User Comfort with Android Background Resource Accesses in Different Contexts

Daniel Votipka and Seth M. Rabin, ***University of Maryland;*** **Kristopher Micinski,** ***Haverford College;*** **Thomas Gilray, Michelle L. Mazurek, and Jeffrey S. Foster,** ***University of Maryland***

https://www.usenix.org/conference/soups2018/presentation/votipka

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

August 12–14, 2018 • Baltimore, MD, USA

ISBN 978-1-939133-10-6

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

User Comfort with Android Background Resource Accesses in Different Contexts

Daniel Votipka, Seth M. Rabin, Kristopher Micinski*, Thomas Gilray,
Michelle M. Mazurek, and Jeffrey S. Foster
University of Maryland, *Haverford College
dvotipka,srabin,tgilray,mmazurek,jfoster@cs.umd.edu; *kmicinski@haverford.edu

ABSTRACT

Android apps ask users to allow or deny access to sensitive resources the first time the app needs them. Prior work has shown that users decide whether to grant these requests based on the context. In this work, we investigate user comfort level with resource accesses that happen in a *background* context, meaning they occur when there is no visual indication of a resource use. For example, accessing the device location after a related button click would be considered an interactive access, and accessing location whenever it changes would be considered a background access. We conducted a 2,198-participant fractional-factorial vignette study, showing each participant a resource-access scenario in one of two mock apps, varying what event triggers the access (*when*) and how the collected data is used (*why*). Our results show that both *when* and *why* a resource is accessed are important to users' comfort. In particular, we identify multiple meaningfully different classes of accesses for each these factors, showing that not all background accesses are regarded equally. Based on these results, we make recommendations for how designers of mobile-privacy systems can take these nuanced distinctions into account.

1. INTRODUCTION

Android apps potentially have access to a range of sensitive resources, such as location, contacts, and SMS messages. As a result, Android and similar systems face a critical privacy and usability trade-off: when should the system ask the user to authorize an app to access sensitive resources? Requesting permissions too often can overburden the user; requesting permission too infrequently can lead to security violations.

There has been significant research into this question, much of which shows that users' access-control decisions depend on the context, including when and why the access attempt is made [7, 27, 30, 38, 43, 45]. However, this prior work has typically focused on individual aspects of context in isolation, such as app behavior at the point of resource-access [30, 45, 46], or the reason the app requires access to the sensitive resource [27]. In particular, much of this work relies on a binary distinction between foreground and background accesses—sometimes defined as whether the app is visible on the screen [45,46], and sometimes defined as whether the resource access is explicitly triggered by a specific user interaction [30, 36]. (Section 2 discusses related work in more detail.)

In this paper, we investigate more deeply how users understand resource uses that occur *in the background*, which we broadly define as not explicitly and obviously caused by a user interaction. We examine whether different kinds of background uses are viewed similarly, or whether more fine-grained distinctions are required for user comprehension.

In our investigation, we consider a broad range of possible background accesses, drawn in part from existing literature and in part from reverse-engineering the behavior of popular apps. We examine the context of these background accesses along two key axes: *when* and *why* the resource access is triggered. We consider four cases for *when*: after an *Interaction* (this is a non-background case, as a control), due to *Prefetching*, by a *Change* to a resource such as the device's location changing, or after an unrelated UI action, which we refer to as a *UI Background* access. We also consider five cases for *why*: to *Personalize* the app, to get data from an app *Server*, to support *Analytics* to improve the app, to provide *Ads*, or for no given reason (*NA*). We consider these cases for a mock *Dating* app and a mock *Ride Sharing* app, and for three sensitive resources: *Location*, *Contacts*, and *SMS* messages.

We performed a 2,198-participant, between-subjects online vignette survey investigating users' comfort across 52 conditions selected from the full-factorial set. Each participant viewed a slideshow of a mock app being used and then a diagram illustrating *when* and *why* the app accessed a selected sensitive resource. The participant was then asked whether they would be comfortable using an app that behaved similarly and whether they would recommend such an app to friends. (Section 3 describes our methodology, and Section 4 reports participant demographics.)

We found that both the *why* and *when* aspects of context played a significant role in users' expressed comfort with background accesses. Background accesses that shared data with third parties for advertising and analytics were more objectionable than accesses providing personalized features, even when data was sent off device to the app developer. Perhaps unsurprisingly, of the third-party accesses, those

USENIX Symposium on Usable Privacy and Security (SOUPS) 2018.
August 12–14, 2018, Baltimore, MD, USA.

associated with with advertising were the least acceptable. Additionally, if no reason for access is provided, participants were likely to assume that the data is accessed for personalization. However, perhaps due to uncertainty about whether this assumption is valid, participants were less comfortable in this case than when personalization was explicitly specified.

While all background accesses were viewed as less acceptable than interactive accesses, participants did not react to all background accesses equally. Participants were significantly more comfortable with background accesses when the app is on-screen than when it is off-screen, even when the resource access is not clearly tied to the app's UI. (Section 5 discusses our results.)

Based on these results, we make several design recommendations: that apps explicitly differentiate uses in different contexts, that systems provide better incentives to explain benign uses to users (e.g., when the data is used only for personalization), and that privacy policies track not just *when* data is used, but also where it flows (to explain *why*). (Section 6 presents our design recommendations.)

2. RELATED WORK

In early versions of Android, users were asked to authorize permissions whenever a new app was installed. Multiple studies showed that users did not understand the privacy risks associated with permissions under this model. Felt et al. found that only 17% of their study participants paid attention to the permissions they granted and just 3% fully understood what those permissions could be used to access [15]. Kelley et al. showed that users found the terms and wording of Android permissions hard to understand [22]. Additionally, other researchers demonstrated that users were unable to make informed decisions on whether to install an app because they did not know the context of the resource use, instead relying on their expectations of the app's behavior [5, 25, 38, 43].

Android M [16] and later versions prompt users to grant or deny access to a permission the first time it is required by the app. This model is commonly referred to as Ask-On-First-Use (AOFU). Andriotis et al. found that users feel they have more control of their privacy with AOFU [1, 2]. Bonne et al. showed that users commonly deny a permission and subsequently run the app to determine whether it is truly required [7].

Unfortunately, further work has shown that even under AOFU, users are still not provided with enough context to make informed decisions [7, 30, 38, 43]. In some cases, AOFU may lead the user to make incorrect decisions due to broken assumptions [30]. On the other hand, users experience warning fatigue when presented with too many permission dialogs [6]. Multiple researchers have shown that including a permission's purpose (e.g., feature personalization, advertising) has a significant effect on user comfort [5, 26–28, 38, 41]. Of this prior work, the study by Lin et al. [27] is most similar to ours. In their study, participants were told that a popular app accesses a specific sensitive resource, and participants were given the purpose of that access. Lin et al. collected comfort ratings from 725 MTurkers for 1,200 different combinations of 837 apps, 6 resources, and 4 purposes (participants could provide responses for multiple combinations). They found that the purpose shown had a significant effect on user comfort. We build on their study of user comfort by testing additional purposes, and we compare each purpose to the case where none is given to determine the effect of not informing the user. Also, we add additional variation in our conditions by testing both *why* and *when* the access occurs to study the relative strength of their effects and determine whether there is some interaction between these variables.

Other work has sought to study how user comfort is affected by the timing of resource uses (e.g., after a button is clicked, whenever the resource changes) [30, 43, 45, 46]. Wijesekera et al. study users *in situ*, measuring the effect of the app, whether the app is on screen, and the resource on whether users grant or deny resource access [45, 46]. Wijesekera et al. found that users were more likely to grant access when the request occurred whenever the app was being used. Our prior work studied what resources users expect apps to access as they interact with the app (i.e., on startup, after a button is clicked, when no interaction is shown) [30]. They find that users expect resources to be accessed directly after a related interaction (i.e., camera is accessed after pressing a button labeled "Take a picture"), but do not always expect accesses that are not tied to an interaction. We expand on these findings by investigating comfort with the latter category of accesses.

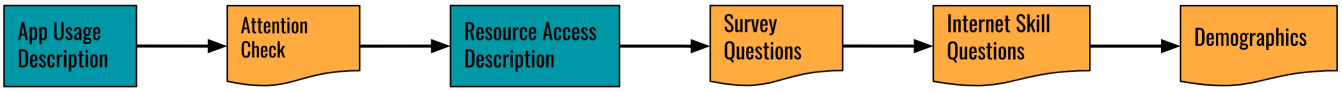
Finally, there has been extensive work on Android permissions more broadly. User comfort has been studied in the context of app recommendation systems [27,27,28,50], which use algorithms to help recommend apps to users based on their privacy preferences. Context has been used to drive static analyses and measure app behavior [9, 12, 19, 47–49]. Lastly, Roesner et al. [36, 37] present Access Control Gadgets, which allows specific buttons in an app to authorize access to specific permissions.

