sciendo

Proceedings on Privacy Enhancing Technologies ; 2020 (1):195–215

Daniel Smullen*, Yuanyuan Feng, Shikun (Aerin) Zhang, and Norman Sadeh*

The Best of Both Worlds: Mitigating Trade-offs Between Accuracy and User Burden in Capturing Mobile App Privacy Preferences

Abstract: In today's data-centric economy, data flows are increasingly diverse and complex. This is best exemplified by mobile apps, which are given access to an increasing number of sensitive APIs. Mobile operating systems have attempted to balance the introduction of sensitive APIs with a growing collection of permission settings, which users can grant or deny. The challenge is that the number of settings has become unmanageable. Yet research also shows that existing settings continue to fall short when it comes to accurately capturing people's privacy preferences. An example is the inability to control mobile app permissions based on the purpose for which an app is requesting access to sensitive data. In short, while users are already overwhelmed, accurately capturing their privacy preferences would require the introduction of an even greater number of settings. A promising approach to mitigating this trade-off lies in using machine learning to generate setting recommendations or bundle some settings. This article is the first of its kind to offer a quantitative assessment of how machine learning can help mitigate this trade-off, focusing on mobile app permissions. Results suggest that it is indeed possible to more accurately capture people's privacy preferences while also reducing user burden.

Keywords: Usability, privacy, profiles, Android, smartphone permissions, contextual integrity

DOI 10.2478/popets-2020-0011
Received 2019-05-31; revised 2019-09-15; accepted 2019-09-16.

***Corresponding Author: Daniel Smullen:** Carnegie Mellon University, E-mail: dsmullen@cs.cmu.edu
Yuanyuan Feng: Carnegie Mellon University, E-mail: yuanyua2@cs.cmu.edu
Shikun (Aerin) Zhang: Carnegie Mellon University, E-mail: shikunz@cs.cmu.edu
***Corresponding Author: Norman Sadeh:** Carnegie Mellon University, E-mail: sadeh@cs.cmu.edu

1 Introduction

Managing one's data privacy is an increasingly challenging task for the average person. As research has shown, not everyone feels the same way about the collection and use of their data, hence the need to provide users with privacy options or settings that enable them to configure data flows and ensure that these flows are aligned with their individual privacy preferences. Regulations such as the EU General Data Protection Regulation (GDPR) actually mandate that users be given proper control over the collection and use of their data such as securing informed consent [31]. Effectively, as data continues to be collected and used in ever more diverse ways, users are also expected to make an increasingly unrealistic number of privacy decisions. This situation reflects a fundamental trade-off between the accuracy at which we capture and enforce people's privacy preferences, and the burden we impose on users to specify their preferences by configuring ever more complex and diverse privacy settings [7, 21].

This trade-off between accuracy and user burden when it comes to capturing people's privacy preferences arises in many domains (e.g., browser privacy settings, Facebook privacy settings, or car privacy settings). A rather prominent example is found in the context of mobile app privacy settings (or *app permissions*), which allow users to control the sensitive APIs an app can access. Prompts appear when apps first request access to sensitive data categories (e.g., contacts, location, audio, calendar, camera, etc.). A permission is the ability for an app to access a specific category of data on the smartphone. Many apps ask users to grant them access to multiple permissions. On average, Android users would have to make over a hundred privacy decisions to configure the permission settings associated with their apps [3, 23]. It is no surprise that the vast majority of users do not take the time to configure many of these settings, even though research shows that they truly care about many of them. Indeed, many users express both surprise and discomfort when asked to take

BY-NC-ND

a look at what their permission settings actually allow [3, 20, 23].

Recent research has shown that, using machine learning techniques, it is often possible to predict many of people's privacy preferences based on a relatively small number of factors such as prior privacy decisions or answers to privacy-related questions [22–24, 28]. This approach offers the promise of helping reduce the number of decisions users have to make by possibly giving users individual recommendations on how they might want to configure their permission settings, or by possibly combining multiple closely correlated privacy decisions for individual users. While research on how to best take advantage of these findings is still ongoing, early results involving the deployment of personalized privacy assistants that use these models to recommend privacy settings to users suggest that such an approach can make a big difference [23]. The question that no one has attempted to answer yet is to what extent more expressive mobile app privacy settings might possibly lend themselves to the construction of privacy preference models with greater predictive power and to what extent these stronger predictive models might in turn help mitigate the greater user burden that would otherwise be associated with the configuration of more expressive privacy settings. In this article, we present a first study aimed at addressing this question. Specifically, we focus on answering this question in the context of mobile app permissions, comparing models with permission settings that take the purpose of permissions into account versus models that do not. We present quantitative results aimed at evaluating this trade-off between accuracy and user burden across a number of parameter configurations. Results suggest that machine learning can indeed help mitigate trade-offs between accuracy and user burden. In particular in the context of models that take the purpose of permissions into account, our study suggests that it is possible to get the "best of both worlds", namely doing a better job at accurately capturing people's privacy preferences while simultaneously reducing the number of privacy decisions they have to make. While similar studies would need to be conducted in other domains to determine whether they might benefit from more expressive privacy settings and to what extent machine learning would help mitigate the potential increase in user burden associated with the introduction of such settings, our results are significant in their own right. They show that greater expressiveness in privacy settings does not have to necessarily translate into greater user burden and that machine learning can help mitigate tradeoffs between user burden and accuracy. In addition, our results also strongly argue for the introduction of purpose-specific permissions in mobile operating systems such as Android and iOS:

- As our results show, people's privacy preferences are strongly influenced by the purpose for which permissions are requested. Regulations such as the EU GDPR further mandate obtaining consent from users for the collection of their data for specific purposes.
- Our results further suggest that, using machine learning, interfaces could be built to mitigate the increase in user burden that would otherwise result from the introduction of purpose-specific mobile app permissions.

Main Contributions

Our study sampled 5964 observations of privacy preferences toward three sensitive Android app permissions (calendar, location, contacts), across a corpus of 108 apps, from a large sample of Android users ($n = 994$) in the United States.

We performed a logistic regression analysis, confirming that purpose has a significant effect on participants' expressed preferences for app permissions, and observed patterns across many factors (including demographics). These patterns are indicative of predictive power. We used these factors to improve recommendation models for privacy preferences, leveraging privacy profiles that incorporate a combination of supervised and unsupervised machine learning (agglomerative hierarchical clusters and conditional inference trees). We empirically determined the number of questions required to successfully profile users and count the instances where additional user input is required to make strong predictions. We measured the differences in efficiency and accuracy between the models which consider purpose and those which do not. We find that models which incorporate purpose make more accurate predictions, and can also reduce the overall user burden, even when compared to other similar state of the art approaches [23]. Using machine learning, our approach demonstrates that it is possible to improve the expressiveness of the Android permissions model without trading efficiency for effectiveness.

We address the following research questions:

1. What is the impact of purpose (and other contextual factors) on the predictive power of machine learning models for Android permission preferences?

2. What effect does this predictive power have on the accuracy of recommendations made by privacy profiles?
3. Can we make better predictions without increasing user burden?

2 Related Work

This research builds on existing work in mobile app privacy management and privacy preference modeling and prediction.

2.1 Evolving Mobile App Privacy

Today, smartphones gather myriad sensitive data about users. Many mobile apps access users' sensitive data not only to deliver their core functionality, but also for advertising [2, 9, 20], and unspecified purposes [35, 43] – a common practice in the age of data economy where personal data is increasingly commoditized [47]. However, users often express surprise and concern when they learn about the extent of such collection and use of sensitive data by mobile apps [18, 20, 29, 38, 44]. A recent study showed that both when and why a data practice occurs influence users' comfort levels, with nuances pertaining to a variety of contextual factors, including purposes in particular [45].

Ask On First Use (AOFU) is the current privacy management scheme of Android, the largest mobile platform. AOFU incorporates some contextual factors which have been shown to influence privacy decision making [4, 8, 26, 42]. However, Android does not support purpose-specific permissions, despite prior work suggesting purpose is an important factor in privacy decision making [22, 37]. Even where notice is informally provided, there is no affordance to grant permissions for one purpose and not another. Mobile ecosystems are now pressured to incorporate purpose-specific mechanisms for controlling data access and obtaining user consent under GDPR. Since GDPR, improvements have been made to AOFU, but meaningfully expressing control over purpose is still elusive [5, 27, 34]. Outside of Android, iOS requires app developers to use text descriptions to explain the purpose for app permissions in request dialogues. However, Tan et al. [40] found that many iOS developers did not meaningfully express accurate purposes and that users tended to allow permissions with some explanations. Another study suggested that participants tended to allow permissions without explanations, but were less permissive with vague explanations [37]. It remains a challenge to effectively incorporate purpose into mobile permissions. Our aim is to explore how to best incorporate purpose into the permissions system, using machine learning to achieve more expressive permissions without increasing user burden.

2.2 Improving Permission Management

AOFU requires that users endure the burden to express their privacy preferences by making permissions decisions when permissions are requested by apps. Under the backdrop of increasing app functionality, both iOS and Android apps require users to make more permissions decisions than ever. Usability research shows that increasing the number of controls leads to serious efficiency problems. Many users are habituated to dismissing prompts, and do not consider their actual preferences when permission management comes in the way of achieving their immediate goals. Often, this results in disappointment, frustration, and a feeling of cynicism or loss of control [1, 22, 36, 37].

Early permission managers have difficulty aligning the increasing number of configurable settings with users' individual preferences [3, 32]. One approach is for mobile platforms to share the burden, by providing default settings based on crowd-sourcing or expert opinions [2, 33]. These solutions cannot sufficiently capture individual users' diverse privacy preferences [24]. Instead, machine learning techniques can be used to model and predict segments of users' preferences. These techniques include collaborative filtering [15, 52] and privacy profiles [10, 50]. Privacy profiles are collections of privacy and sharing rules that reflect the preferences of similar-minded people [19, 50]. It is possible to identify privacy profiles that sufficiently capture diverse privacy preferences in both social media settings [10, 51] and mobile app privacy management [22, 24]. A recent field study demonstrates that profiles can provide effective app privacy recommendations for basic permissions without overwhelming users [23]. Building on the prior work, we provide an unambiguous measure of the effectiveness and efficiency, to compare privacy profile-based models that include purpose and those which do not.

