EXAMINING USERS' APPLICATION PERMISSIONS ON ANDROID MOBILE DEVICES

By

MUHAMMAD IRTAZA SAFI
B.S Lahore University of Management Sciences 2016

A thesis submitted in partial fulfilment of the requirements
for the degree of Master of Science
in the Department of Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2018

Major Professor: Pamela Wisniewski

ABSTRACT

Mobile devices have become one of the most important computing platforms. The platform's portability and highly customized nature raises several privacy concerns. Therefore, understanding and predicting user privacy behavior has become very important if one is to design software which respects the privacy concerns of users. Various studies have been carried out to quantify user perceptions and concerns [23,36] and user characteristics which may predict privacy behavior [21,22,25]. Even though significant research exists regarding factors which affect user privacy behavior, there is gap in the literature when it comes to correlating these factors to objectively collected data from user devices. We designed an Android application which administered surveys to collect various perceived measures, and to scrape past behavioral data from the phone. Our goal was to discover variables which help in predicting user location sharing decisions by correlating what we collected from surveys with the user's decision to share their location with our study application. We carried out logistic regression analysis with multiple measured variables and found that perceived measures and past behavioral data alone were poor predictors of user location sharing decisions. Instead, we discovered that perceived measures *in the context of past behavior* helped strengthen prediction models. Asking users to reflect on whether they were comfortable sharing their location with apps that were already installed on their mobile device was a stronger predictor of location sharing behavior than general measures regarding privacy concern or past behavioral data scraped from their phones. This work contributes to the field by correlating existing privacy measures with objective data, and uncovering a new predictor of location sharing decisions.

This thesis is dedicated to my family who have supported me throughout my graduate journey.

ACKNOWLEDGMENT

I would like to express my gratitude to my advisor Dr. Pamela Wisniewski, for the exceptional guidance and encouragement she provided throughout this study.

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF ACRONYMS (or) ABBREVIATION x

CHAPTER ONE: INTRODUCTION AND LITERATURE REVIEW 1

Introduction 1

Related Work 4

Research Framework 9

CHAPTER TWO: METHODS 16

Overview 16

Study Design 16

Operationalization of Measures 17

Data Collection and Recruitment 22

Study Procedure 23

System Implementation 29

Data Analysis Approach 30

CHAPTER THREE: RESULTS 32

Descriptive Statistics 32

Simple Effects Analysis 33

Logistic Regression Results 35

CHAPTER FOUR: DISCUSSION AND CONCLUSION 41

Summary of Main Findings 41

Implications for Mobile Privacy Research 42

Implications for Design 43
Limitations and Future Research 43
Conclusion 44
APPENDIX: IRB APPROVAL FOR HUMAN RESEARCH 45
LIST OF REFERENCES 48

LIST OF FIGURES

Figure 1: Location comfort question..... 24

Figure 2: Revoke location question 25

Figure 3: Allow location question..... 26

Figure 4: Pre location request 27

Figure 5: Location request screen 28

Figure 6: Completion screen..... 29

LIST OF TABLES

Table 1: Descriptive statistics of variables 31

Table 2: Simple Effects Analysis..... 33

Table 3: Comparison of means 34

Table 4: Perceived Constructs Only (Nagelkerke R squared = 13.9%)..... 35

Table 5: Behavioral Scraped Data (Nagelkerke R squared = 2.1%)..... 36

Table 6: Percieved In Context of Past Behavior Only (Nagelkerke R squared = 16%) 37

Table 7: Combined Model (Nagelkerke R squared = 21.3%) 38

Table 8: Summary Of Models..... 39

LIST OF ACRONYMS (or) ABBREVIATION

CPM	Communication Privacy Management
IUIPC	Internet Users Information Privacy Concerns
MUIPC	Mobile Users Information Privacy Concern.
RQ	Research Question.