3. METHODOLOGY

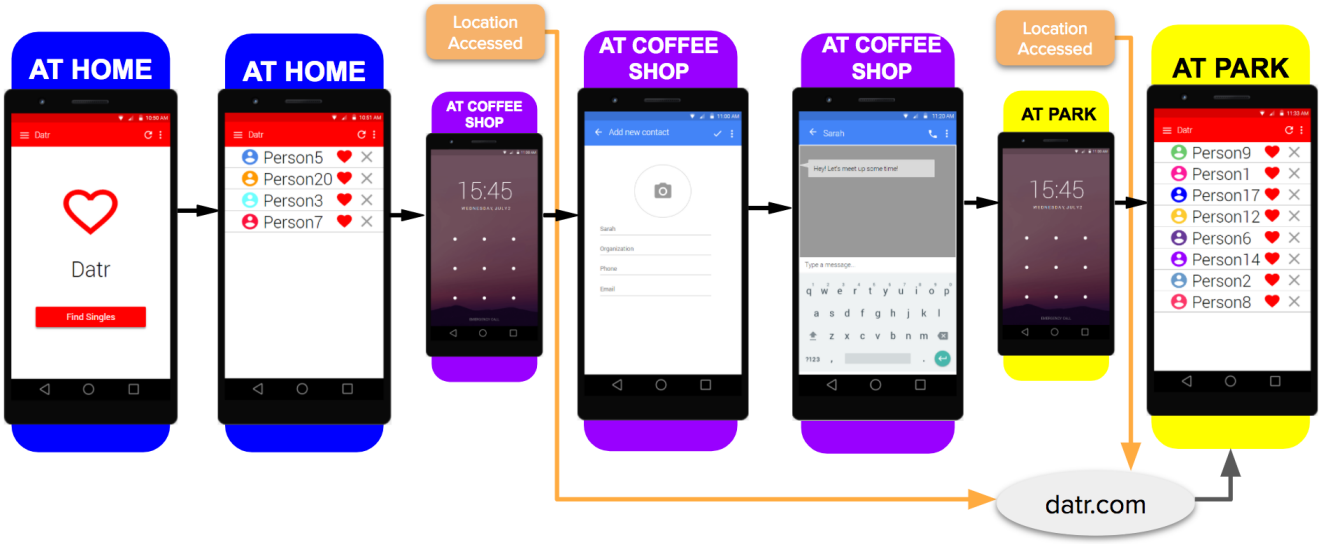
Our study focuses on *background* resource accesses, by which we mean accesses with no obvious, immediate triggering action by the user. For example, accessing device location whenever it changes is a background use because it occurs without the user's direct, immediate intervention. Whereas accessing device location after the user clicks a button is a foreground or *interactive* access. We describe the exact background usage scenarios we study in Section 3.2. From Nissenbaum's theory of Privacy as Contextual Integrity [31] and prior work [5, 26–28, 30, 38, 41, 43, 45, 46], we expect that users' comfort should be significantly affected by the context of an access, including whether it is in the background.

There are potentially many different kinds of background accesses. To determine which accesses to study, we reviewed prior work on common app behaviors [5, 25, 27, 33, 35, 38]. We also manually reverse engineered a small set of Android apps and investigated their background access patterns. In particular, we selected 20 popular apps from our prior analysis that we identified as having background resource accesses [30]. For each app, we used our tool, AppTracer, to locate those background accesses. We then manually examined the app's code, as decompiled with JEB [39], to understand the background access patterns.

Based on this analysis, we decided to study two dimensions of background accesses: *when* the event is triggered and *why*



(a) Survey procedure. Blue rectangles represent description portions of the survey. Orange, curved boxes represent question portions.



(b) Sample vignette for Datr app. In the app usage description, the orange boxes and arrows, and the gray circle, are not shown. They are added in in the resource access description step, along with the follow textual description: "While Jane was using Datr, the app behaved in the following way: Whenever Jane's location changed, Datr learned about the change to her location and sent her updated location to datr.com. datr.com then used her updated location along with other updates it had collected on Jane previously to create a list of recommended singles based on places she has traveled in the past."

Figure 1: User study survey procedure and sample vignette.

the data is accessed. For simplicity, we refer to *when* and *why* as the *access context*. As an example, consider an app that accesses device location every time the device moves and sends this data to a third party advertiser. The *when* is changing device location, and the *why* is advertising.

3.1 Study Overview

We performed a between-subjects, fractional-factorial vignette study [4]. Participants were recruited from the Amazon Mechanical Turk crowd-sourcing service. All participants were at least 18 years old and located in the United States. After completing the vignette study, participants were paid $1.20. Participants took on average 4 minutes and 46 seconds to complete the survey. This study was approved by our organization's ethics review board. Participants were asked for their opinions regarding a given app's functionality and behavior, but we did not explicitly mention privacy or the possible sensitivity of specific resources.

Before beginning the main study, we piloted the survey with nine participants selected from a convenience sample, chosen in part for varying levels of technical knowledge. For each pilot, we asked participants to "think aloud" as they read the prompts and answered each question. We iteratively updated our survey following each pilot, eventually reaching the final instrument detailed below.

Figure 1a describes the survey procedure. First, the participant is shown a short description of the app. We used two mock apps in our study: Datr (*Dating*) and Ridr (*Ride Sharing*). Datr presents users with a list of singles they might like to meet, similar to popular dating apps like Tinder and Bumble. Ridr allows users to request rides to a selected destination, similar to popular ride sharing apps like Uber and Lyft. We do not attempt to fully study the effect of app type on user comfort, but instead simply include two apps from different categories to provide some insight into whether an effect may exist.

We begin each survey by describing how Jane, a fictional character, might use the app. We do so by showing a sequence of screenshots in which Jane first uses the app at home; then travels to a coffee shop and uses other apps; and then travels to the park and uses the app again. Figure 1b shows an example. Note that in this first step, the orange boxes and gray circle in the figure are omitted from the vignette. In this example, Jane opens the Datr app at home, presses the "Find Singles" button, and sees a list of recommended singles. Then Jane travels to a coffee shop to meet a friend. While at the coffee shop, she adds the friend's contact information to her address book and receives a text message. Finally, Jane travels to a park and re-opens Datr, which shows a new list of recommended singles. Vignettes for Ridr are similar, except Jane presses "Request Ride" and is shown a list of recommended destinations.

After viewing the description of the app's use, participants answer a simple question about the app's functionality as an attention check (second item in Figure 1a). We include this check to reduce the risk of invalid responses.

Next, participants are informed of the app's "behind-the-scenes" access context where they are shown the same series of app screens with additional indicators showing *when* and *why* the resource access occurred. Figure 1b shows that Datr collects Jane's location whenever it changes (i.e., *when* she goes to the coffee shop and later to the park) and sends it to `datr.com`. The screenshots are also accompanied by a textual description, listed in the caption of Figure 1b, explaining the scenario. The server then returns a list of recommended singles, to be displayed the next time she opens the app, based on Jane's location. For each vignette, the set and order of app pages shown is the same, but we vary the *when* and the *why* (i.e., the orange boxes, gray circles, arrows, and text explanations).

For all scenarios where information is sent to a server, we show an arrow from the app to a circle labeled with a suggestive domain name. Kang et al. found that this was the most common convention used by non-technical users to draw information transmitted over a network [21].

After describing the app's access context, we ask the participants a series of five-point Likert-scale questions.[1] We begin by asking whether participants believe the access context is likely ("Very unlikely" to "Very likely") to appear in popular apps and whether they agree ("Disagree" to "Agree") the behavior makes the app more useful. We ask these questions because we expect that participants' comfort with an access context is likely affected by their prior exposure to similar scenarios and perceived usefulness of the behavior to the user [32, pg. 133-140].

Next, we directly assess the participant's level of comfort by asking whether they would feel "Very uncomfortable" to "Very comfortable" using an app with the described access context. Additionally, we ask whether they would be "Very unlikely" to "Very likely" to recommend an app with the described access context to a friend who is looking for an app with the given functionality. We add this question to indirectly measure participant comfort.

Some participants were assigned to a condition in which the *why* is not given. In these conditions, we ask participants to provide a short, open response description of why they think the access occurred.

Finally, we conclude with a set of questions about the participants' Internet skill level and demographics. We measure Internet skill using the seven-question scale proposed by Hargittai and Hsieh [17]. Each question on the scale asks participants to rate their familiarity with a different Internet-related term, from "No understanding" to "Full understanding."

3.2 Conditions and Hypotheses

Within this study design, we developed a set of conditions varying over four variables: the *app*, the *resource* being accessed, *why* the app accessed the resource, and *when* the resource was accessed. Table 1 lists the levels for each variable. Conditions consist of one level from each column. As detailed below, we selected a subset of possible combinations to arrive at a final set of 52 conditions, which were assigned round-robin to participants. The condition levels we selected for *when* and *why* map directly to the hypotheses we investigate.

App	Resource[1]	Why[2]	When[3]
Dating	*Loc*	*Personalize*	*Int* [5,6]
Ride Sharing	*Con*	*Server*	*Pre* [5,6]
	SMS [5]	*Analytics* [5,6,7]	*UI-Bg*
		Ads	*Change*
		NA [4,5,6,7]	

[1] *Con* - Contacts, *Loc* - Location, *SMS* - SMS
[2] *Ads* - Advertising, *Analytics* - Debugging/Analytics, *Server* - Server, *Personalize* - Personalize, *NA* - Not Given
[3] *Change* - On Change, *UI-Bg* - UI Background, *Int* - Interactive, *Pre* - Prefetch
[4] Never used with *Int*, and *Pre*
[5] Never used with *Ride Sharing*
[6] Never used with *SMS*
[7] Never used with *Int*

Table 1: Possible values for each variable in tested conditions.