2.3 Specifying or Inferring Purposes

Contextual integrity theory argues that privacy should be evaluated in context [30], and the purpose for data collection is a fundamental contextual factor [20, 25, 37]. Recent research has used various contextual factors, such as the extent of data practices and foreground versus background data access to model finer-grained app privacy preferences [4, 25, 48, 49]. Many researchers have attempted to incorporate purposes into privacy preference modeling and prediction [22, 23, 25]. Since purpose-specific control options currently do not exist on mobile app platforms, researchers must either assume or infer the purpose for permissions, or provide separate ways for app developers to express them outside of the permissions model.

Purpose specification and inference approaches rely on static code analysis of apps [9, 14, 46] or context-aware and network traffic analysis [4, 17, 49]. Some tools rely on taxonomies to specify and infer categories of purposes. TaintDroid [9] categorized purposes into legitimate or non-legitimate. Others have adopted more complicated taxonomies with hierarchies of purpose categories [17, 22, 25, 46]. Data Controller Indicators [44] distinguished among core functionality, non-core functionality, and marketing.

Solving the purpose specification and inference problem is outside the scope of our work. Our aim is to explore a hypothetical Android permissions model which enables users to express control over permissions subject to purpose. A recent study showed that a granular taxonomy of purposes affected smartphone users' comfort levels in a nuanced way [45]. However, prior work also suggests that complicated purpose explanations can often confuse users [40]. To avoid potential confusion, we decided to use a simplified taxonomy revised from prior work [20, 23, 44] to include three basic purposes: internal (to provide basic app functionality), advertising, and other/unspecified. We also designed a purpose-independent null condition that is aligned with AOFU to elicit users' preferences to allow or deny permissions, where control with respect to purpose is not expressible and no purpose is mentioned. We provided clear definitions of each purpose to participants, based on the understanding that the way in which expressible purposes (or the lack thereof) may be subjectively interpreted is variable. This variability may also be influenced by other factors, including demographics and smartphone usage behavior, which were measured in post-survey questionnaires.

3 Methodology

We elicited the privacy preferences of Android users using a large-scale IRB approved survey with 994 participants. Participants were recruited and compensated via the Amazon Mechanical Turk platform. A consent form with screening questions was presented prior to participation and data collection. All participants were required to be Android smartphone users located in the United States, and at least 18 years old. Participants were required to affirm that they meet all required criteria when signing the consent form, otherwise, they were ineligible to participate and were removed from the participant pool. Additionally, we designed attention check questions to reconfirm the answers to the screening questions elsewhere in the survey.

First, participants were asked about their preferences independent of purpose, where no purpose was expressible. We generally refer to these preferences and their associated analyses as *purpose-independent*. Next, participants were asked to reconsider their preferences under three expressed purposes: internal, advertisement, and unspecified/other purposes. These are referred to as *purpose-specific*.

Survey responses were first analyzed using logistic regression. We believe that purpose as a contextual factor may impact two usability aspects: efficiency and effectiveness [16]. *Efficiency* is the measure of user burden, such as time and energy required for privacy management. *Effectiveness* is the measure of the accuracy and completeness of a particular privacy management scheme. Next, we applied machine learning techniques to evaluate if privacy profile-based models could improve app privacy permission management in terms of efficiency and effectiveness. We generated agglomerative hierarchical clusters for similar individuals in our dataset, aggregating their preferences into privacy profiles. A *profile* is a model of either (*app category* × *permission*) recommendations or, (*app category* × *permission* × *purpose*) recommendations.[1] Once an unknown individual has been matched to a profile (referred to as *profiling*), the profile can be queried for recommendations across all permissions and app categories. Conditional inference decision trees are used to perform profiling and to evaluate the number

1 We refer to purpose-specific and purpose-independent permissions as permissions generically, for brevity. All models either contain purpose-specific or purpose-independent permissions, but not both.

of questions needed to profile. Profiles and the decision trees used in profiling are static models. Once trained, they do not continue to learn from profiling or queries.

One way that privacy profiles differ from traditional classifiers and recommendation systems is that in some cases profiles cannot make a recommendation for a particular permission. This can be due to sparse data or lack of consensus. Where the clusters of individuals that make up a profile have greater than a specified threshold for consensus about a preference, the profile makes a recommendation. In our study, we tested multiple thresholds between 70% and 90%. Where recommendations cannot not be made (known as *null recommendations*), we default to AOFU: the user is directly asked whether to allow or deny a permission instead. Traditional measures of classifier performance (such as precision and recall) are limited to evaluate our techniques, since they cannot account for null recommendations. We employ two alternative measures of performance. Our measure of *effectiveness* is the accuracy in cases where recommendations are made that coincide with participants' surveyed preferences, divided by the total number of recommendations made. *Efficiency*, in contrast, is the measure of user interactions required to both perform profiling plus the number of AOFU-style preference elicitation prompts (in cases of null recommendations).

3.1 Survey Design

We designed our survey to collect data about participants' app permission preferences through a large number of realistic vignette scenarios. It consists of a main survey and a post-survey demographic questionnaire. The first subsection is a primer on Android permissions in layperson's terms, which we revised from the Android developers' documentation [12]. The primer gave participants of varying technical fluency basic knowledge about Android permissions, necessary to complete the survey. The primer also explained the 3 sensitive permissions we asked about in the survey (i.e. contacts, location, and calendar), and the 3 general categories of purposes we considered: (1) internal, which is required for the app to deliver its basic functionality; (2) advertising, including personalized advertisement, generally collecting, and analyzing data about the user; and (3) unspecified, including any other unspecified or unknown purpose.

The second subsection elicits app permission preferences. Participants answered questions about their privacy preferences towards 6 different Android apps randomly selected from a pool of 108 Android apps. We curated the pool by randomly selecting 54 popular apps ($> 5M$ downloads) and 54 less popular apps (between $50K$ and $5M$) across all app categories in the Play Store. 3 popular apps and 3 less popular apps were shown to each participant in randomized order. The distribution of apps from each category roughly approximated the frequency in the Play Store at the time of surveying. Apps were revealed along with questions about the participants' privacy preferences towards permissions to allow or deny each permission.

First, we showed a screenshot of an app from the Google Play Store, in the identical format seen on a typical Android device. To simulate a realistic app download scenario, we instructed participants to examine the screenshot as they would normally do when making the decision to download and install an app on their phone. Following the app screenshot, we asked questions about their familiarity with the app, their frequency of use, and their preferences to allow or deny the app access to the three permissions. Throughout the survey, participants could hover over information icons to see the definition of each permission as introduced in the primer. These questions serve as the baseline of participants' purpose-independent privacy preferences – no mention of purposes is made, and the specific purpose for the permission is not expressible. Last, we asked participants about their preferences to allow or deny the app access to the same permissions in 3 scenarios, where the 3 purposes described in the primers are expressed. For each app, we collected 12 binary preferences (allow or deny) in total: 3 purpose-independent, and 9 purpose-specific. A generic version of the survey instrument (without app-specific information or Play Store screenshots) can be found in the appendix.

The post-survey questionnaire asked about participants' demographics and smartphone usage behavior, including frequency of use, number of apps installed, number of apps used. These questions helped to determine the likely number of permissions decisions a typical participant would encounter. The number of privacy-related surveys participants had previously completed was also measured. All responses were mutually exclusive categorical factors. In total, the instrument sampled 16 control factors, including app familiarity, app usage, demographics, and smartphone usage. Traditional rankings of privacy awareness such as IUIPC were omitted from our survey, due to lack of statistical significance in prior work [23]. Additionally, we embedded attention check questions throughout the survey. We withdrew participants who failed to correctly answer 2 or more

attention check questions, and their responses were automatically discarded. Participants who completed the survey were each compensated $3.00 for the 15-minute nominal survey duration. The comprehensive list of factors and their statistical significance in regression models can be seen in the appendix.

3.2 Feature Selection with Logistic Regression

To test the statistical significance of all 16 control factors included in the survey, we used a matrix of binomial mixed-effects multiple regression models. Each model was fit by maximum likelihood (Laplace Approximation) [6]. We modeled the random identifier assigned to each participant and the names of the apps they were shown as random effects. The 12 outcome variables and 16 factors were modeled as fixed effects. *A priori* power calculations were performed using G*Power [11] to determine the required number of participants and error rates to achieve a statistical power of 95%. We assumed a small effect size ($f2 = 0.03$) with an error probability of $\alpha = 0.05$ ($Power = 1 - \beta = 0.95$), which required $n = 873$ to achieve noncentrality of $\lambda = 26.19$, a critical F score of $F = 1.76$, and an expected actual power of $Power = 0.950$.

Bonferroni-corrected hypothesis tests were used to determine whether any of the tested predictors were influenced by the control factors. Each regression model in the matrix was tested using χ^2 analysis of variance (ANOVA) against a random-effects-only null model. The design matrix consisted of each permission on one axis, with all fixed effects on the opposing axis. 12 independent hypothesis tests were performed on each fixed effect, one for each of the tested predictors (permissions). Fixed effects which were shown to be statistically significant ($pr\chi^2 \leq 0.05$) were kept as features for further analysis with machine learning. Fixed effects with weak or no significance ($pr\chi^2 > 0.05$) are reported on in our results, but were not included in later models – these may have some limited predictive power.

Once the design matrix was tested, the purpose-specific models were subjected to ANOVA against the purpose-independent models. This tested the hypothesis that, given the same effects, the outcomes (expressed preferences) differ, depending on the purpose. The purpose-independent model was used as the null model. By determining that there was a statistically significant difference between the models in the design matrix, hypothesis testing confirmed whether the affordances related to purpose influenced the participants' expressed preferences. If the null hypothesis was rejected, there are measurable differences in responses when purpose is expressible. Based on the rejection of the null hypotheses, we show that purpose is a significant factor in the regression model. By examining the fixed effects coefficients in each regression model, we can quantify the impact of of each factor on likelihood to allow or deny, based on the levels of each factor. These manifest as changes in regression β-coefficients per differences in age, app category, and so on.