CHAPTER ONE: INTRODUCTION AND LITERATURE REVIEW

Introduction

Since the advent of the iPhone in 2007, smartphone usage has seen tremendous growth. According to Statista, the number of smartphones sold grew from 122 million in 2007 to 1.5 billion in 2017 [37]. This growth in device sales has in turn fueled the growth of the mobile app industry. Mobile app revenue is projected to reach 188 billion dollars by 2020 [38] and mobile users now account for a majority (52%) internet traffic share [35]. The rise in mobile internet traffic share has promoted the growth of new social networks and communication apps such as Instagram, Snapchat, WhatsApp and Facebook. As of Q2 2018, Facebook had 2.23 billion monthly active users and 90% of its revenue came from mobile devices [39].

In order to monetize free services, social media companies such as Facebook charge money to run product advertisements for different businesses. To increase the effectiveness of the advertisements, social media companies mine user data to 'learn' individual user characteristics to deliver highly targeted adverts. For instance, a business owner can specify target characteristics as precise as age, gender, sexual orientation, political outlook, religion, field of study, relationship status and much more [40]. So successful is this ad-based business model that as Facebook's profits grew from $230 million in 2009 to $15.9 billion in 2018 [41], the Economist declared data to be more valuable than oil [34]. As a result, the marketing and commoditization of user data has jeopardized user privacy [16]. Companies dealing with user data have been frequently involved in practices which violate users' basic assumption of privacy. The first three months of 2018 alone have seen ten major data breaches/scandals [7]. The worst

of the scandals being Facebook's Cambridge Analytica debacle where the company's lax data policies allowed a third party firm (Cambridge Analytica) to gather data on 87 million Facebook users [13]. Understandably, this has led to increased user anxiety about their privacy. An online survey of ten thousand consumers by IBM [42] found that 73% of respondents think that businesses care more about profits than users' security needs and that 60% of respondents are more concerned about cybersecurity than a potential war.

Although such evidence suggests that users' concern about privacy are high, research has revealed discrepancies between people's stated privacy concerns and their actual privacy behaviors. Known widely as the privacy paradox [6], this phenomenon has made it difficult to predict user behavior based on their stated privacy preferences. As such researchers have devised many scales to predict and explain users' privacy intentions [36][22][21][30], but there is still a gap when it comes to correlating these perceptions with objectively collected data about users' privacy decisions. Meanwhile, computational scientists have taken a different approach by attempting to predict users' privacy behavior based on their past behavioral data. Past behavioral data can be things like the number of apps installed, the kinds of apps installed, the kinds of permissions granted etc. Computational approaches have generally used objective variables collected from the device but have not correlated them with well known privacy measures and constructs.

In this study, we or goal was to bridge the gap between social and computational science approaches to studying and predicting end users' mobile privacy behaviors. We developed a mobile app for Android that combined perceived privacy constructs (using a survey embedded in the app via the Qualtrics API) and scraped behavioral data (collected unobtrusively by the app) regarding the apps participants had installed on their Android smartphones and the permissions

users' granted to these apps. As a dependent variable, we asked users to share their location permission with the app as part of the user study and captured whether this permission was granted or denied. Having both perceived measures and objective data allowed us to study the correlation between stated attitudes to privacy and actual behavior. Additionally, we were able to ask users if they were comfortable with their past location sharing decisions as collected from the phone (this was the *percieved in context of past behavior*) variable.

After collecting the data, we first carried out a simple effects analysis to figure out the variables most closely related to the location sharing decision. We then carried out logistic regression in a step by step fashion to come up with the strongest model to predict if the user would grant our app the location permission.

We found that most perceived measures on their own are weak predictors of user location sharing behavior and that even objective past privacy behavior scraped from phone metadata were not significantly linked to the decision to grant location permission. We discovered that using a *perceived in context of past behavior* variable (the percentage of apps to which the user was comfortable giving their location permission) significantly improved the predictive power of our model. The model using only perceived variables explained 13.9% of the variance, the model with only behavioral scraped data explained 2.1% of the variance, and the combined model with percieved measures and the percieved in context of past behavior variable explained 21.3% of the variance.

Predicting location sharing decisions better can help researchers understand the factors behind user privacy decisions. App makers can also use this knowledge to understand how users will

react to their app asking for the location and can then choose to not ask the permission for a better user experience.

Related Work

Privacy on Mobile Smartphones

Approaches to studying privacy on mobile devices have generally been divided in to 'social' and 'computational'. Social approaches have focused on conducting survey studies to gauge user behavior whereas computational approaches have focused more on predicting user behavior using on device data.