Reasons for resource access. We used five variations for *why* the app collected the sensitive resource. In the personalize (*Personalize*) case, users were told the app collected data to provide personalized features. Additionally, this case stated that no data was sent off device (i.e., to the app's server or any third party). Server (*Server*) is similar, but users were told data was first sent to the app's own server to support personalization. For example, the *Dating* app sends the user's information to the server to retrieve a list of personalized dating matches. Debugging/Analytics (*Analytics*) stated that the app shared data with a third-party for debugging crashes and collecting analytics to improve the app. For Advertising (*Ads*), participants were told the app sent their collected data to a third-party advertiser to improve ad targeting.

From these variations, it can be seen that this context variable also implicitly includes a *who* component. For simplicity, we only consider the general function of the *who* data is shared with and use generic domain names (e.g., *ads.com* for *Ads*). An investigation of the effect of a specific advertisement or analytics provider on user comfort is beyond this scope of this paper.

We also include a Not Given (*NA*) case, in which the participant was not given a reason for data collection.

These scenarios map to our first two hypotheses, as follows.

H1. The provided reason for resource access affects participants' comfort levels.

Within this broad hypothesis, we test three sub-hypotheses concerning specific categories of possible reasons for resource access.

H1a. Users are more comfortable if their information is kept on their device.

[1]The exact wording for each question is in Appendix A.

H1b. Users are more comfortable if their information is not shared with a third party.

H1c. Users are more comfortable if their information is only shared with a third party to improve general app functionality, as opposed to advertising.

Notice that each of these hypotheses represents an increasing degree of willingness to share information.

We test *H1a–c* by searching for divergence among our *why* levels. If *H1a* is true, then we would expect to see a gap in comfort between *Personalize* and *Server*. Similarly, *H1b* indicates a divide between first party (i.e., *Personalize* and *Pre*) and third party (i.e., *Analytics* and *Ads*) sharing. Finally, *H1c* is true if there is a significant difference in comfort between *Ads* and the other levels.

H2. Users are more comfortable if given a reason for background use.

While *H1* investigates how different reasons for resource access compare in terms of user comfort, *H2* asks how these different explanations compare to the lack of an explanation. Prior work in psychology has shown that people are generally more accommodating when given a reason for a request, no matter how vague [24]. *H2* tests whether this is the case for background resource accesses. If *H2* does not hold, then perhaps in some cases the reason for access can be omitted from access notifications or requests, reducing cognitive burden on users without causing undue discomfort.

Triggers for resource access. We considered five variations in *when* the app requests resources. UI-Interactive (*Int*) describes the case where an app accesses a resource after a directly related UI event (e.g., a button click). We include this case, which is not a background access, as a control that mimics the interactive resource use patterns described in our previous work, which led users to expect resource accesses [30].

Prefetch (*Pre*) is similar to *Int*, as the UI indicates that the resource is accessed. However, the actual resource access occurs prior to the UI event (on startup of the application), so that the accessed data is ready to present when the user performs the UI action. In the *Pre* case, there is no visual indication of access when the data is collected, but the user is eventually made aware. UI-Background (*UI-Bg*) presents the same behavior as *Int*—access after a UI event—except the UI event is unrelated to the resource access. Finally, On Change (*Change*) describes an app that accesses a sensitive resource directly after that resource has been modified (e.g., the user changes location or adds/deletes a contact). Note that a *Change* access—unlike *UI-Bg* and *Int*—can occur whether the app is or is not currently in use.

These variations map to our final hypothesis:

H3. Users have different comfort levels when resource accesses are triggered by different events.

Prior work has considered two dichotomous categorizations of access triggers: On-screen vs. off-screen [45] and interactive vs. non-interactive [30]. We use the following sub-hypotheses to understand user comfort across and between these categorizations, with more fine-grained distinctions.

H3a. Users are more comfortable with sensitive resource accesses when they are interactive.

H3b. Users are more comfortable with sensitive resource accesses when there is an explicit foreground visual indicator of use, even if the use occurs before the indicator.

H3c. In the absence of an explicit foreground visual indication of use, users are more comfortable when the app is on-screen than when it is off-screen.

To examine *H3a*, we compare *Int* to all the other levels. To examine *H3b*, we compare *Pre* to the other background levels. Finally, to examine *H3c* we compare *UI-Bg* to *Change*.

Apps and resources. As stated previously, we use two mock apps, Datr (*Dating*) and Ridr (*Ride Sharing*). We selected three resources which we found in our prior work to be used with both foreground and background interaction patterns: Location (*Loc*), Contacts (*Con*), and Text Messages (*SMS*) [30]. We test multiple resources because prior work has shown that grant/deny rates varied between permission types [7].

Final condition set. Because the full-factorial combination of all levels of each variable creates too many conditions to be feasibly tested, we discarded combinations that were redundant, logically inappropriate, or less relevant to our hypotheses. After this reduction, we were left with 52 final conditions.

First, we removed any condition that includes *Int* or *Pre* together with *NA*. Since a reason for the resource access is directly presented to the user through the UI (i.e., the button text clearly states that the resource is accessed to provide personalization of a feature), *Int-NA* and *Pre-NA* are redundant with *Int-Personalize* and *Pre-Personalize*, respectively. Therefore, *NA* is only included with *UI-Bg* and *Change*. This is shown in Table 1 by the orange highlight of *NA* and indicated by the superscript 4.

As we do not intend to completely investigate the effect of app type and resource, we restrict the *Ride Sharing* and *SMS* conditions to only include levels where we expect to observe the largest variation in comfort. Specifically, with *Ride Sharing*, we do not test the resource *SMS*; the *why* levels *Analytics* and *NA*; and the *when* levels *Int* and *Pre*. In Table 1 all the highlighted levels are never considered with *Ride Sharing* as indicated by the superscript 5.

For *SMS*, we do not test the *why* levels *Analytics* and *NA* or the *when* levels *Int* and *Pre*. The levels that are never associated with *SMS* are highlighted in blue, orange, and yellow and indicated by the superscript 6.

Finally, due to the similarity in presentation between *Pre* and *Int*, we limit the levels included with *Int* to only those where we expect to observe the largest variation in comfort. Therefore, we do not consider *Analytics* with *Int*. In Table 1, we highlight in blue—and indicate with the superscript 7—the levels that are never included with *Int* due to this rule.

3.3 Statistical Analysis

For all Likert-scale questions, we use an ordered logistic regression (appropriate for ordinal data) [29] to estimate the

effect of the assigned condition on the participant's comfort, likelihood to recommend the app to others, perceived usefulness of the app behavior, and perceived likelihood that this behavior occurs in popular apps.

For each question, our initial regression model included all the factors and interactions given in Table 2. We applied the standard technique of centering the numerical factor (Internet skill) around its mean before analysis to promote interpretability [10]. To determine the optimal model, we calculated the Bayseian Information Criterion (BIC)—a standard metric for model fit [34]—on all possible combinations of the given factors. To avoid overfitting, we selected the model with the minimum BIC. This process was completed for each regression separately.

Additionally, to understand what participants believed about the reason for data collection when none was explicitly given (i.e., the *NA* level of the *when* factor), we performed an open coding of participants' free responses. Two researchers individually reviewed each response in sets of 30 and iteratively developed the codebook. The coders reached a Krippendorff's α of 0.831 after three rounds of pair coding (i.e., 90 responses), which is within the recommended bounds for coding agreement [18]. The remaining responses were divided evenly and each coded by a single researcher.

3.4 Limitations

Our reliance on mock apps for our controlled experiment limits the ecological validity of our study. We chose this setting because it allows us to reason about the statistical effect of specific factors on participant comfort. Additionally, using mock apps allows us to disregard possible confounding factors such as participants' prior experience with an app or its developer's reputation. However, in this controlled setting, users may be less concerned about their privacy than if their real data were at risk. They may also overstate their discomfort because they are not actually using the app and therefore not placing emphasis on the functionality benefits gained by allowing access to their personal data [40]. To partially account for this, we ask about comfort both directly and indirectly (i.e., would they recommend the app to a friend) and include a description of the app functionality that is dependent on the given access context. Additionally, we only rely on comparative, rather than absolute, results when analyzing responses.

Limiting our study to two types of apps and restricting the resources and access contexts tested is likely to cause us to miss potential factors, especially interactions between factors that affect user comfort. For example, users are likely to expect different types of apps to use resources differently depending on the app's functionality, and these differences in expectations are likely to affect comfort. In an attempt to reduce this problem, we selected conditions based on a review of prior work and manual app reverse engineering.

For each finding from our open-response questions, we report the percentage of participants that expressed a concept. However, a participant not mentioning a specific idea does not necessarily indicate disagreement. Instead, they may have simply failed to state it, or they may not have thought it the most likely possibility. Therefore, our results from open-response questions should be interpreted as measuring what was at the front of participants' thoughts as they responded to the questions.