3.3 Building and Evaluating Privacy Profiles

Logistic regression allowed for a systematic and principled approach to feature selection for machine learning models. We built privacy profiles that can make recommendations for individuals based on the features included in the model. Profiles can further tailor predictions about privacy preferences to representative segments of Android users and mitigate the need to ask additional questions to get personalized recommendations. Participants' responses were clustered and aggregated using individual feature vectors comprised of the fixed effects found to be significant in logistic regression, across each permission and app category. The survey dataset was divided into a validation set and training set for machine learning with a 90/10 split, using 10-crossfold validation. A summary of the approach can be seen in Figure 1.

3.3.1 Clustering and Profile Generation

Our approach employs agglomerative hierarchical clustering to cluster similar individuals, which is a parametric unsupervised machine learning method. Each collection of profiles for a given k hyperparameter is referred to as a single model. The parameter k refers to the number of profiles to be used, dividing up the training data into k clusters, and matching the individuals in the test data to these clusters during profiling. A generalization of Gower's distance was used to measure individual dissimilarity, applying a standardization and normalization based on the data type of each element in the feature vectors agglomerated into clusters. Gower's distance is a harmonized dissimilarity metric, suitable for the mixed categorical and binary data types in individuals' survey response feature vectors [13]. Features

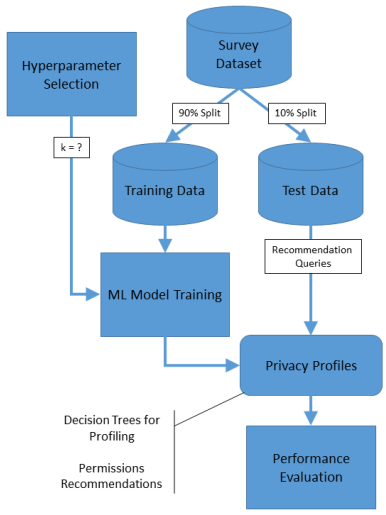


Fig. 1. Our approach to training and evaluating privacy profiles. Recommendation queries originate from unknown individuals in the held-out test portion of the dataset in each crossfold. The process is repeated for each k in the hyperparameter sweep.

related to permissions preferences were transformed into range-normalized quantitative scales from 0 to 1, where 1 was most likely to allow, and 0 was most likely to deny. This transformation was necessary to aggregate the likelihood for an individual to allow a permission for a particular app category, rather than individual apps. Where users were shown multiple apps from the same Google Play Store category, the mean across all apps in the same category was used. Participants were surveyed about apps which were distributed uniformly randomly across all Google Play Store categories, thus, individuals were shown apps according to the same frequency as the distribution of categories in the Play Store. There are over 2 million apps in the Google Play Store, belonging to 25 mutually exclusive categories at the time of data collection. Since making predictions about individual apps would require information about all apps in the Play Store, we instead make generalized recommendations based on their categories. To arrive at a normalized measure of individual preference to allow/deny permissions based on app category, rather than for individual apps, between-subjects responses collected about multiple apps were bucketed based on their app categories and the resulting mean was used.

Transformed and aggregated app category preferences are compared using Manhattan distance. For the remaining features (i.e. demographics, which are all mutually-exclusive nominal categorical responses), each feature is converted to mutually exclusive binary vectors and the Dice coefficient is used. The cluster hierarchy is cut at the specified k, organizing all individuals into k separate clusters with minimal feature dissimilarity. Each participants' feature vector is labeled with their cluster number, 1 through k. Unique individual identifiers (i.e. survey completion codes) are omitted, clustered preferences are aggregated, and the result is k profiles. The agreement thresholds where the aggregated preferences for a cluster are sufficient to make a recommendation were tested at 70%, 80% and 90%. Individuals in the test set are profiled, queried, and the recommendations made by their profile are compared against the individual's expressed preferences from their survey responses to evaluate accuracy.

To evaluate our parametric hierarchical clustering approach, we created a mapping of the hyperparameter space with respect to effectiveness and efficiency. The process of clustering, building profiles, and counting the number of traversed nodes in the inference trees was repeated for each even value of k from 2 through 40, in a hyperparameter sweep. We then repeated this same hyperparameter sweep, simulating and evaluating the performance for up to 36 apps, which is known to be a common number of apps installed and used on Android smartphones [41] and was consistent with our survey results. The sweep was performed at each of the three agreement thresholds – 80% was found to be optimal for both efficiency and effectiveness. The sweep and simulation results are reported on in section 4.3.

3.3.2 Evaluation Procedure and Assumptions

We evaluate the ability of privacy profiles to make accurate recommendations, while limiting the number of user interactions. Since each survey participant was shown 6 apps (to limit fatigue and maximize ecological validity), our initial evaluation was limited to scenarios with a maximum of 6 apps per user. Further evaluation using simulations was performed, based on the Bootstrap distribution of the performance characteristics of profiles built using our 6-app dataset. These simulations were used to predict and evaluate the effectiveness and efficiency of our approach with the expected number of apps found in real-world scenarios, using conservative assumptions.

One part of evaluating efficiency is characterizing the number of interactions required up-front to profile users. The formula for calculating efficiency (E), in terms of the number of user interactions, is given as $E = profile + ask$, where $profile$ represents the number of questions required to profile an individual. The

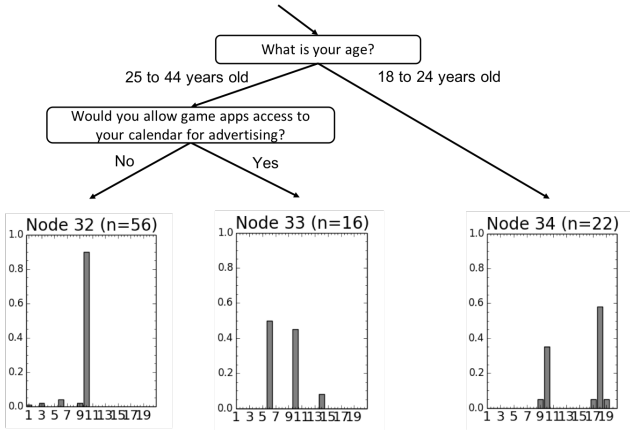


Fig. 2. Detail of one branch in the $k = 20$ decision tree for the purpose-specific model, showing 2 decision nodes and 3 leaf nodes. The remainder of the tree is not shown.

second part, *ask*, counts the number of instances where a user would need to answer an AOFU prompt in the absence of a recommendation.

Profiling uses conditional inference decision trees to re-estimate the regression relationship between clusters and individual preferences. Trees are composed of unidirectional connected decision nodes based on the most statistically significant model features: app categories, demographic factors, and permissions from the design matrix used in the logistic regression analysis. The permutation tests used in the tree generation are based on Strasser and Weber's method, using Bonferroni-corrected p-values [39]. Significance is the same as the original logistic regression models ($\alpha = 0.05$). The length of the tree path traversals from root to leaf nodes are used to characterize the number of questions required to profile an unknown individual from the test dataset. The decision nodes in the trees are questions that must be answered by the participant which determine which profile they should be assigned to. The answers are known *a priori* from their survey responses. The leaf nodes represent a probabilistic conclusion about which profile that the individual ought to be assigned to (see Figure 2). By counting the number of decision nodes required to arrive at a leaf node, we can directly observe the number of user interactions required to profile an individual.

Regardless of the number of recommendations a privacy profile is queried about, profiling need only occur once per individual user. For any given individual's profile, the ability to make a recommendation does not change based on the number of queries it undergoes or the number of recommendations it makes over time. Profiles and the decision trees used in profiling are static. Therefore, with respect to efficiency, no additional assumptions are required for our analysis or evaluation. Privacy profiles can be queried for recommendations *ad infinitum*, and can be asked to make recommendations for an unlimited number of new apps without the need to profile individuals more than once. As such, the number of interactions required for profiling is always constant for any given individual, and efficiency can only decrease proportionally to the number of instances where no recommendation is made. Querying a profile about additional apps for a particular individual introduces opportunities to make more recommendations and possibly null recommendations, decreasing efficiency.

Measuring effectiveness is based on the proportion of correct recommendations, and is not sensitive to the number of user interactions. Effectiveness is essentially a measure of profile accuracy (A), given as $A = (C + null)/Q$ where C is the number of correct recommendations, $null$ is the number of instances where recommendations were not made, and Q is the number of queries for recommendations. Based on this formula, in an instance where no recommendations can be made, the accuracy is assumed to be 100%, as we must assume that the AOFU user interactions that would take place in lieu of a recommendation always elicit user preferences accurately. Accuracy is not penalized when recommendations are not made, so accuracy alone is not the sole measure of effectiveness. Evaluation must also consider the contextual factors that are included in the model, and determine what the acceptable trade-off in efficiency is for a given accuracy requirement, or vice versa.

To simulate the effectiveness of a profile when queried about an arbitrary number of apps, we must make an additional conservative assumption; that the expected accuracy of the profiles' recommendations for an arbitrary number of apps lies within the Bootstrap distribution of accuracy for our 6-app dataset. We use the mean of this distribution for 6 apps when simulating querying profiles for 36 apps, for all values of k in our hyperparameter sweep. This is a reasonable assumption given that the profiles that are being queried in our simulation are the same static profiles that were trained and evaluated with 6 apps, subjected to additional queries.

Because of our limited assumptions, our analysis, simulation, and evaluation are conservative. Our results show that privacy profiles can help mitigate the need for additional interactions by users to elicit their preferences as more apps are installed, in many circumstances.

In contrast, the current AOFU model in Android always requires the maximum number of additional interactions to elicit preferences for new apps, in all circumstances.

4 Results

In this section, we report the results of our survey, findings made during our regression analysis, and the evaluation of our privacy profile-based models.

4.1 Survey Responses

In total, our survey gathered 1092 responses. 98 participants' responses were removed. Of those 98, 38 were removed due to withdrawal or incomplete surveys (3% withdrawal rate). 60 responses were rejected due to failure to correctly answer several attention checks (6% overall rejection rate). Rejected responses were analyzed for evidence of systematic survey abuse; the mean time for responses was approximately 13 minutes, similar to the overall expected duration of the survey based on pilots. However, the median time for rejected responses was only 8 minutes, with a standard deviation of 12.2 minutes. Among all respondents, there was a mean of 0.24 erroneous responses to attention check questions, with median 0 errors, and standard deviation of 0.94 errors per survey. When respondents did fail attention checks, they failed most of them. Based on this data, we observe that overall most participants did not fail any attention checks, and approximately 95% of respondents made no mistakes on attention checks. This seems to suggest that most participants filled out the survey in earnest, and were paying close attention.