Mobile Privacy and Behavioral Intent

Survey based studies on mobile privacy have been largely conducted in the fields of Information Systems and Human-Computer Interaction. A key theme of this work has been trying to predict user behavior based on survey responses. Survey questionnaires generally try to measure the user's attitude towards specific topics using specially designed survey constructs. A commonly accepted practice in this research is that user behavior can be predicted based on users' attitudes and behavioral intent is a consequence of the theory of planned behavior [1]. The theory states that behavioral beliefs inform user attitudes towards behaviors which then lead to behavioral intention which directly impacts user behavior. Researchers have thus developed various constructs to quantify different aspects of user beliefs, feelings, intentions, and attitudes towards mobile privacy in order to understand how their behavior affects privacy decisions.

Smith et al's[28] work was one of the first scales developed to measure user concern for information privacy. The work introduced a fifteen item instrument which measured concerns regarding data collection, unauthorized secondary use (the use of information for purposes other than it was collected for), improper access (the access of data by individuals not permitted to access it), and errors in data handling (data leaks and privacy breach accidents).One of the recent works in this area has been the Measuring Mobile Users Concern for Information Privacy (MUIPC

) scale by Xu et al [36]. This is a three-factor scale which determines users' concern for information privacy by measuring their concern regarding the misuse of their shared data, the degree of intrusion felt by them, and their perceived surveillance. MUIPC uses the Communication Privacy Management (CPM) theory as a its theoretical foundation. The CPM theory states that individual information sharing decisions are based on the perceived benefits and costs of information disclosure [43]. CPM introduces the concept of boundary coordination which involves the data sharer and recipient to coordinate the permeability rules, ownership rules, and linkage rules between themselves to minimize the data sharer's privacy concerns. The three factors of MUIPC capture these boundary rules as follows: permeability (percieved survellience), ownership (perceived intrusion), linkage (secondary use of personal information). MUIPC has since been referenced by several studies aiming to understand different facets of user behavior [9][26][15].

Similarly, work done by Page et al [22] has shown how measuring specific communication styles can influence decisions to share location. The study identified the FYI Communication style as the style where users would rather infer availability and social information about others than interact with them in person. Users ranking high on the FYI Communication style were more

likely to share their location in location sharing social networks. Further work by Page et al[23] found that the desire to preserve boundaries was the main source of privacy concerns regarding location sharing. The study found that when people thought that location sharing services would change their relationships with others, they also got concerned about being forced to interact with others or being inundated with information from other people. Heavy social media users were found to be less concerned about privacy boundaries. Barkhus et al [5] distinguished between location tracking services (services based on other parties tracking the user) and position aware services (services which rely on the device's knowledge of it's location). They found that even though user perceived both of these services as equally useful, location tracking services produced far more concern for privacy. Guha et al[12] shed light on the practice of deceptive location sharing (i.e users sharing incorrect location information) and found that users engaged in this practice due to various concerns about privacy, and as acts of boundary and impression management.

Additionally, researchers have profiled power users as users who use technology to the fullest extent, can easily adapt to technological changes, and feel that technology is an integral part of their life[21] .Power users were found to prefer customization (tailoring interfaces to learnt user preferences) given that the customization was user initiated. If the customization was done automatically by the system power users felt a loss of agency and did not feel as positively towards the presented content [30].