Since we ask participants to consider app behaviors that occur in the background, it is possible that participants may not completely understand the scenario. However, we attempted to mitigate this issue by using diagrams similar to those drawn by non-technical users to represent network communication [4]. During our pilot interviews, we specifically asked participants to describe what was occurring in the displayed scenario to ensure comprehension, and revised the diagrams accordingly. Finally, all participants who were not shown a reason for the resource access (i.e.,*NA* level of the *why* variable) were asked to state why they thought the app accessed the resource. We did not observe any responses indicating participants misunderstood the scenario.

As is common for any online studies and self-reported data, it is possible that some participants do not approach the survey seriously, and some may try to make multiple attempts at the survey. We limit repeat attempts by collecting participants' MTurk ID and compare these to future attempts to restrict access. Though MTurk has been found to produce high-quality data generally [8, 11, 23, 44], the U.S. MTurker population, from which we drew participants, is slightly younger and more male, tech-savvy, and privacy-sensitive than the general population [20]. This restricted population may affect the generalizability of our results.

However, we consider comparisons between conditions to be valid because each of these limitations apply similarly across all conditions.

4. PARTICIPANT DEMOGRAPHICS

A total of 2,797 participants attempted our survey. Of these, 2,328 (83.2%) finished. From these, we removed several participants who had previously taken the survey. We also removed 121 participants (5.2%) who failed an attention check. We ultimately had 2,198 total responses, with between 40 and 45 responses per condition.

Demographics for our participants are summarized in Table 3. Participants were more male and more white than the U.S. population, as is expected from MTurk. Additionally, our participants' average Internet skill of 32.2 was slightly higher than the mean score of 30.5 recorded by Hargittai and Hsieh on a more general population several years ago [17]. The vast majority of participants use smartphones regularly. The proportion of accepted participants who own a smartphone (99%) is well above the reported U.S. average of 79% reported by Pew [42]. The majority of participants (97%) also considered themselves to have at least "Average" smartphone expertise on a five-point scale from "Far below average" to "Far above average."

5. RESULTS

In our online vignette study, we found that both *why* and *when* resource accesses occurred had a significant effect on user comfort. Additionally, we found that there are several meaningful classes of accesses for each part of the access context.

With respect to *why* the access occurred, we observed that users were more comfortable when data was shared with the app developer (*Personalize* and *Server*) than a third-party (*Analytics* and *Ads*). Further, within third-party sharing, users are more comfortable when data is shared for app an-

Factor	Description	Baseline
When	The context regarding when the sensitive data is accessed	*Int*
Why	The reason the app collected the sensitive data	*NA*
App type	The type of app displayed in the vignette	*Dating*
Resource	The sensitive resource accessed in the vignette	*Loc*
Internet skill	Participant's score on Hargittai and Hsieh's Internet skill scale [17]	0
Smartphone Use	Time per day using a smartphone	0-3 hrs/day
Resource:When	The interaction between the Resource and When variables	*Loc*:*Int*
Resource:Why	The interaction between the Resource and Why variables	*Loc*:*NA*
When:Why	The interaction between the When and Why variables	*Int*:*NA*

Table 2: Factors used in regression models. We compared categorical variables individually to the given baseline. Candidate models were defined using all possible combinations of factors. The final model was selected by minimum BIC.

Metric	Percent
Gender	
Male	54
Female	46
Education	
B.S. or above	49
Some college	39
H.S. or below	13
Age	
18-29 years	34
30-49 years	55
50-64 years	9
65+ years	1

Metric	Percent
Ethnicity	
Caucasian	78
African Am.	10
Asian	1
Hispanic	7
Smartphone Use	
9+	10
6-9	13
3-6	38
0-3	39
No smartphone	<1

Table 3: Participant demographics. Percentages may not add to 100% because we do not include "Other" or "Prefer not to answer" percentages for brevity and selection of multiple options was possible for some questions (i.e., ethnicity).

alytics (to improve the functionality of the application) as opposed to sharing data for advertising. Additionally, if no reason for access was provided, we found that users were less comfortable than they would be if told the data never left their device (*Personalize*), but slightly more comfortable than having their data shared with advertisers (*Ads*).

For *when*, as expected, users are the most comfortable when accesses occur interactively, directly after a UI event (*Int*). Non-interactive (background) accesses can further be divided into two classes: participants were more comfortable if the access occurred when the app was on-screen (*Pre* and *UI-Bg*) compared to off- screen (*Change*). Detailed descriptions of these results are given below.

Interpreting regression results. The majority of our key findings are drawn from our regression analysis over the users' comfort (Table 4a). We also discuss regression analyses for willingness to recommend an app with a given behavior (Table 4b) and belief that the app's behavior is useful (Table 5). Overall, these regressions produced very similar significance results. Our discussion will therefore focus primarily on comfort results.

All three regression tables show (as groups of rows) the variables included in the final selected model. For each categorical variable, we present the base case first. We selected base cases that we expected to produce the highest levels of comfort. For *why*, we selected *Personalize* because it involves the least data sharing. For *when*, we selected *Int* because it is the most interactive, which has been shown to correlate with user expectation of resource access [30]. For resource, we selected *Loc* based on prior work that suggests users are more comfortable with apps accessing location than other sensitive resources [14, 27].

In the odds ratio (OR) column, we show the variable's observed effect. For categorical variables, the OR is the odds of comfort increasing one unit on our Likert scale when changing from the base case to the given parameter level. For the numeric variable (Internet skill), the OR represents the odds of comfort increasing one unit on our Likert scale, per one-point increase in Internet skill. The OR for the base case (categorical) and the average Internet skill (numeric) is definitionally 1.0. For each value, we also give the 95% confidence interval for the odds ratio (CI) and the associated p-value.

As an example, the odds ratio for *Pre* in Table 4a indicates that a user who is assigned to *Pre* rather than *Int*—assuming all other variables are the same—would lead to a 0.64× likelihood of increasing one unit in comfort. Because this effect is less than one, participants are less likely to report higher comfort levels for *Pre* than *Int*. In short, users are less comfortable with *Pre*. Furthermore, *Pre*'s CI indicates that the "true" odds ratio is between 0.48 and 0.87 with 95% confidence. The p-value of 0.004 is less than our significance threshold of 0.05, so we consider this difference between *Int* and *Pre* to be significant.

5.1 H1 and H2: Reasons for resource access

For *H1* and *H2*, we primarily focus on the *why* variable, shown in the first section of Tables 4a and 4b.

Data leaving the device (H1a). We first consider whether resource accesses in which data remains on the device (*Personalize*) are more comfortable for users than those in which data is transferred to the app company's server (*Server*). The first two rows of each Tables 4a and 4b indicate that *Personalize* and *Server* are not significantly different from each other. This is illustrated in Figure 2, which shows participants' Likert responses to the main comfort question, grouped according to the *why* scenario they were shown. In the *Personalize* condition, 44% selected comfortable or very comfortable, compared to 42% in the *Server* condition.

Variable	Value	Odds Ratio	CI	p-value
Why	*Personalize*	–	–	–
	Server	0.88	[0.72, 1.09]	0.240
	Analytics	**0.49**	**[0.37, 0.64]**	**< 0.001***
	Ads	**0.34**	**[0.28, 0.42]**	**< 0.001***
	NA	**0.58**	**[0.43, 0.80]**	**< 0.001***
When	*Int*	–	–	–
	Pre	**0.64**	**[0.48, 0.87]**	**0.004***
	UI-Bg	**0.72**	**[0.55, 0.94]**	**0.014***
	Change	**0.34**	**[0.26, 0.44]**	**< 0.001***
Resource	*Loc*	–	–	–
	Con	**0.33**	**[0.28, 0.39]**	**< 0.001***
	SMS	**0.12**	**[0.09, 0.16]**	**< 0.001***
Internet Skill	0	–	–	–
	+1	**0.95**	**[0.94, 0.97]**	**< 0.001***

*Significant effect – Base case (OR=1, by definition)

(a) Comfort

Variable	Value	Odds Ratio	CI	p-value
Why	*Personalize*	–	–	–
	Server	0.91	[0.74, 1.11]	0.351
	Analytics	**0.51**	**[0.39, 0.67]**	**< 0.001***
	Ads	**0.27**	**[0.22, 0.33]**	**< 0.001***
	NA	**0.58**	**[0.43, 0.80]**	**< 0.001***
When	*Int*	–	–	–
	Pre	**0.62**	**[0.46, 0.84]**	**0.002***
	UI-Bg	**0.68**	**[0.52, 0.88]**	**0.004***
	Change	**0.30**	**[0.23, 0.39]**	**< 0.001***
Resource	*Loc*	–	–	–
	Con	**0.33**	**[0.28, 0.38]**	**< 0.001***
	SMS	**0.13**	**[0.10, 0.18]**	**< 0.001***
Internet Skill	0	–	–	–
	+1	**0.96**	**[0.94, 0.98]**	**< 0.001***

*Significant effect – Base case (OR=1, by definition)

(b) Likelihood to Recommend

Table 4: Summary of regressions over participant comfort and likelihood to recommend apps with different access contexts.