4.2 Logistic Regression and Feature Selection

For each of the 3 purpose-independent permissions, and the 9 purpose-specific permissions, logistic regression models identified clear patterns of significance in the fixed effects factors. The final design matrix contained only the factors which were shown to be strong predictors based on strong statistical significance ($pr\chi^2 < 0.05$). These included Familiarity with App, App Category, App Usage Frequency, Age, Education Level, Participant City Size, Marital Status, and Number of Apps Used.

Factors with marginal or weak significance were discarded; these included Gender, Employment Status, Smartphone Usage Frequency, Smartphone Usage Duration, Android Version, and Participant's Number of Recently Completed Privacy Surveys. Gender showed weak significance across all permissions, regardless of purpose. Surprisingly, Marital Status was a very strong predictor across all permissions preferences. In particular, participants who were divorced, widowed, or never married were most similar and were more likely to deny permissions broadly. Participants who were married or separated were more likely to allow, in comparison to those who were divorced. While Employment Status was generally a very weak predictor, it was observed to have strong significance for Calendar permissions. Participants who are not working because they were retired or disabled were more likely to deny, while those who are students, paid employees, laid off, or otherwise looking for work were more likely to allow.

The variance in Participant Smartphone Usage Frequency and Participant Smartphone Usage Duration was the likely explanation for the observation that these two factors had very weak significance, with only marginal significance (if any) in many cases. This suggests that this aspect of smartphone usage behavior is not a useful predictor. However, Number of Apps Installed and Number of Apps Used appeared to be a very strong predictor in almost all cases; there is a clear trend where participants are more likely to allow access to permissions if they have many apps installed, and if they report that they use many of them. Participants with small numbers of apps installed and used were more likely to deny permissions in many cases, perhaps because participants who are more privacy conscious download fewer apps.

Android Version and Participant's Number of Recently Completed Privacy Surveys had too little significance to observe any clear response trend. Participants' number of recently completed privacy-related surveys did not appear to correlate with any particular characteristic of responses, nor did Android version. Many participants self-reported outdated Android versions, including some which do not support AOFU, which suggests that they may not have the technical knowledge to determine what version they had.

The significance levels of all factors, and all permissions, can be found in the appendix. The results of the final ANOVA are summarized in Table 1. The null hypothesis is that there are no differences in responses between the purpose-specific and purpose-independent models. The purpose-independent model

	Contacts			
	Df	χ^2	χDf	$pr(>\chi^2)$
Null vs. Internal	**57**	**≈ 0**	**0**	**≈ 1**
Null vs. Advertisement*	**57**	**1039.1**	**0**	$\leq 2.2 \times 10^{-16}$
Null vs. Other/Unspec.*	**57**	**1577.6**	**0**	$\leq 2.2 \times 10^{-16}$
	Calendar			
Null vs. Internal	**57**	**≈ 0**	**0**	**≈ 1**
Null vs. Advertisement*	**57**	**1292.1**	**0**	$\leq 2.2 \times 10^{-16}$
Null vs. Other/Unspec.*	**57**	**2025**	**0**	$\leq 2.2 \times 10^{-16}$
	Location			
Null vs. Internal	**57**	**≈ 0**	**0**	**≈ 1**
Null vs. Advertisement*	**57**	**1180.7**	**0**	$\leq 2.2 \times 10^{-16}$
Null vs. Other/Unspec.*	**57**	**1952.1**	**0**	$\leq 2.2 \times 10^{-16}$

Table 1. ANOVA of purpose-independent regression models (Null) versus purpose-specific models. Those marked with an asterisk reflect the rejected null hypothesis.

is the null model, which is subjected to ANOVA versus the purpose-specific models across the three permissions. The alternative hypothesis is that the purpose-specific information has measurably different responses. It is clear based on the rejection of the null hypotheses that there are measurable differences when comparing purpose-independent to purpose-specific regression models, except in the case of Internal. One possible explanation is that participants already assumed that the app permissions which were purpose independent are already declared because they are ostensibly for Internal purposes. Regardless, rejecting the null hypothesis provides strong evidence that the participants' purpose-independent expressed preferences do not intersect with their purpose-specific expressed preferences in a significant number of instances. There is a significant difference in expressed preferences between the two types of affordances – this implies that users would clearly benefit from the ability to express purpose-specific preferences. We elaborate on this in section 4.3.

4.3 Analyzing and Evaluating Privacy Profiles

Here we present the measures of efficiency and effectiveness in our analysis using machine learning. We use hierarchical clustering to build privacy profiles and conditional inference decision trees to do profiling. We then evaluate the recommendation accuracy and number of additional interactions required by each model.

Our hyperparameter sweep of k values showed two dominant tendencies of values for k, for both the purpose-independent and purpose-specific models. There appeared to be a relationship between efficiency and effectiveness as the agreement threshold changes; as the threshold is raised, the profiles are able to make slightly more accurate recommendations, at the cost of an increased number of user interactions. In our results, we report on the hyperparameter sweep with an agreement threshold of 80%, as it was found to be the best trade-off between efficiency and effectiveness. We found that with an agreement threshold of 70%, the mean accuracy of our recommendations decreased by approximately 5%, and the average number of additional interactions decreased by 3 nearly uniformly for all values of k. At a threshold of 90%, the mean accuracy increased by 5%, and the average number of additional interactions increased by 3 for all values of k.

As can be seen in Figure 3, the accuracy overall appears to be within the range of approximately 75% to 90% across all values of k in both models. The accuracy of the purpose-specific model is a few percent higher on average, particularly at $k > 14$, even though it incorporates additional context.

While the difference in effectiveness between the two models is not particularly large, there are larger differences in efficiency. This can be seen at a glance in Figure 4a and 5a. Questions about the overall "best" models and k are best answered using the scatterplots in Figure 4b and 5b. In these graphs, the x-axis is the overall efficiency measure, showing the number of user interactions in the expected case to perform profiling and recommendations for 6 apps (the sum of the two lines in Figure 4a and 5a respectively). The y-axis represents the overall recommendation accuracy (seen in Figure 3). The individual points are labeled with the value of k, colored by model type. What is evident is the relationship between accuracy and the number of user interactions. Where the highlighted regions show the central tendency of the two models, one can observe that the purpose-specific model consistently shows fewer user interactions for proportionally higher accuracy overall.

There are outliers from the highlighted areas in Figure 4b. In particular, it's worth noting $k = 2$ and $k = 4$ are outliers in both models, appearing to suggest that the best accuracy/efficiency trade-off might occur with very small numbers of profiles. However, using such a small number of profiles is impractical for the same reasons identified in prior work, which found that a single set of defaults or a very small number of profiles are too internally heterogeneous to generalize well [22, 23]. With 36 apps, small values of k prove far worse than they appeared in Figure 4b with only 6 apps, because they cannot make recommendations in a much higher percentage of instances. With low values of k, it takes

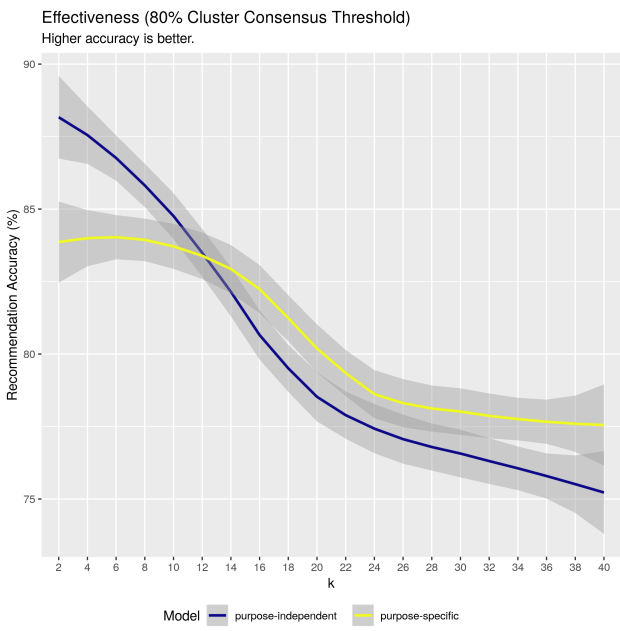


Fig. 3. There is a minimal difference in accuracy between the purpose-specific and purpose-independent models. The effectiveness is higher for most values of k in the purpose-specific model.

a very small number of questions to profile individuals, the highest numbers of additional interactions are required as these profiles seldom make any recommendations. As such, they sub-optimally trade higher accuracy for many more additional user interactions.

Profiles with small values of k are far more timid about making recommendations due to lack of consensus, but make accurate recommendations in limited cases when they can. In contrast, in terms of user interactions, they will always be worse than more personalized models with more clusters. Recall that once trained, profiles do not change, so the number of questions required to profile an individual is always the same regardless of the number of queries they are subjected to. As one would expect, the overall number of average user interactions increases in the 36-app simulation (Figure 5b, but the points under the red and blue highlighting show a much more consistent manifold and the outliers now fall within the central tendency.

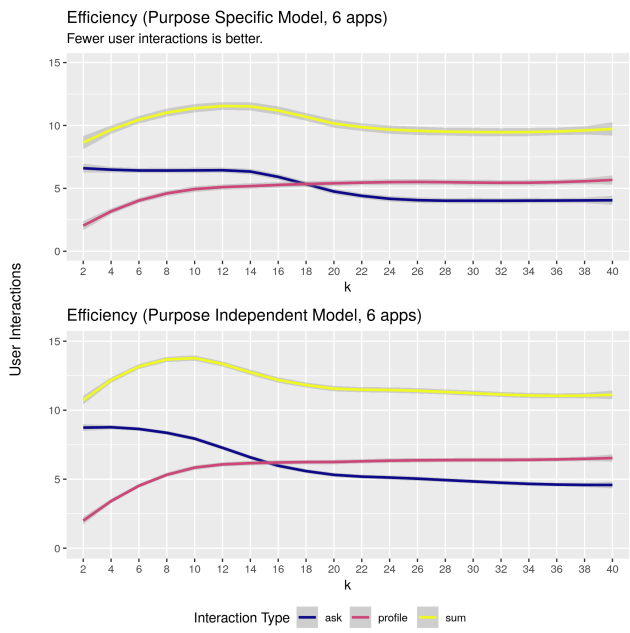
We observe that while it is easy to profile individuals with a very small number of profiles, the true cost in additional user interactions comes from the profiles' inability to make recommendations afterward. Recall that efficiency is the measure of user interactions required to both profile users, and additionally to ask their preferences when a recommendation cannot be made for a particular app. As the number of profiles increases, the number of questions required to profile individuals increases, but this increase flattens out substantially after $k > 10$. This can be seen in Figure 4a and 5a as well. Note that there is a similar inflection point in the decreasing trend for additional interactions where recommendations could not be made.

