014), 296 – 303.
28. Shilton, K., and Martin, K. E. Mobile privacy expectations in context. In *Proc. of Communication, Information and Internet Policy* (2013).
29. Shklovski, I., Mainwaring, S. D., Skúladóttir, H. H., and Borgthorsson, H. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proc. ACM CHI* (2014), 2347–2356.
30. Tan, J., Nguyen, K., Theodorides, M., Negrón-Arroyo, H., Thompson, C., Egelman, S., and Wagner, D. The effect of developer-specified explanations for permission requests on smartphone user behavior. In *Proc. ACM CHI* (2014), 91–100.
31. Toch, E. Crowdsourcing privacy preferences in context-aware applications. *Personal and Ubiquitous Computing* (2014), 129–141.
32. Toch, E., Cranshaw, J., Drielsma, P. H., Tsai, J. Y., Kelley, P. G., Springfield, J., Cranor, L., Hong, J., and Sadeh, N. Empirical models of privacy in location sharing. In *Proc. ACM UBICOMP* (2010), 129–138.
33. Urban, J., Hoofnagle, C., and Li, S. Mobile phones and privacy. In *UC Berkeley Public Law Research Paper* (2012).
34. Venkatanathan, J., Ferreira, D., Benisch, M., Lin, J., Karapanos, E., Kostakos, V., Sadeh, N., and Toch, E. Improving users' consistency when recalling location sharing preferences. In *Proc, of INTERACT (Springer)*. 2011, 380–387.
35. White House. Consumer data privacy in a networked world: A framework for protecting privacy and promoting innovation in the global digital economy. 15–19.
36. White House. Big data and privacy: A technological perspective. 40–41.
37. World Economic Forum. Unlocking the value of personal data: From collection to usage.
38. Xie, J., Knijnenburg, B. P., and Jin, H. Location sharing privacy preference: Analysis and personalized recommendation. In *Proc. IUI '14*, IUI '14 (2014), 189–198.