Variable	Value	Odds Ratio	CI	p-value
Why	*Personalize*	–	–	–
	Server	0.93	[0.76, 1.15]	0.502
	Analytics	**0.47**	**[0.36, 0.61]**	**< 0.001***
	Ads	**0.22**	**[0.18, 0.27]**	**< 0.001***
	NA	**0.44**	**[0.33, 0.61]**	**< 0.001***
When	*Int*	–	–	–
	Pre	**0.69**	**[0.51, 0.94]**	**0.018***
	UI-Bg	**0.63**	**[0.48, 0.83]**	**< 0.001***
	Change	**0.33**	**[0.25, 0.43]**	**< 0.001***
Resource	*Loc*	–	–	–
	Con	**0.41**	**[0.35, 0.49]**	**< 0.001***
	SMS	**0.27**	**[0.21, 0.36]**	**< 0.001***
Internet Skill	0	–	–	–
	+1	**0.97**	**[0.95, 0.99]**	**0.002***

*Significant effect – Base case (OR=1, by definition)

Table 5: Summary of regression over participant beliefs regarding the usefulness of different access contexts.

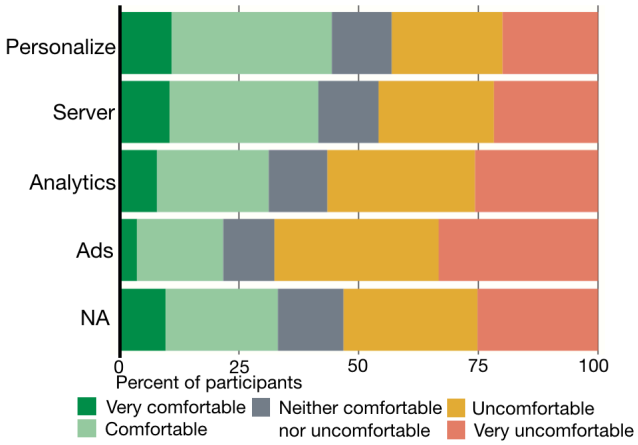


Figure 2: Likert-scale comfort organized by reason for resource access.

Table 5 shows the results of our logistic regression for whether the app's behavior is useful. This provides additional insight into participant preferences, as users may be more willing to tolerate uncomfortable behavior if it is useful. As shown in this table, the *Personalize* and *Server* conditions also do not differ significantly from each other in perceived usefulness.

We therefore conclude that *H1a* does not hold. This corroborates, at a larger scale, the findings of Shklovski et al., who showed that users were comfortable sharing information off device if it was only used by the app's developer [38].

First vs. third parties (H1b). We next consider whether participants responded to first-party accesses (*Personalize* and *Server*) differently than third-party accesses (*Analytics* and *Ads*). Figure 2 shows that participants were overall less comfortable with the third party accesses. Across our two first-party conditions, 43% of participants responded comfortable or very comfortable, compared to only 25% across our two third-party conditions.

As shown in Table 4a, differences between first- and third-party explanations are statistically significant. The *Analytics* and *Ads* conditions are associated with significantly less comfort than the base *Personalize* case. Further, the confidence intervals for *Analytics* and *Ads* do not overlap with that for *Server*, indicating that the two third-party conditions are each significantly different from the first-party *Server* condition as well. The same significance relation holds for app recommendations, as shown in Table 4b. In terms of effect size, the relative odds ratios among the first- and third-party conditions indicate that participants in third-party conditions were between one-third and two-thirds as likely to report a higher level of comfort than were the first-party participants. For example, participants were 0.6× as likely to report a higher level of comfort for *Analytics* than for *Server* (0.49/0.88), and 0.4× as likely for *Ads* than *Server* (0.34/0.88). The effect sizes for willingness to recommend were similar: 0.6× (0.51/0.91) and 0.3× (0.27/0.91), respectively.

A similar analysis of Table 5 shows that participants found

the behavior of apps in the third-party conditions (*Analytics*, *Ads*) to be significantly less likely to be seen as useful than the behavior of apps in the first-party conditions (*Personalize*, *Server*).

Overall, we conclude that *H1b* holds, and that the difference between first- and third-party accesses is meaningful.

Analytics vs. advertising (H1c). We find partial evidence to support *H1c*, which concerns the difference between our two third-party conditions. With respect to our main comfort question (Table 4a), the confidence intervals between *Analytics* and *Ads* overlap, indicating no significant difference between the two. However, comparing confidence intervals in Table 4b does show that participants were significantly more likely to recommend the app in the *Analytics* condition than in the *Ads* condition. *Ads* participants were only 53% (.27/.51) as likely to report a higher level of recommendation. Perhaps unsurprisingly, a parallel reading of Table 5 indicates that participants also found *Analytics* more useful than *Ads*.

Perception when no *why* is provided (H2). To test *H2*, we compare the *NA* condition, in which no reason is provided, to all the other *why* conditions. Overall, we find that *H2* holds partially; a lack of explanation is more comfortable than some explanations, but less comfortable than others.

Inspection of Figure 2 suggests that the *NA* condition falls in the middle of the pack in terms of expressed comfort; 33% of participants in this condition reported being comfortable or very comfortable with this behavior.

Referring again to the top section of Table 4a, we see that this "middle" impression is reflected in our statistical analysis. The *NA* condition is worse than the most comfortable case, with a point estimate of 0.58× the likelihood of higher comfort compared to the baseline *Personalize* condition. On the other hand, comparison of odds ratios suggests that the *NA* condition is slightly (but significantly) better than the worst case (*Ads*). Comparing odds ratios with the other *why* levels, we see that *NA* is not significantly different from *Server* or *Analytics*. The same trend—worse than *Personalize* but better than *Ads*—holds as well for responses to the recommendation question (Table 4b). With respect to the usefulness of the app's behavior, Table 5 indicates that *NA* scenarios were seen as less useful than *Personalize* and *Server*, but not different than *Analytics* or *Ads*.

We asked participants in the *NA* condition to provide an open-ended explanation for the resource accesses they were shown. Figure 3 shows how many participants (grouped according to the type of resource access they were shown) provided each of the most common reasons, according to our manual coding. (Note that an individual participant could provide more than one reason, so totals are greater than 100%.)

By far the most common response (76% of all *NA* participants) was that resource accesses were used for personalization. For example, one participant said Jane's location was accessed to "find singles that are nearby." The second-most common response was advertising (24% of all *NA* participants).

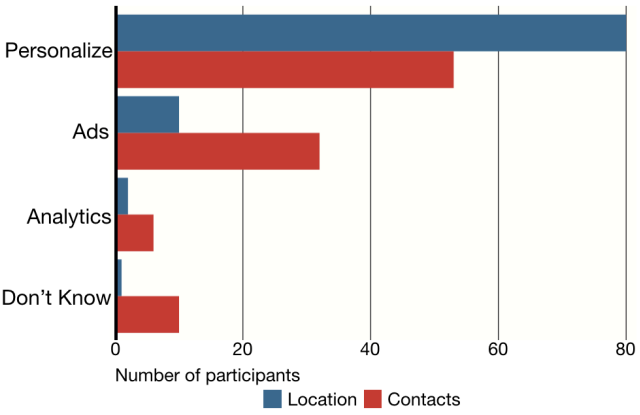


Figure 3: Number of participants who believed the app was collecting their data. Note, these codes are not mutually exclusive, so one participant could express multiple reasons for data access.

Our regression results suggest that a lack of explanation (*NA*) is less comfortable and useful than *Personalize*, even though most participants' assumed the resource access was actually for personalization. Because participants generally did not distinguish between on- and off-device personalization, these personalization responses can be considered roughly equivalent to either our *Personalize* or *Server* conditions. One potential explanation is that the uncertainty associated with a lack of explanation creates some discomfort, even when participants assume that the underlying explanation is acceptable.

Summary of *why* results. Overall, our results for *H1* and *H2* suggest that both who sensitive data is shared with and why matter: accesses used only by the app company for personalization are most comfortable, followed by third-party accesses associated with analytics, with third-party accesses for advertising least comfortable.

5.2 H3: Triggers for Resource Access

We next examine the effect of our *when* variable on users' responses, shown in the second group of results in Tables 4a, 4b, and 5, labeled *when*.

Interactive vs. non-interactive accesses (H3a). We first compare our three non-interactive triggers to the *Int* control condition, to validate that interactive accesses are more comfortable. We find that, as expected, *H3a* does hold. As shown in Tables 4a, 4b, and 5, we find that *Int* is associated with statistically significantly higher levels of comfort, willingness to recommend, and usefulness compared to every other *when* condition. Point estimates range from 1.4× (1/0.72, *UI-Bg*) to 2.9× (1/0.34, *Change*) more likely to report a higher comfort level.

Figure 4, which shows participants' Likert responses to the comfort question organized by *when* condition, illustrates this comfort gap between *Int* and the other *when* conditions.