4.3.1 Choosing k for a Given User Interaction Budget

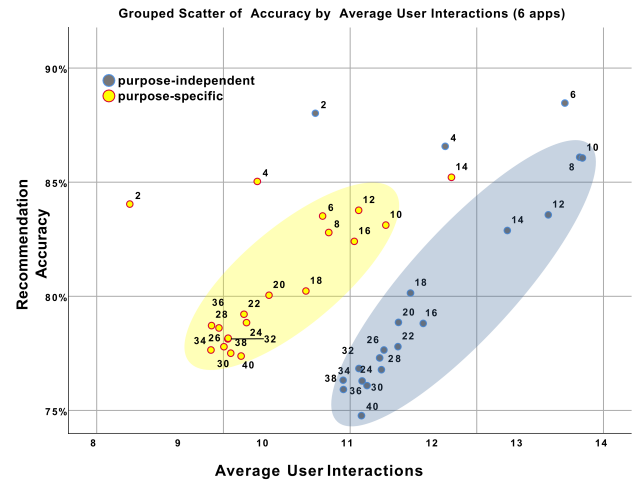
Instantiating a profile-based privacy assistant requires that one model (either purpose-specific or purpose-independent) be chosen, using a single value of k. Ideally, one would choose a value which is best suited to a desired efficiency and effectiveness trade-off – either by choosing an upper limit for the number of user interactions or achieving a particular accuracy percentage.

Choosing one of the values seen in the purpose-specific model (yellow-highlighted point cloud) in Figure 4b or 5b would be most ideal overall, as they lie within a Pareto-optimal grouping with higher accuracy and fewer user interactions. A helpful way to frame this is to describe the graph in terms of the maximum accuracy that can be achieved for a given limit on interactions for 36 apps. Note that there is no single optimal k value overall; the choice must be made based on either a maximum budget of user interactions, or a target accuracy figure. Thus, for a budget of around 30 user interactions, we can see that the purpose-specific model is optimal at $k = 28$ or $k = 24$, achieving an accuracy of around 78%. In contrast, the purpose-independent model has no value of k which can work within this budget. For a budget of around 40 user interactions, the purpose-specific model is optimal at $k = 16$, achieving an accuracy of around 83%. In contrast, the purpose-independent model is optimal at $k = 18$, achieving an accuracy of only 80%. For a budget of around 50 user interactions, the purpose-specific model is optimal at $k = 14$, achieving an accuracy of 85%, well under budget (with only around 45 interactions). In contrast, the purpose-independent model is optimal at $k = 12$, achieving an accuracy of around 83%. It is worth noting that the purpose-independent model is able to achieve the highest accuracy at $k = 6$ – in this case, the accuracy is misleading, because the model makes recommendations in the fewest circumstances, requiring the highest budget of 60 user interactions.

Our assumptions (detailed in section 3.3.2) allow the possibility to achieve maximal 100% accuracy by abandoning privacy profiles altogether and resorting to AOFU. In the case of 36 apps, 3 permissions, and 3 purposes, $36 \times 3 \times 3 = 324$ user interactions would be

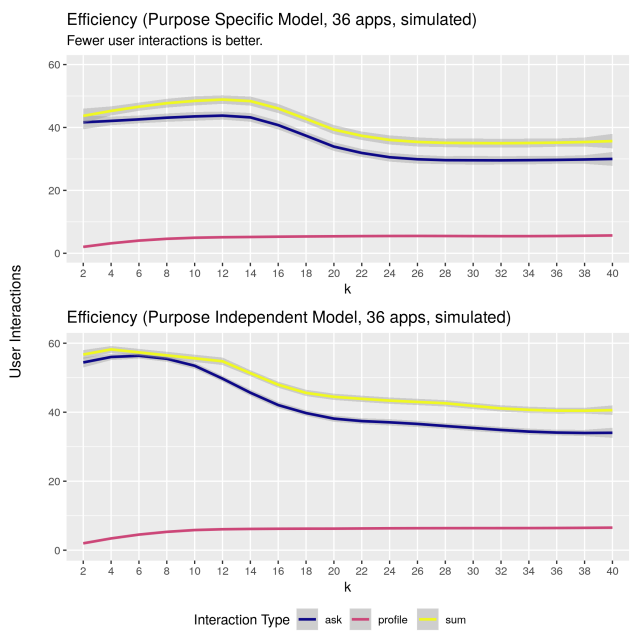


(a) This graph breaks apart the efficiency measurement into the number of interactions required to profile a user (profile), and the number of additional interactions required when recommendations cannot be made (ask). The sum of the two is also shown.

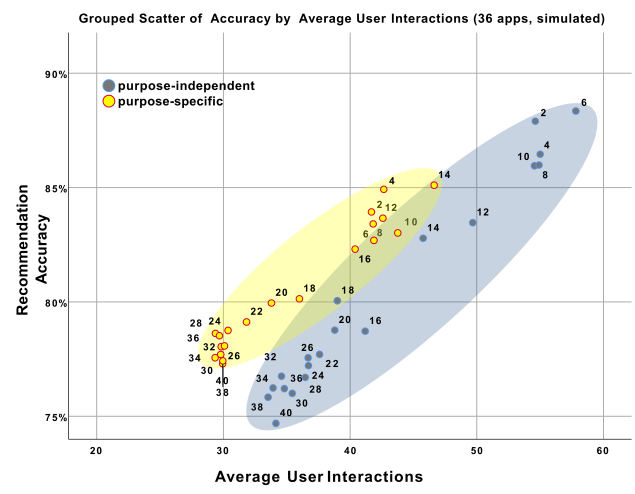


(b) This plot shows the overall relationship between efficiency and effectiveness at different values of k, using the data for 6 apps. Higher accuracy and fewer user interactions are more desirable.

Fig. 4. Hyperparameter sweep for 6 apps.



(a) The number of interactions required to profile a user is static, as can be seen when comparing with the *profile* line in Figure 4a. The *ask* line increases proportionally to the number of apps; both models are queried for more recommendations, but the purpose-independent model makes fewer recommendations, and must ask additional questions more often.



(b) The overall relationship between efficiency and effectiveness. Note that the number of user interactions varies, as can be seen between the two figures 4a and 5a.

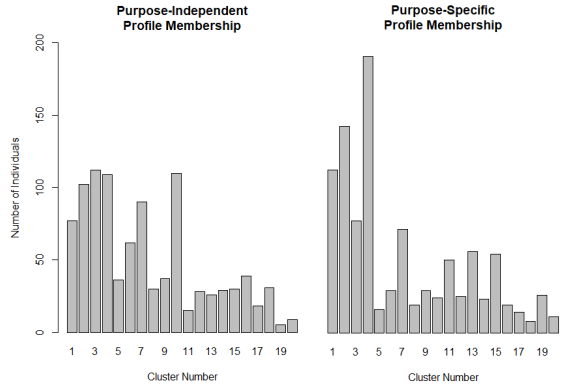
Fig. 5. Hyperparameter sweep for 36 apps, simulated using the Bootstrap distribution from 6 apps.

required. In contrast, profiles would require new interactions in only 17% of instances in the worst case (28 interactions at $k = 6$ in the purpose-independent model), and only 8% of instances in the best case ($k = 34$ in the purpose-specific model).

4.3.2 Example at $k = 20$

To highlight differences in the characteristics of the purpose-specific and purpose-independent models, we show an illustrative example at one value of k. We can see in Figure 6a that at $k = 20$, the purpose-independent model shows several dominant clusters. This suggests that a large proportion of users fit into a small number of dominant categories, however, the remaining clusters still account for the majority of users overall. In the purpose-specific model (Figure 6b) we observe that there is a greater tendency towards a single dominant cluster, but there is greater variability in the proportions of the other clusters. A similar trend in cluster membership was observed across other values of k, particularly those where $k > 8$. This supports the idea that while many individuals generally tend towards similar preferences, there is broad variability that can be better expressed along the extra dimension of purpose. Variability makes preferences more heterogeneous when individuals are clustered among a small number of profiles. This observation is further supported by measures of intra-cluster similarity. Larger numbers of profiles are more internally homogeneous in the purpose-independent model. Comparing silhouette coefficients, we see the average silhouette coefficient for the purpose-independent model in the $k = 20$ example is 0.03, and for the purpose-specific model is -0.07. Both coefficients suggest overlapping clusters, which is unremarkable considering the nuances of individual preferences, but the purpose-specific model has slightly less internal homogeneity. However, balanced against the efficiency of higher values of k, it is clear that small numbers of profiles, while able to achieve accurate recommendations, fail to make recommendations in a large number of instances. Models with low values of k group everyone into a small number of clusters that are largely heterogeneous.

The $k = 20$ example is again illustrative of the kind of predictive power that purpose-specific profiles are capable of. Given 3 permissions, and 25 app categories, there are a total of 1350 recommendations made across $k = 20$ profiles (75 per profile), with 326 null recommendations (where profiles achieved less than 80% consensus on allowing or denying), and 1024 recommendations. Once an individual is profiled, recommendations can be made in approximately 76% of circumstances overall, with only very small differences among more or less popular app categories. There are on average 18 null recommendations per profile, and 57 recommendations per profile across all app categories. The purpose-independent model can profile an individual within an approximate range of 3 to 7 questions, across all values of k. The purpose-specific model can profile an individual within a range of 2 to 6 questions on average.



(a) Purpose-independent model. **(b)** Purpose-specific model.

Fig. 6. Cluster membership histogram for $k = 20$ profiles. Note the the overall flatter characteristic for 6a, and the tendency towards more dominant clusters for 6b.