Computational Approaches to Mobile Privacy

Computational approaches to mobile privacy have generally focused on designing tools to detect dangerous behavior, designing systems to increase user awareness about privacy threatening behavior, and using machine learning to classify malware which threatens user data. Researchers have built several tools that rely on program analysis to unearth privacy risk. For instance, Artz et al [4] developed FlowDroid, a tool which uses static taint analysis to track data paths through Android applications. Many other researchers have relied on similar approaches to uncover privacy risks. For instance Bartel et al[16] conducted program analysis to find how information travels between different components of applications, identifying several instances of unwarranted data leakages. Furthermore, works such as [3] have applied machine learning classifiers to detect android malware. Wang et al [31] used text mining to infer the purpose of permission use in mobile apps. The CPM theory [43] states that users balance the risk and rewards in making data sharing decisions, Wang et al's work builds a machine learning model which classifies why the location permission is needed based on text strings in decompiled app code. This helps users make more informed decisions regarding permission sharing. Similar to this, Li et al [18] built an automated system which maps the permissions to the part of the app UI which requests it so that the users are more aware of why the permissions are being requested. Almuhimidi et al [2] designed a permissions manager for Android devices which periodically sent 'nudges' to users to remind them of the different kinds of data being collected, and the frequency with which it was being collected. The study resulted in 58% of participants restricting some of their permissions. Fawaz et al's LP Guardian tool [10] is one of the more comprehensive solutions in this domain. The software works intelligently to make sure apps only access location when the user expects, anonymizes location in the background for certain

services and limits user profiling by different apps without significantly affecting the user experience. Overall, this body of work relies on providing the user more information regarding the kinds of privacy risks present and the actual usage of their shared data in order to help them make more informed decisions.

Combining Social and Computer Science Approaches

The above-mentioned approaches to mobile privacy have been either entirely computational (i.e., using objective variables to make machine learning models or designing software to identify privacy and security threats), or entirely survey based (i.e., using perceived measures collected from surveys to predict user privacy behavior). There is a dearth of research that has merged these two approaches by collecting both survey data and scraping behavioral data to try to predict user privacy behavior. For instance, work done by Lin et al. [44] studied app permissions by downloading apps through the google play store and then using static analysis to study sections of specific code where permissions were used. Participants were then recruited to answer survey questions about the downloaded apps. In this way, the researchers were able to correlate user preferences with app permissions and come up with a set of privacy profiles. Similarly, Ghosh et al [11] used phone metadata, such as call frequency and call length to try to predict user privacy behaviors. They found that higher call response rate, higher missed call rate, and higher number of new contacts were associated with a low concern for privacy.

Our work falls into the category of combining social science and computational approaches. Unlike Lin et al. [44], we try to predict actual user behavior (whether the user would grant location permission) or not. Furthermore, while Lin et al [35] ask generalized questions regarding apps to participants, we tailor our questions to the specific apps on the participants'

devices. This novel contribution is what we consider asking perceived measures *in the context of past behavior*. For example, instead of asking users how comfortable they are sharing their location permission with mobile apps on their phone, we scraped the app manifest of installed applications on their mobile smartphone and asked them whether or not they were comfortable with each app installed on their phone having access to their location. We believe this new type of variable represents a hybrid construct that combines the strengths of social and computational science in a way that helps us better understand end-user privacy behaviors. Overall, we make the following contributions to the field of end-user mobile privacy:

1) We evaluate the perceived measures which best predict user location sharing behavior.
2) We show that behavioral scraped data is not a good predictor of user location sharing behavior.
3) We show that perceived measures *in the context of past behavior* can help improve models for predicting user location sharing behavior.

Research Framework

In this study, we predict if the user will share their location (LOCATION_GIVEN) with our app based on the following three distinct classes of independent variables:

1) *Perceived measures*
2) *Behavioral scraped variables/data*
3) *Perceived measures in context of past behavior*.

Perceived variables are variables that directly measure how a user feels about a certain topic, behavior or action. Perceived measures are generally computed by having the user answer survey

questions which are grouped into sets called constructs. Constructs are created by researchers to conceptualize latent variables (i.e., variables which are inferred from other observed variables). A construct must be valid [8] (i.e., it must pass various construct validity tests so that it is confirmed to be measuring what it claims to measure). We used the Cronbach alpha to measure construct validity and set a cut off of 0.6 to discard constructs. An example of a pre-validated construct is the Behavioral Intention construct used by Xu et al. [36] to quantify the degree to which users plan on disclosing personal information and using mobile apps.

In contrast, behavioral or meta data can be scraped unobtrusively from the users' device. This data is objective in that it only collects data on the device without the user's subjective input. Studies done by computer scientists often use objective variables since they can design software to access this data [25]. An example of behavioral scraped data being used to make predictions is Seneviratne et al.'s[25] work which uses the kinds of installed apps on the phone to predict user traits such as religion, gender and relationship status.