Importance of visual indicator (H3b). We next consider whether an explicit foreground indication of use can increase user comfort, even if the indication happens after

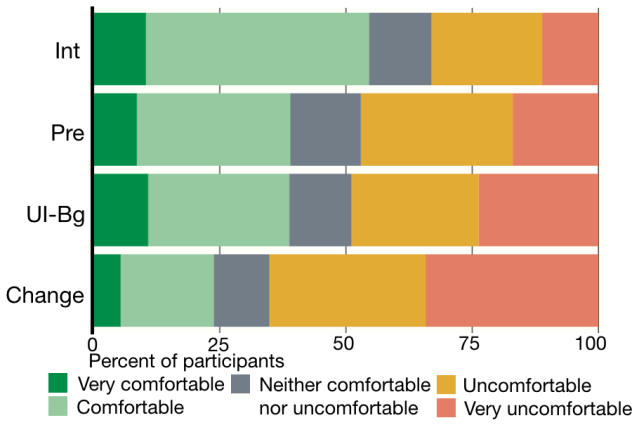


Figure 4: Likert-scale comfort organized by when the resource access occurred.

the access. In particular, we compare the *Pre* condition to the other background *when* conditions.

Comparison of odds ratios in Table 4a suggest that *H3b* holds partially: *Pre* is associated with significantly higher comfort levels than *Change* (1.9×, 0.64/0.34), but is not significantly different from *UI-Bg*. The same pattern holds for willingness to recommend and for usefulness, shown in Tables 4b and 5.

On-screen vs. off-screen (H3c). Finally, we compare the two *when* conditions without visual indicators: *UI-Bg*, which only includes background accesses while the app is on-screen, and *Change*, which includes accesses while the app is off-screen. We find that, as might be expected, off-screen accesses are significantly less comfortable, meaning *H3c* holds.

This finding can be observed visually in Figure 4, which shows that only 24% of participants were comfortable or very comfortable with the *Change* scenario. As shown by comparing odds ratios in Table 4a, this difference is significant: the point estimate suggests that *Change* is only 0.47× as likely to be associated with higher comfort as *UI-Bg* (0.34/0.72). Tables 4b and 5 exhibit the same significance relation for willingness to recommend and usefulness, respectively.

Summary of *when* results. Taken together, our results for *H3a–H3c* suggest three distinct classes of access triggers: interactive accesses, non-interactive (background) accesses that occur when the app is on screen, and background accesses that occur when the app is off screen.

5.3 Other Findings

As described in Section 3.3, our regression analysis included several other covariates beyond *why* and *when*. The final two groups of results shown in Table 4a indicate that the resource shown (*Loc*, *Con*, or *SMS*) and the participants' Internet skill both had significant effects on comfort. In particular, participants reported the highest levels of comfort with the baseline *Loc* resource. Access to *Con* was also significantly (2.8×, 0.33/0.12) more likely to comfortable than access to *SMS*. This aligns with prior work from Felt et al. [14]. Additionally, we observed that users who scored higher on Hargittai and Hsieh's Internet skill scale [17] were significantly likely to be less comfortable (OR 0.95, $p < 0.001$). This means that a participant with the maximum possible score of 35 would be about 0.87× ($0.95^{2.8}$) as likely to express increased comfort as a participant with the mean score of 32.2. This result is analogous to Liccardi et al.'s finding that users with less understanding of how apps operate were more likely to download apps requiring additional significant permissions [25].

Resource type and Internet skill exhibited similar significance relations in willingness to recommend and usefulness (Tables 4b and 5, last two sections), with one exception: accesses to *SMS* were not viewed as significantly less useful than accesses to *Con*.

Notably, none of the interactions we considered (Table 2) appeared in the final minimum-BIC model for any of our outcome variables. This suggests that these variables—most importantly, the *when* and *why* context factors—can be considered independent from each other.

Similarly, app type was not included in any of the final models, meaning we did not observe a significant difference between participants' responses to the Datr and Ridr apps. However, because we only tested two apps, we cannot conclude that the app has no effect on user comfort.

Finally, we observed that a large percentage of participants stated they were uncomfortable or very uncomfortable in all the tested *why* and *when* conditions. In fact, *Int* was the only condition where the majority of participants expressed comfort. In practice, of course, users do use apps with these sorts of background uses. One explanation could be that participants tend to over-report the magnitude of their privacy concerns. Alternatively, users may in practice continue to use apps that violate their privacy preferences because the utility outweighs the cost.

6. DESIGN RECOMMENDATIONS

Based on the results of our study, we make several recommendations for app developers, designers of mobile-privacy systems, and third-party app auditors:

Developers should provide context-sensitive access descriptions. When no reason for an access is given, we found that users are too generous in their assumptions about access context. For example, in absence of explanation, users will tend towards assuming data is being used for personalization (although with slightly lower comfort, perhaps due to uncertainty). If an access is actually used for advertising (or worse, for both advertising and personalization), users might authorize more access than they are actually comfortable with. On the other hand, if data is actually used only for personalization or remains on the device, providing this information could allow the user to feel more comfortable allowing a request than they otherwise would.

Both in Android and iOS, by default whenever an app requests permission to access a sensitive resource (i.e., on first use of the resource), no reason is given for that access. Both systems allow developers to provide a reason, but in practice very few developers take advantage of this feature [41]. Users should be skeptical of any access presented without an explanation, since developers are disincentivized to explain accesses that are used for advertising. Perhaps the Android

API could require an explanation from a fixed set of options, or even default to a "may be used for advertising" explanation if the developer fails to provide a reason. Legitimate developers could presumably be incentivized to provide accurate information to avoid charges of fraud or deceptive practices (analogous to privacy policies).

Further, Tan et al. found that many developers did not include description strings because they did not think they were useful [41]. Our results provide evidence for the utility of these descriptions and could be used to inform design of description strings to ensure users are only shown information relevant to their decisions.

Privacy support agents should consider nuanced variations of context. Because it is unlikely that all app developers will act altruistically, several systems have been proposed to help users make informed decisions according to their privacy preferences, with context in mind [28, 30, 46]. In each, the authors group various access contexts together. We found that such groupings may be insufficiently nuanced. For example, Wijesekera et al. learn user preferences based on whether the app is on- or off-screen at the time of access [46]. This grouping conflates *Int*, *Pre*, and *UI-Bg*, which all occur when the app is on-screen, but were associated with significantly different user comfort levels in our study. Our prior work makes another split, recommending that interactive accesses be treated differently from those that are not associated with user interaction [30]. Again, this oversimplifies user comfort with non-interactive accesses: our results show significant differences between *Pre* and *UI-Bg*, in one class, and *Change* in another.

As a positive example, the privacy assistant developed by Liu et al. divides reasons for resource access into first-party, analytics, and advertisement bins [28]. This grouping accounts for the differences in user comfort we observe in the *why* context. Future such systems should attempt to accurately capture nuanced resource-access classes across both *when* and *why*.

Third-party app auditors should focus on our presented tiers of context. Finally, we believe the job of app auditors can be simplified by concentrating on the most significant contextual classes when investigating app behavior. For example, auditors can focus their efforts on data that is shared off device, because this is most likely to cause user discomfort.

This also highlights the need for tools to support helping auditors answer questions specific to the tiers of context we found. For example, our results underscore the importance of data flow analyses such as Taintdroid [13] and FlowDroid [3].

7. CONCLUSION

In this work, we used a 52-condition, 2,198-participant vignette study to examine how the context of a sensitive resource access in Android—defined as both *when* and *why* the access occurs—affects user comfort with that access. In particular, we examined whether users think similarly about different kinds of background resource accesses, or whether there are important distinctions that determine users' comfort with those accesses.

We found that both *when* and *why* a sensitive resource access occurs have a statistically significant effect on user comfort, and that there are meaningful differences between classes of accesses within both access context variables. While users are most comfortable with interactive accesses, they also make a distinction between non-interactive accesses occurring when an app is on- compared to off-screen. Similarly, users are more comfortable with first-party than third-party accesses, but also make a distinction between third-party accesses for analytics as compared to advertising. We recommend that designers of mobile-privacy systems not only consider both *when* and *why* a resource access is requested, but also respect nuanced distinctions that influence user comfort.

8. ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful feedback. This research was supported in part by a UMIACS contract under the partnership between the University of Maryland and DoD, and by a Google Research Award.

9. REFERENCES

[1] P. Andriotis, S. Li, T. Spyridopoulos, and G. Stringhini. A comparative study of android users' privacy preferences under the runtime permission model. In T. Tryfonas, editor, *Proceedings of the 5th International Conference on Human Aspects of Information Security, Privacy and Trust*, HAS '17, pages 604–622. Springer International Publishing, 2017.

[2] P. Andriotis, M. A. Sasse, and G. Stringhini. Permissions snapshots: Assessing users' adaptation to the android runtime permission model. In *Proceedings of the 8th IEEE International Workshop on Information Forensics and Security*, WIFS '16, pages 1–6, Dec 2016.

[3] S. Arzt, S. Rasthofer, C. Fritz, E. Bodden, A. Bartel, J. Klein, Y. Le Traon, D. Octeau, and P. McDaniel. Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps. In *Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation*, PLDI '14, pages 259–269, New York, NY, USA, 2014. ACM.