5 Discussion and Limitations

In this section, we discuss the implications of our results as well as potential limitations of our work. We remark upon the practical implications of our work, towards mobile app privacy management, including modeling, prediction, and recommendations. We find that there are implications for our work towards permissions-based notice and control methods broadly.

5.1 Potential Limitations of Our Work

Due to the limits of the data collected in our survey, the specific profiles built using this data are not sufficient to make recommendations for all apps and all app categories. At the same time, our results indicate that our approach is capable of making accurate recommendations in most circumstances. This simply reflects the fact that our limited resources enabled us to only collect data about a limited subset of apps in the Google Play

Store. In addition, some app categories are more popular than others. For example, Games is the predominant category of apps in general. Other more esoteric categories (such as Health apps) have more limited numbers of apps. The Google Play Store is in a constant state of flux, with new apps being added, apps being removed, and their categorization being modified over time. There are also limits to the number of apps which can practically be surveyed with a given participant without introducing survey fatigue, which would diminish the validity of elicited preferences. We limited our survey to 6 apps per participant, which allowed us to demonstrate our approach empirically without an overly burdensome survey, and generalized those results using simulations. Our findings suggest that our approach would be increasingly effective for users with more than 6 apps, due to the static number of user interactions required to profile a given individual. In other words, the number of user interactions required to assign a user to a profile does not change with the number of apps the user has on his or her phone. Once the user has been assigned to a profile, the number of recommendations that can be generated for that user increases with the number of apps the user has on his or her phone. If you were to think of the number of interactions required to assign a profile to a user, you could say that there is a greater return on investment for a user with more apps. With only 6 apps, this return on investment is still small compared to what it would be for a user with 30 or 50 apps. We defer further exploration of our approach with large numbers of apps to future work.

During the course of our survey, we observed changes to the categories of some apps on the Play Store, which may suggest a limitation to our approach. App developers are ultimately responsible for app categorization, which can potentially result in variability in interpretations of app categories. For example, the Waze app is categorized as "Maps & Navigation", while Google Maps (which has ostensibly the same functionality) is categorized as "Travel & Local". The categorization can also change, based on the developer changing it or the Play Store categories changing. The result is that the ability for an app category to generalize well to all apps that are categorized as such may be limited. However, our observation was that the categories that were selected for apps seemed to be meaningful representations of the apps in terms of their overall theme and functionality. In general, this limitation does not appear to prevent our approach from making good recommendations – there are a small overall number of categories for apps to belong to, and they are mutually exclusive. One can imagine that if there was a stricter categorization scheme imposed by the Play Store, this limitation would be mitigated. Yet another approach would be to automatically generate our own categories (e.g., mining the text found in descriptions of apps in the Google Play store and possibly also taking into account other factors such as the particular permissions requested by apps).

We also noted that there were differences in the descriptions of the permissions as they are presented in the Play Store versus Android developer documentation. We revised permission descriptions from developer documentation, as it provided the most internally consistent description. In our study, we control for the variability of interpretation by not priming participants about the category for the apps which were surveyed. We also analyzed correlations across the app category factor in post-hoc tests. Survey participants may have taken app categories into consideration, but our method can only infer whether this factor influenced their responses to allow or deny permissions, not what their interpretation was.

It is not possible to enumerate and measure the significance of every possible contextual factor that may have had an influence on participants' responses. There might be confounding factors which were not controlled for, such as reports of privacy violations in the media at different points during our study. We are not aware of any such factor having had a meaningful impact on the data we collected.

Finally, we acknowledge that our study is based on self-reported privacy preferences and that such preferences are not necessarily representative of user behavior, namely the way users would actually configure the permission settings considered in the study. In fact, prior research has shown that both self-reported privacy preferences and actual configuration of privacy settings are malleable. This includes research on nudging users to more carefully consider the configuration of mobile app permissions [3]. The vignettes presented to participants in our study share similarities with those used in similar studies of people's privacy preferences (e.g. [29]). The selection of a methodology that focuses on collecting self-reported privacy preferences simply reflects the difficulty of collecting actual user behavior, given that current mobile app permission settings do not support purpose-specific permissions.

5.2 Models With Purpose Are More Usable

Our results strongly suggest that there is value in the added expressiveness inherent in modeling permissions subject to purpose; they are more efficient and more effective. In particular, we find that there are significant differences in how users would express their preferences when they have the ability to distinguish among purposes. We also find evidence to support the conclusion in prior work, that the inability to express purposes may lead users to assume that permissions are required for apps to provide their basic functionality. We refer to this category of purposes as "internal purposes" in our study, as others have done in the past [20, 23, 44]. Many prior studies have found that purpose is an important contextual factor for a variety of reasons, and our work supports this conclusion [24, 45]. We also find diverse preferences along the axis of purpose, particularly for specific purposes such as advertising.

In incorporating purpose into permissions models, we find that there is a dramatic increase in the user burden required to elicit comprehensive preferences using the existing AOFU model – trivially, the number of questions increases proportionally to the number of purposes. Our methodology shows great promise in reducing user burden, such that the average user would experience less burden even with this new dimension, compared to the current model that cannot express purposes. While it is outside the scope of our study, there still exists a problem wherein the purpose for app permissions needs to be specified, and no canonical set of purposes has been specified for the Android platform. Based on our findings, we believe that such a taxonomy of purposes needs to be formalized and incorporated into the Android and iOS permissions systems. Perhaps assisted by the platform itself through inference, or using other tools, it would be possible to extend our methodology across a range of specific purposes. The benefit of our approach would be compounded in these instances, as the number of potential purposes could potentially be numerous.

5.3 Implications for GDPR Compliance

Our findings about the importance of the purpose for which an app requests a given permission is consistent with regulatory requirements such as those introduced by the EU GDPR, where consent for the collection of sensitive data is qualified by one or more specific purposes. Current regulation however does not explicitly consider the possibility of using machine learning to support data subjects in making privacy decisions. Yet, as already discussed earlier in this article, the burden on users associated with the increasing number of privacy decisions they have to make has become unrealistic. Machine learning offers the prospect of mitigating this burden, whether by providing recommendations to users, aggregating some decisions that are closely correlated for a given user (e.g., based on a privacy profile) or some other yet-to-be explored approach (e.g. dialog with a privacy assistant). Future UX research will need to look at how to best take advantage of machine learning functionality to empower users to regain control over their data, enabling them to more accurately specify their privacy preferences without imposing an unrealistic burden on them and without taking away their sense of autonomy.

5.4 Sacrificing Privacy to Gain Privacy

We note that, beyond the collection of mobile app privacy preferences, some of the additional information used in our predictive model is also sensitive (e.g. marital status, employment status). Additional research will be required to check whether people are comfortable disclosing this information for the purpose of getting help configuring their mobile app permission settings. Addressing this question would require comparing to what extent they might be more comfortable sacrificing user burden (i.e., spending more time configuring more app permissions) in return for revealing less information about themselves. Our intuition tells us that with strong guarantees that this information would be solely used for the purpose of helping them configure their mobile app permission settings, many users would likely be comfortable sharing this information with a privacy assistant, but this will need to be verified. In the worst case, we would end up with models that have somewhat less predictive power, or would need to ask users additional questions about their mobile app privacy preferences.

6 Conclusion

In this work, we administered a survey which collected participants' Android permissions preferences for a variety of apps under two conditions: one with purpose-

specific permissions and another with permissions that extend across all possible purposes. We analyzed responses using logistic regression and machine learning. Our aim was to discover whether machine learning could help mitigate the trade-off between effectiveness and efficiency when it comes to configuring Android app permissions. In doing so, we found that we can achieve the best of both worlds; app permissions can be made more expressive and thus more effective, but without sacrificing efficiency and overburdening users. In this paper, this is accomplished using machine learning to assign users to privacy profiles and using these profiles to infer many permissions for each user. In general, we would expect to see similar results with other machine learning techniques (e.g. collaborative filtering techniques or techniques such as those discussed in prior work [24]). In examining the studied contextual factors, we found that preferences change significantly subject to the more expressive permissions which incorporate purpose. There is also evidence that participants cannot distinguish between cases where purpose is unspecified and cases where the purpose is "internal" (for the app to provide basic functionality). This finding is consistent with prior research results. The added dimension of purpose, as well as the other tested contextual factors, do indeed make the Android permissions model more effective – namely the resulting permission settings can be configured to better align with people's actual privacy preferences. Beyond regulatory requirements, these findings further advocate for the introduction of mobile app permission settings that enable users to differentiate between different purposes.

Our results further show that models of people's mobile app privacy preferences that take into account the purpose(s) for which apps request permissions have greater predictive power than models that ignore purpose information. We leverage this additional predictive power to overcome the increase in user burden that would otherwise result from the introduction purpose-specific mobile app permission settings.

Acknowledgments

This study was supported in part by grants from DARPA and AFRL under the Brandeis project on Personalized Privacy Assistants (FA8750-15-2-0277) and by grants from the National Science Foundation Secure and Trustworthy Computing program (CNS-15-13957, CNS-1801316, CNS-1914486). The US Government is authorized to reproduce and distribute reprints for Governmental purposes not withstanding any copyright notice. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the NSF, DARPA, AFRL or the US Government.

References

[1] Alessandro Acquisti and Jens Grossklags. Privacy and rationality in individual decision making. *IEEE Security Privacy*, 3(1):26–33, Jan 2005.

[2] Yuvraj Agarwal and Malcolm Hall. ProtectMyPrivacy: Detecting and mitigating privacy leaks on ios devices using crowdsourcing. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '13, pages 97–110, New York, NY, USA, 2013. ACM.

[3] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. Your location has been shared 5,398 times!: A field study on mobile app privacy nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 787–796, New York, NY, USA, 2015. ACM.

[4] Panagiotis Andriotis, Martina Angela Sasse, and Gianluca Stringhini. Permissions snapshots: Assessing users' adaptation to the android runtime permission model. In *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6, Dec 2016.

[5] Paul Bankhead. Providing a safe and secure experience for our users. https://android-developers.googleblog.com/2018/10/providing-safe-and-secure-experience.html, Oct 2018. Accessed: 2019-02-24.

[6] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[7] Michael Benisch, Patrick Gage Kelley, Norman Sadeh, and Lorrie Faith Cranor. Capturing location-privacy preferences: Quantifying accuracy and user-burden tradeoffs. *Personal Ubiquitous Comput.*, 15(7):679–694, October 2011.