[4] C. Atzmüller and P. M. Steiner. Experimental vignette studies in survey research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6(3):128, 2010.

[5] R. Balebako, J. Jung, W. Lu, L. F. Cranor, and C. Nguyen. "little brothers watching you": Raising awareness of data leaks on smartphones. In *Proceedings of the 9th Symposium on Usable Privacy and Security*, SOUPS '13, pages 12:1–12:11, New York, NY, USA, 2013. ACM.

[6] R. Böhme and J. Grossklags. The security cost of cheap user interaction. In *Proceedings of the 2011 New Security Paradigms Workshop*, pages 67–82. ACM, 2011.

[7] B. Bonné, S. T. Peddinti, I. Bilogrevic, and N. Taft. Exploring decision making with android's runtime permission dialogs using in-context surveys. In *Proceedings of the 13th Symposium on Usable Privacy*

and Security, SOUPS '17, pages 195–210, Santa Clara, CA, 2017. USENIX Association.

[8] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5, 2011.

[9] K. Z. Chen, N. M. Johnson, V. D'Silva, S. Dai, K. MacNamara, T. R. Magrino, E. X. Wu, M. Rinard, and D. X. Song. Contextual policy enforcement in android applications with permission event graphs. In *Proceedings of the 20th Annual Network and Distributed System Security Symposium*, NDSS '13, page 234, San Diego, CA, 2013. Internet Society.

[10] J. L. Devore. *Probability and Statistics for Engineering and the Sciences*. Cengage Learning, 2015.

[11] J. S. Downs, M. B. Holbrook, S. Sheng, and L. F. Cranor. Are your participants gaming the system?: Screening mechanical turk workers. In *Proceedings of the 28th ACM Conference on Human Factors in Computing Systems*, CHI '10, pages 2399–2402, New York, NY, USA, 2010. ACM.

[12] K. O. Elish, D. Yao, and B. G. Ryder. User-centric dependence analysis for identifying malicious mobile apps. In *Proceedings of the 1st Workshop on Mobile Security Technologies*, MoST '12, San Jose, CA, 2012. IEEE Press.

[13] W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth. Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones. In *Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation*, OSDI'10, pages 393–407, Berkeley, CA, USA, 2010. USENIX Association.

[14] 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.

[15] A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner. Android permissions: User attention, comprehension, and behavior. In *Proceedings of the 8th Symposium on Usable Privacy and Security*, SOUPS '12, pages 3:1–3:14, New York, NY, USA, 2012. ACM.

[16] Google. *Requesting Permissions at Run Time*, 2016.

[17] E. Hargittai and Y. P. Hsieh. Succinct survey measures of web-use skills. *Social Science Computer Review*, 30(1):95–107, 2012.

[18] A. F. Hayes and K. Krippendorff. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1):77–89, 2007.

[19] J. Huang, X. Zhang, L. Tan, P. Wang, and B. Liang. Asdroid: Detecting stealthy behaviors in android applications by user interface and program behavior contradiction. In *Proceedings of the 36th International Conference on Software Engineering (ICSE)*, ICSE 2014, pages 1036–1046, New York, NY, USA, 2014. ACM.

[20] R. Kang, S. Brown, L. Dabbish, and S. Kiesler. Privacy attitudes of mechanical turk workers and the u.s. public. In *Proceedings of the 23rd USENIX Security Symposium*, USENIX Security '14, pages 37–49, San Diego, California, USA, 2014. USENIX Association.

[21] R. Kang, L. Dabbish, N. Fruchter, and S. Kiesler. "my data just goes everywhere:" user mental models of the internet and implications for privacy and security. In *Proceedings of the 11th Symposium On Usable Privacy and Security*, SOUPS '15, pages 39–52, Ottawa, 2015. USENIX Association.

[22] P. G. Kelley, S. Consolvo, L. F. Cranor, J. Jung, N. Sadeh, and D. Wetherall. A conundrum of permissions: Installing applications on an android smartphone. In *Proceedings of the 16th International Conference on Financial Cryptography and Data Security*, FC '12, pages 68–79, Berlin, Heidelberg, 2012. Springer-Verlag.

[23] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with mechanical turk. In *Proceedings of the 26th ACM Conference on Human Factors in Computing Systems*, CHI '08, pages 453–456, New York, NY, USA, 2008. ACM.

[24] E. J. Langer, A. Blank, and B. Chanowitz. The mindlessness of ostensibly thoughtful action: the role of "placebic" information in interpersonal interaction. *Journal of personality and social psychology*, 36(6):635, 1978.

[25] I. Liccardi, J. Pato, D. J. Weitzner, H. Abelson, and D. De Roure. No technical understanding required: Helping users make informed choices about access to their personal data. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, MOBIQUITOUS '14, pages 140–150, ICST, Brussels, Belgium, Belgium, 2014. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[26] J. Lin, S. Amini, J. I. Hong, N. Sadeh, J. Lindqvist, and J. Zhang. Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 14th ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 501–510, New York, NY, USA, 2012. ACM.

[27] J. Lin, B. Liu, N. Sadeh, and J. I. Hong. Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In *Proceedings of the 10th Symposium On Usable Privacy and Security*, SOUPS '14, pages 199–212, Menlo Park, CA, 2014. USENIX Association.

[28] B. Liu, M. S. Andersen, F. Schaub, H. Almuhimedi, S. A. Zhang, N. Sadeh, Y. Agarwal, and A. Acquisti. Follow my recommendations: A personalized privacy assistant for mobile app permissions. In *Proceedings of the 12th Symposium on Usable Privacy and Security*, SOUPS '16, pages 27–41, Denver, CO, 2016. USENIX Association.

[29] P. McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2):109–142, 1980.

[30] K. Micinski, D. Votipka, R. Stevens, N. Kofinas, M. L. Mazurek, and J. S. Foster. User interactions and permission use on android. In *Proceedings of the 35th*

ACM on Human Factors in Computing Systems, CHI '17, New York, NY, USA, 2017. ACM.

[31] H. Nissenbaum. Privacy as Contextual Integrity. *Washington Law Review*, 79:119–157, 2004.

[32] H. Nissenbaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 2009.

[33] P. Pearce, A. P. Felt, G. Nunez, and D. Wagner. Addroid: Privilege separation for applications and advertisers in android. In *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, ASIACCS '12, pages 71–72, New York, NY, USA, 2012. ACM.

[34] A. E. Raftery. Bayesian model selection in social research. *Sociological methodology*, pages 111–163, 1995.

[35] A. Razaghpanah, R. Nithyanand, N. Vallina-Rodriguez, S. Sundaresan, M. Allman, C. Kreibich, and P. Gill. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem. In *Proceedings of the 24th Network and Distributed System Security Symposium*, NDSS '18, San Diego, California, USA, 2018. Internet Society.

[36] T. Ringer, D. Grossman, and F. Roesner. Audacious: User-driven access control with unmodified operating systems. In *Proceedings of the 23rd ACM Conference on Computer and Communications Security*, Vienna, Austria, oct 2016. ACM.

[37] F. Roesner, T. Kohno, A. Moshchuk, B. Parno, H. J. Wang, and C. Cowan. User-driven access control: Rethinking permission granting in modern operating systems. In *2012 IEEE Symposium on Security and Privacy*, pages 224–238, May 2012.

[38] I. Shklovski, S. D. Mainwaring, H. H. Skúladóttir, and H. Borgthorsson. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2347–2356, New York, NY, USA, 2014. ACM.

[39] P. Software. Jeb decompiler, 2017. (Accessed 5-19-2017).

[40] M. Spence and R. Zeckhauser. Insurance, information, and individual action. In *Uncertainty in Economics*, pages 333–343. Elsevier, 1978.

[41] J. Tan, K. Nguyen, M. Theodorides, H. Negrón-Arroyo, C. Thompson, S. Egelman, and D. Wagner. The effect of developer-specified explanations for permission requests on smartphone user behavior. In *Proceedings of the 32nd ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 91–100, New York, NY, USA, 2014. ACM.

[42] P. R. C. I. . Technology. March 7 - april 4, 2016 – libraries, 2018. (Accessed 1-15-2018).

[43] C. Thompson, M. Johnson, S. Egelman, D. Wagner, and J. King. When it's better to ask forgiveness than get permission: Attribution mechanisms for smartphone resources. In *Proceedings of the 9th Symposium on Usable Privacy and Security*, SOUPS '13, pages 1:1–1:14, New York, NY, USA, 2013. ACM.

[44] M. Toomim, T. Kriplean, C. Pörtner, and J. Landay. Utility of human-computer interactions: Toward a science of preference measurement. In *Proceedings of the 29th ACM Conference on Human Factors in Computing Systems*, CHI '11, pages 2275–2284, New York, NY, USA, 2011. ACM.

[45] P. Wijesekera, A. Baokar, A. Hosseini, S. Egelman, D. Wagner, and K. Beznosov. Android permissions remystified: A field study on contextual integrity. In *Proceedings of the 24th USENIX Security Symposium*, USENIX Security '15, pages 499–514, Washington, D.C., Aug. 2015. USENIX Association.