[8] Bram Bonné, Sai Teja Peddinti, Igor Bilogrevic, and Nina Taft. Exploring decision making with android's runtime permission dialogs using in-context surveys. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 195–210, Santa Clara, CA, 2017. USENIX Association.

[9] William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P. Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N. Sheth. TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Trans. Comput. Syst.*, 32(2):5:1–5:29, June 2014.

[10] Lujun Fang and Kristen LeFevre. Privacy wizards for social networking sites. In *Proceedings of the 19th International*

Conference on World Wide Web, WWW '10, pages 351–360, New York, NY, USA, 2010. ACM.

[11] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2):175–191, May 2007.

[12] Google. Android permissions overview. https://developer.android.com/guide/topics/permissions/overview, Jan 2019. Accessed: 2019-02-24.

[13] John Gower. A general coefficient of similarity and some of its properties. *Biometrics*, 27(4):857–871, 1971.

[14] Peter Hornyack, Seungyeop Han, Jaeyeon Jung, Stuart Schechter, and David Wetherall. These aren't the droids you're looking for: Retrofitting android to protect data from imperious applications. In *Proceedings of the 18th ACM Conference on Computer and Communications Security*, CCS '11, pages 639–652, New York, NY, USA, 2011. ACM.

[15] Qatrunnada Ismail, Tousif Ahmed, Apu Kapadia, and Michael K. Reiter. Crowdsourced exploration of security configurations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 467–476, New York, NY, USA, 2015. ACM.

[16] ISO/IEC 25022:2016. https://www.iso.org/standard/35746.html, Jun 2016.

[17] Haojian Jin, Minyi Liu, Kevan Dodhia, Yuanchun Li, Gaurav Srivastava, Matthew Fredrikson, Yuvraj Agarwal, and Jason I. Hong. Why are they collecting my data?: Inferring the purposes of network traffic in mobile apps. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(4):173:1–173:27, December 2018.

[18] Patrick Gage Kelley, Sunny Consolvo, Lorrie Faith Cranor, Jaeyeon Jung, Norman Sadeh, and David Wetherall. A conundrum of permissions: installing applications on an android smartphone. In *International conference on financial cryptography and data security*, pages 68–79. Springer, 2012.

[19] Bart P Knijnenburg. Information disclosure profiles for segmentation and recommendation. In *Symposium on Usable Privacy and Security (SOUPS)*, 2014.

[20] Jialiu Lin, Shahriyar Amini, Jason I. Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 501–510, New York, NY, USA, 2012. ACM.

[21] Jialiu Lin, Michael Benisch, Norman Sadeh, Jianwei Niu, Jason Hong, Banghui Lu, and Shaohui Guo. A comparative study of location-sharing privacy preferences in the united states and china. *Personal Ubiquitous Comput.*, 17(4):697–711, April 2013.

[22] Jialiu Lin, Bin Liu, Norman Sadeh, and Jason I. Hong. Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 199–212, Menlo Park, CA, 2014. USENIX Association.

[23] Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Shikun (Aerin) Zhang, Norman Sadeh, Yuvraj Agarwal, and Alessandro Acquisti. Follow my recommendations: A personalized privacy assistant for mobile app permissions. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 27–41, Denver, CO, 2016. USENIX Association.

[24] Bin Liu, Jialiu Lin, and Norman Sadeh. Reconciling mobile app privacy and usability on smartphones: Could user privacy profiles help? In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 201–212, New York, NY, USA, 2014. ACM.

[25] Kirsten Martin and Katie Shilton. Putting mobile application privacy in context: An empirical study of user privacy expectations for mobile devices. *The Information Society*, 32(3):200–216, 2016.

[26] Kristopher Micinski, Daniel Votipka, Rock Stevens, Nikolaos Kofinas, Michelle L. Mazurek, and Jeffrey S. Foster. User interactions and permission use on android. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 362–373, New York, NY, USA, 2017. ACM.

[27] Scott R. Moore, Huangyi Ge, Ninghui Li, and Robert W. Proctor. Cybersecurity for android applications: Permissions in android 5 and 6. *International Journal of Human–Computer Interaction*, 0(0):1–11, 2018.

[28] Jonathan Mugan, Tarun Sharma, and Norman Sadeh. Understandable learning of privacy preferences through default personas and suggestions. http://reports-archive.adm.cs.cmu.edu/anon/isr2011/abstracts/11-112.html, Aug 2011.

[29] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. Privacy expectations and preferences in an iot world. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*, pages 399–412, 2017.

[30] Helen Nissenbaum. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.

[31] Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L119:1–88, May 2016.

[32] Leysia Palen and Paul Dourish. Unpacking "privacy" for a networked world. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 129–136, New York, NY, USA, 2003. ACM.

[33] Bahman Rashidi, Carol Fung, and Tam Vu. Dude, ask the experts!: Android resource access permission recommendation with recdroid. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, pages 296–304, May 2015.

[34] Srikar Reddy. Android runtime permissions, recent policy changes and security vulnerabilities. https://medium.com/finbox/android-runtime-permissions-recent-policy-changes-and-security-vulnerabilities-935c5fc88f3d, Dec 2018. Accessed: 2019-02-24.

[35] Joel Rosenblatt. Uber data-scraping, surveillance detailed by ex-manager. https://www.bloomberg.com/news/articles/2017-12-15/uber-data-scraping-surveillance-detailed-in-ex-manager-s-letter, 2017. Accessed: 2019-02-24.

[36] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*, SOUPS'15, pages 1–17, Berkeley, CA, USA, 2015. USENIX Association.

[37] Fuming Shih, Ilaria Liccardi, and Daniel Weitzner. Privacy tipping points in smartphones privacy preferences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 807–816, New York, NY, USA, 2015. ACM.

[38] Irina Shklovski, Scott D. Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 2347–2356, New York, NY, USA, 2014. ACM.

[39] Helmut Strasser and Christian Weber. On the asymptotic theory of permutation statistics. *Mathematical Methods of Statistics*, 8, 02 1970.

[40] Joshua Tan, Khanh Nguyen, Michael Theodorides, Heidi Negrón-Arroyo, Christopher Thompson, Serge Egelman, and David Wagner. The effect of developer-specified explanations for permission requests on smartphone user behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 91–100, New York, NY, USA, 2014. ACM.

[41] Eric Thompson. App annie blog. https://www.appannie.com/en/insights/market-data/global-consumer-app-usage-data/, May 2017.

[42] Lynn Tsai, Primal Wijesekera, Joel Reardon, Irwin Reyes, Serge Egelman, David A. Wagner, Nathaniel Good, and Jung-Wei Chen. Turtle guard: Helping android users apply contextual privacy preferences. In *SOUPS*, 2017.

[43] Jennifer Valentino-DeVries, Natasha Singer, Michael H. Keller, and Aaron Krolik. Your apps know where you were last night, and they're not keeping it secret. https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html, 2018. Accessed: 2019-02-24.

[44] Max Van Kleek, Ilaria Liccardi, Reuben Binns, Jun Zhao, Daniel J. Weitzner, and Nigel Shadbolt. Better the devil you know: Exposing the data sharing practices of smartphone apps. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 5208–5220, New York, NY, USA, 2017. ACM.

[45] Daniel Votipka, Seth M. Rabin, Kristopher Micinski, Thomas Gilray, Michelle L. Mazurek, and Jeffrey S. Foster. User comfort with android background resource accesses in different contexts. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, pages 235–250, Baltimore, MD, August 2018. USENIX Association.

[46] Haoyu Wang, Jason Hong, and Yao Guo. Using text mining to infer the purpose of permission use in mobile apps. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 1107–1118, New York, NY, USA, 2015. ACM.

[47] Sarah Myers West. Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society*, 58(1):20–41, 2019.

[48] Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David Wagner, and Konstantin Beznosov. The feasibility of dynamically granted permissions: Aligning mobile privacy with user preferences. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 1077–1093, May 2017.

[49] Primal Wijesekera, Joel Reardon, Irwin Reyes, Lynn Tsai, Jung-Wei Chen, Nathan Good, David Wagner, Konstantin Beznosov, and Serge Egelman. Contextualizing privacy decisions for better prediction (and protection). In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 268:1–268:13, New York, NY, USA, 2018. ACM.

[50] Shomir Wilson, Justin Cranshaw, Norman Sadeh, Alessandro Acquisti, Lorrie Faith Cranor, Jay Springfield, Sae Young Jeong, and Arun Balasubramanian. Privacy manipulation and acclimation in a location sharing application. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 549–558, New York, NY, USA, 2013. ACM.

[51] Pamela Wisniewski, Bart P Knijnenburg, and Heather Richter Lipford. Profiling facebook users' privacy behaviors. In *Symposium on Usable Privacy and Security (SOUPS)*, 2014.

[52] Jierui Xie, Bart Piet Knijnenburg, and Hongxia Jin. Location sharing privacy preference: Analysis and personalized recommendation. In *Proceedings of the 19th International Conference on Intelligent User Interfaces*, IUI '14, pages 189–198, New York, NY, USA, 2014. ACM.

Appendix

Main Survey

Your Android phone has settings, called app permissions, that allow you to control what data your apps can access. When a given permission setting is denied for a particular app, that app is not able to access the corresponding data.

Examples of app permissions include permission to access your location, your contacts, your calendar, and more. Some apps request certain permissions for **internal purposes**, namely to deliver their basic functionality. They might not work properly if the permission is not granted. Some apps request certain permissions for other purposes, such as showing you **personalized advertisements, or generally collecting and analyzing data about you**. Some apps request certain permissions for **unspecified purposes, or for no particular purpose at all**.

The following are the permissions we will ask you questions about in our survey.

Calendar
If you allow an app access to the Calendar permission, the app is allowed to:
- Read your calendar data, including your events, appointments, and any other data you have added to your calendar
- Make changes to your calendar

Location
If you allow an app access to the Location permission, the app is allowed to:
- Access and record your approximate location (using the cellular network and WiFi networks in your vicinity)
- Access and record your precise location (using your phone's GPS, as well as the cellular network and WiFi networks in your vicinity)

Contacts
If you allow an app access to the Contacts permission, the app is allowed to:
- Read what contacts you have saved on your device, and the contact entries' information (such as phone number, email address, social media account IDs)
- Write and make changes to contact entries you have saved on your device, and add new contact entries
- Access the different accounts you have saved on your device, including your Google accounts (and their associated contacts), Facebook accounts, and any other accounts you have registered on your phone which have information that can appear on your contact list

If you need help to remember what the different permissions allow access to, look for the ⓘ icon in the survey - you can hover over it with your mouse and read the description.

On the following page, you will see a screenshot of the Google Play Store page for an app. Look through and think about things that you would normally look for when making the decision to download and install an app on your phone.

Please note that you will see a screenshot and not an interactive webpage. You can't click on things in the screenshot, and it won't react to your mouse cursor like the real website would. *(The user is presented with a screenshot of an app's Google Play Store page.)*
Approximately how often do you use this app now?
- ○ Daily
- ○ Weekly
- ○ Monthly
- ○ Yearly
- ○ I uninstalled it and/or I don't use it anymore
- ○ I have never used this app before

What is the name of the app in the screenshot? *(Attention check question.)*

How familiar are you with the functionality and features of the app? *(Likert scale response from Not familiar at all to Extremely familiar.)*

Suppose the app asks for access to permissions on your phone. Think carefully about the app and what you would do if the app requested these permissions. For each permission, tell us whether you would allow access or deny access.
If you need help to remember what the permissions allow access to, look for the ⓘ icon in the survey - you can hover over it with your mouse and read the description.

	I would allow access	I would deny access
Calendar ⓘ		
Location ⓘ		
Contacts ⓘ		

In this section you will be asked questions about whether you would allow access to the Calendar, Location, and Contacts permissions for the app, for different purposes. Although you have already answered general questions about your preferences to allow or deny this app's permissions earlier in the survey, note that the following questions are different. Take a moment to carefully consider how you would answer them, given the more specific information they are asking about.

Given the 3 different purposes listed below, would you allow or deny this app access to your **Calendar**?

	I would allow access to Calendar	I would deny access to Calendar
Suppose the app requests access to your Calendar for internal purposes, which allow the app to deliver its basic functionality...		
Suppose the app requests access to your Calendar for personalized advertisements, or generally collecting and analyzing data about you...		
Suppose the app requests access to your Calendar for any other, unspecified or unknown purpose...		

Given the 3 different purposes listed below, would you allow or deny this app access to your **Location**?

	I would allow access to Location	I would deny access to Location
Suppose the app requests access to your Location for internal purposes, which allow the app to deliver its basic functionality...		
Suppose the app requests access to your Location for personalized advertisements, or generally collecting and analyzing data about you...		
Suppose the app requests access to your Location for any other, unspecified or unknown purpose...		

Given the 3 different purposes listed below, would you allow or deny this app access to your **Contacts**?

	I would allow access to Contacts	I would deny access to Contacts
Suppose the app requests access to your Contacts for internal purposes, which allow the app to deliver its basic functionality...		
Suppose the app requests access to your Contacts for personalized advertisements, or generally collecting and analyzing data about you...		
Suppose the app requests access to your Contacts for any other, unspecified or unknown purpose...		

Post-Survey Questions

- Please select the category which best describes your age range.
- What is the highest level of school you have completed or the highest degree you have received?
- What is your gender assigned at birth?
- What is your ZIP code?
- What category best describes where you live?
- Please select the category which best describes your marital status.
- Which statement best describes your current employment status?
- Approximately how long have you been using a smartphone?
- What kind of smartphone do you use?
- Approximately how frequently do you use a smartphone?
- Approximately how many apps do you have installed on your smartphone?
- Approximately how many apps on your smartphone have you used in the past week?
- What Android version do you have on your smartphone?
- Not including this survey, approximately how many privacy-related surveys have you completed in the past year?
- How did you find the privacy-related surveys you completed in the past?

Logistic Regression Table

Factors with very strong significance ($pr\chi^2 \leq 0.01$) are marked with two asterisks. Factors with strong significance ($pr\chi^2 \leq 0.05$) are marked with a single asterisk.

Factors	*df*	Calendar		Location		Contacts		Calendar × Internal		Calendar × Ads		Calendar × Other	
		χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
App Familarity	4	312.5	$\leq$0.01**	341.23	$\leq$0.01**	308.76	$\leq$0.01**	238.26	$\leq$0.01**	222.32	$\leq$0.01**	169.75	$\leq$0.01**
App Usage Freq.	5	450.22	$\leq$0.01**	413.52	$\leq$0.01**	394.26	$\leq$0.01**	262.22	$\leq$0.01**	350.66	$\leq$0.01**	248.2	$\leq$0.01**
App Category	24	82.747	$\leq$0.01**	138.02	$\leq$0.01**	76.292	$\leq$0.01**	74.084	$\leq$0.01**	56.923	$\leq$0.01**	0	N.S.
Age	3	27.96	$\leq$0.01**	6.4626	N.S.	29.984	$\leq$0.01**	11.065	$\leq$0.05*	27.596	$\leq$0.01**	6.1866	N.S.
Education	7	24.61	$\leq$0.01**	42.512	$\leq$0.01**	29.992	$\leq$0.01**	19.97	$\leq$0.01**	37.923	$\leq$0.01**	0	N.S.
Gender	2	1.0494	N.S.	1.329	N.S.	0.3974	N.S.	0.262	N.S.	0.6321	N.S.	0.8382	N.S.
City Size	3	31.3	$\leq$0.01**	17.002	$\leq$0.01**	27.854	$\leq$0.01**	28.013	$\leq$0.01**	38.87	$\leq$0.01**	8.4015	$\leq$0.05*
Marital Status	5	40.956	$\leq$0.01**	45.998	$\leq$0.01**	59.944	$\leq$0.01**	26.021	$\leq$0.01**	65.398	$\leq$0.01**	19.835	$\leq$0.01**
Employment	8	13.672	N.S.	12.362	N.S.	15.509	N.S.	6.9251	N.S.	23.735	N.S.	0	N.S.
Phone Usage Freq.	3	12.741	$\leq$0.01**	5.5258	N.S.	11.169	$\leq$0.05*	9.1237	$\leq$0.05*	9.0246	$\leq$0.05*	6.1913	N.S.
Phone Usage Duration	3	30.637	N.S.	15.135	$\leq$0.01**	30.347	N.S.	15.516	N.S.	24.36	N.S.	10.594	N.S.
# Apps Installed	3	17.775	$\leq$0.01**	20.769	$\leq$0.01**	22.761	$\leq$0.01**	10.305	$\leq$0.05*	9.801	$\leq$0.05*	2.4036	N.S.
# Apps Used	3	70.021	$\leq$0.01**	49.999	$\leq$0.01**	87.794	$\leq$0.01**	26.238	$\leq$0.01**	96.416	$\leq$0.01**	43.322	$\leq$0.01**
Android Version	9	40.407	N.S.	35.401	N.S.	34.189	N.S.	37.186	N.S.	35.28	N.S.	8.0093	N.S.
# Privacy Surveys	3	5.8137	N.S.	12.443	$\leq$0.01**	20.343	$\leq$0.01**	7.7106	N.S.	20.689	$\leq$0.01**	6.535	N.S.

Factors	*df*	Location × Internal		Location × Ads		Location × Other		Contacts × Internal		Contacts × Ads		Contacts × Other	
		χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
App Familarity	4	238.8	$\leq$0.01**	271.31	$\leq$0.01**	232.24	$\leq$0.01**	265.09	$\leq$0.01**	216.17	$\leq$0.01**	186.84	$\leq$0.01**
App Usage Freq.	5	246.27	$\leq$0.01**	415.82	$\leq$0.01**	336.8	$\leq$0.01**	297.34	$\leq$0.01**	372.51	$\leq$0.01**	238.93	$\leq$0.01**
App Category	24	127.62	$\leq$0.01**	86.111	$\leq$0.01**	53.21	$\leq$0.01**	74.404	$\leq$0.01**	18.726	N.S.	0	N.S.
Age	3	2.0877	N.S.	13.709	$\leq$0.01**	5.8693	N.S.	7.9095	$\leq$0.05*	8.7713	N.S.	4.6471	N.S.
Education	7	25.153	$\leq$0.01**	39.567	$\leq$0.01**	28.776	$\leq$0.01**	24.874	$\leq$0.01**	0	N.S.	4.8749	N.S.
Gender	2	0.7467	N.S.	1.4218	N.S.	5.1683	N.S.	2.432	N.S.	0.4015	N.S.	0.4795	N.S.
City Size	3	15.654	$\leq$0.01**	31.437	$\leq$0.01**	12.221	$\leq$0.01**	25.16	$\leq$0.01**	19.745	$\leq$0.01**	6.3056	N.S.
Marital Status	5	23.716	$\leq$0.01**	57.664	$\leq$0.01**	36.609	$\leq$0.01**	38.645	$\leq$0.01**	36.342	$\leq$0.01**	19.447	$\leq$0.01**
Employment	8	7.3579	N.S.	19.532	N.S.	0	N.S.	6.931	N.S.	0	N.S.	0	N.S.
Phone Usage Freq.	3	8.7472	$\leq$0.05*	9.8276	$\leq$0.05*	7.9295	$\leq$0.05*	5.7587	N.S.	6.6361	N.S.	6.0856	N.S.
Phone Usage Duration	3	8.0061	N.S.	23.226	$\leq$0.01**	13.106	$\leq$0.01**	20.373	$\leq$0.01**	18.491	$\leq$0.01**	7.9586	N.S.
# Apps Installed	3	14.084	$\leq$0.01**	11.648	$\leq$0.01**	5.589	N.S.	8.7129	N.S.	1.7364	N.S.	1.7594	N.S.
# Apps Used	3	16.346	$\leq$0.01**	91.588	$\leq$0.01**	81.582	$\leq$0.01**	35.207	$\leq$0.01**	60.961	$\leq$0.01**	37.508	$\leq$0.01**
Android Version	9	26.003	N.S.	54.676	N.S.	26.133	$\leq$0.01**	24.407	N.S.	16.092	N.S.	0	N.S.
# Privacy Surveys	3	14.175	$\leq$0.01**	26.155	$\leq$0.01**	11.267	$\leq$0.05*	8.3121	$\leq$0.05*	11.466	$\leq$0.01**	6.9094	N.S.