[46] P. Wijesekera, J. Reardon, I. Reyes, L. Tsai, J.-W. Chen, N. Good, D. Wagner, K. Beznoso, and S. Egelman. Contextualizing privacy decisions for better prediction (and protection). In *Proceedings of the 36th ACM on Human Factors in Computing Systems*, CHI '18, New York, NY, USA, 2018. ACM.

[47] W. Yang, X. Xiao, B. Andow, S. Li, T. Xie, and W. Enck. Appcontext: Differentiating malicious and benign mobile app behaviors using context. In *Proceedings of the 37th IEEE International Conference on Software Engineering*, volume 1 of *ICSE '15*, pages 303–313, Florence, Italy, May 2015. ACM.

[48] Z. Yang, M. Yang, Y. Zhang, G. Gu, P. Ning, and X. S. Wang. Appintent: Analyzing sensitive data transmission in android for privacy leakage detection. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS '13, pages 1043–1054, New York, NY, USA, 2013. ACM.

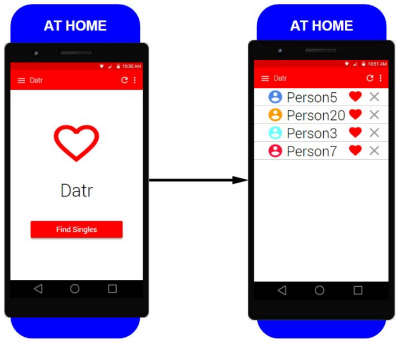
[49] Z. Yang, M. Yang, Y. Zhang, G. Gu, P. Ning, and X. S. Wang. Appintent: Analyzing sensitive data transmission in android for privacy leakage detection. In *Proceedings of the 2013 ACM Conference on Computer and Communications Security*, pages 1043–1054. ACM, 2013.

[50] H. Zhu, H. Xiong, Y. Ge, and E. Chen. Mobile app recommendations with security and privacy awareness. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 951–960, New York, NY, USA, 2014. ACM.

APPENDIX

A. SURVEY QUESTIONNAIRE

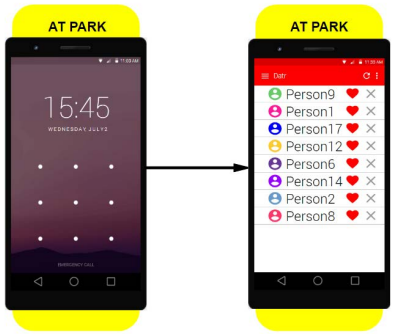
App usage description and attention check.



While at home, Jane decides to use Datr to look for other singles. She opens the app and presses the button "Find

Singles". The app then shows her a screen with a list of recommended singles.

Jane closes the app and travels to a nearby coffee shop where she meets her friend, Sarah. As they get ready to leave, Jane realizes she does not have Sarah's contact information. Jane adds Sarah to her phone's contacts and Sarah sends Jane a text message to remind her that they should meet again some time.



After leaving the coffee shop, Jane heads to the park. She decides to check Datr again and is presented with a new list of singles.

1. Which of the below options best describes the set of steps Jane would have to take to indicate that she is interested in Person 9?
   (a) Press the heart-shaped icon next to Person 9
   (b) Press the X icon next to Person 12
   (c) Press the reload symbol

Resource access description and study questions.

While Jane was using Datr, the app behaved in the following way:

Whenever Jane's location changed, Datr learned about the change to her location and sent her updated location to datr.com. datr.com then used her updated location along with other updates it had collected on Jane previously to create a list of recommended singles based on places she has traveled in the past.

For the remaining questions, we're going to ask you about an app like Datr that collects your location whenever your location changes and sends it to its server to provide personalized features and does not send your location to other parties.

1. Do you think popular dating apps collect your location whenever your location changes and send it to its server to provide personalized features and do not send your location to other parties?
   (a) Very likely
   (b) Likely
   (c) Neither likely nor unlikely
   (d) Unlikely
   (e) Very unlikely
2. Please indicate your level of agreement with the following statement: A dating app like Datr is more useful when it collects your location whenever your location changes and sends it to its server to provide personalized features and does not send your location to other parties?
   (a) Agree
   (b) Somewhat agree
   (c) Neither agree nor disagree
   (d) Somewhat disagree
   (e) Disagree
3. Suppose you were interested in using a dating app like Datr. How would you feel about using a dating app that collects your location whenever your location changes and sends it to its server to provide personalized features and does not send your location to other parties?
   (a) Very comfortable
   (b) comfortable
   (c) Neither comfortable nor uncomfortable
   (d) Uncomfortable
   (e) Very uncomfortable
4. Suppose you know someone who wants to use a dating app like Datr. If you had to recommend an app for them to use, would you recommend an app that collects your location whenever your location changes and sends it to its server to provide personalized features and does not send your location to other parties?
   (a) Very likely
   (b) Likely
   (c) Neither likely nor unlikely
   (d) Unlikely

(e) Very unlikely

Note: For the none case we also included the following free response question

5. Please provide a short description of why you think Datr is interested in knowing Jane's location.

Internet skill questionnaire.

1. How familiar are you with the following computer and Internet-related items? (Items: Reload, Bookmark, Advanced Search, Favorites, Tagging, Preference Settings, PDF) (Choices: No Understanding, Little Understanding, Good Understanding, Full Understanding)

Demographics.

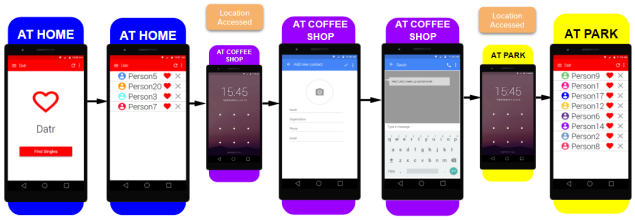
1. What is the highest level of school you have completed or the highest degree you have received? (Choices: Less than high school degree, High school graduate (high school diploma or equivalent including GED), Some college but no degree, Associate degree (2-year), Bachelor's degree (4-year), Master's degree, Doctoral degree, Prefer not to answer)
2. Please specify the gender with which you most closely identify. (Choices: Male, Female, Other, Prefer not to answer)
3. Please specify your ethnicity. (Choices (may choose multiple): Hispanic or Latino, Black or African American, White, American Indian or Alaska Native, (Asian, Native Hawaiian, or Pacific Islander), Other, Prefer not to answer)
4. Please specify your age.
5. Please select the response option that best describes your household income in 2017, before taxes. (Choices: Less than $5,000, $5,000 - $14,999, $15,000 - $29,999, $30,000 - $49,999, $50,000 - $74,999, $75,000 - $99,999, $100,000 - $149,999, $150,000 - $199,999, $200,000 or more, Prefer not to answer)
6. Please select the response option that best describes your current employment status. (Choices: Working for payment, Unemployed, Looking after home/family, Student, Retired, Unable to work due to permanent sickness or disability, Other(specify), Prefer not to answer)
7. How many hours a day do you use your smartphone? (Choices: 0-3, 3-6, 6-9, 9+, Unsure, I do not own a smartphone)
8. Rate your expertise using a smartphone. (Choices: Far above average, Somewhat above average, Average, Somewhat below average, Far below average)

B. EXAMPLE SCENARIOS

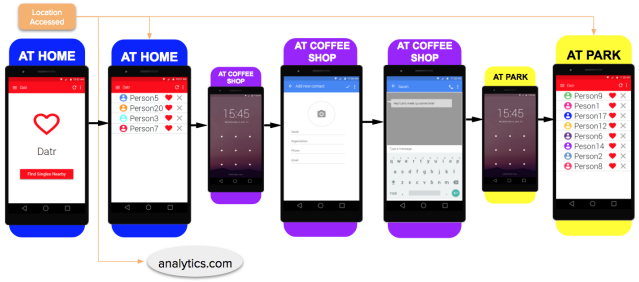
Here, we give a few representative examples of the different resource access scenarios shown to users along with the description of app behavior provided.

(Rideshare, Contacts, UI Background, Advertising)

When Jane presses the button "Request Ride", Ridr learned the contacts in her contact list and sent her contact list to a third party advertiser (advertising.com). advertising.com then used her contact list to better target advertisements to her in the future.



(Dating , Location, On Change, Not Given)
Whenever Jane's location changed, Datr learned about the change and her new location.



(Dating, Location, Interactive, Debugging/Analytics)
When Jane pressed the button "Find Singles Nearby", Datr learned her current location and sent her location to a third party website (analytics.com). analytics.com then used this location information along with other location data it had collected on Jane previously to fix bugs and other problems in the app. Datr also used her location to create a list of recommended singles based on places she has traveled in the past.

(Dating, Contacts, Prefetch, Personalize)
When Jane opened Datr, Datr learned the contacts in her contact list and used her contact list to create a list of recommended singles nearby. Datr creates this list ahead of time so that the list can be displayed quickly if Jane presses the "Find Singles Based On Contacts" button (instead of having to wait a few seconds after the button is pressed). Datr only uses Jane's contact list to personalize her recommendations and does not send her contact list to any other parties (i.e., datr.com or advertisers).