While perceived variables are measured in surveys isolated from the context of actual use, and behavioral scraped variables are objective data that do not take into account a users' perceptions or attitudes, we define perceived measures in *the context of past behavior* as a combination of perceived and scraped behavioral data. As such, we propose the following high-level research questions:

1) How well do pre-validated perceived measures predict users' actual privacy behavior (i.e., location sharing)?
2) How well does behavioral scraped data predict users' actual privacy behavior?

3) How well do perceived in context of past behavior variables predict users' actual privacy behavior?
4) Can these three types of variables be combined to improve the ability to predict users' actual privacy behavior?

Dependent Variable: End-User Privacy Behavior

A number of studies such as have attempted to predict end-user's privacy behaviors in mobile contexts. Lin et al[19] found that if developers were clearer in revealing the purpose of requesting sensitive privacy permissions, users were more likely to have positive feelings regarding the experience of granting permissions. Furthermore, Shih et al[27] found that users become more privacy aware restrictive and apprehensive in granting permissions when a vague purpose was given for the permission request. The above research shows that user privacy concerns have a direct impact on their privacy sharing behavior, if software designers can anticipate these privacy concerns, they can design better experiences which preserve user privacy and help retain users.

The dependent variable that we chose as a measure of user privacy behavior was whether the user agreed to give their location (GPS location) to our study app. We frame this as a 'Yes/No' choice. Location services are the cornerstone of personalized mobile content and form an important part of how targeted advertisements are delivered. Social networks such as Tinder and Foursquare and mapping applications like Google Maps all rely on the user sharing their location with the community or app. Therefore, we think that finding models to better predict if the user

shares their location can go a long way in helping such applications design better use experiences.

Perceived Measures

We decided to choose the following constructs as our perceived measures. The constructs are described with their questions in more detail in the appendix.

Behavior Intention: Developed by Xu et al [33], in our study, this construct measures the user's intention to disclose personal information and to use mobile apps in the next 12 months. The theory of planned behavior states that user behavior can be predicted using attitudes towards behavior and behavioral intent [1]. Xu et al [36] showed that increased user privacy concern reduced the user's behavioral intention to disclose personal information and to use mobile apps. Therefore, we want to measure the users' intention to use new apps and to share their data with new apps as this would correlate with if they decide to share their location.

Perceived Surveillance: Developed by Xu et al [36], this construct quantifies the user's perception of if they are being surveilled and if too much information is being collected about them. Perceived surveillance is one of the factors of MUIPC developed by Xu et al [36], this measure is rooted from the dimension of 'collection' from Malhotra et al's Internet Users' Information Privacy Concerns (IUIPC) scale[20]. Malhotra et al note that data collection is the starting point of various privacy concerns. The dimension of collection in IUIPC measures the degree to which a person is concerned about the specific data that others have relative to the value of benefits received. As the CPM [43] theory states that user privacy decisions are based

on based on an assessment of the perceived benefits and risk associated with the decision, we therefore add the perceived survellience measure to quantify this user's perception of the balance of survellience and benefit.

Perceived Intrusion: Developed by Xu et al [32], this construct quantifies the user's perception of intrusion caused by using mobile apps. Xu et al [32] conducted a survey based study and found that perceived intrusion shaped individuals' views about the privacy practices of a specific website. We therefore use this measure in order to correlate it to the privacy decision of sharing location. We want to see if the percieved intrusion felt due to use of mobile apps has an effect on user location sharing behavior.

Secondary Use of Personal Information: Developed by Smith et al[28], this construct quantifies the degree of concerns users have about their information being used for purposes other than what it was collected for. Solove et al [29] notes that "the potential for secondary use generates fear and uncertainity over how one's information will be used in the future, creating a sense of powerlessness and vulnerability." Xu et al [36] also make secondary use a factor of their MUIPC scale. We wanted to see how the perception of secondary use affects location sharing behavior, therefore we include this measure in our survey.

FYI About Myself: Developed by Page et al [22], this constructs quantifies the 'FYI Communication Style'. People with the FYI communication style prefer to keep in touch with others without having to personally interact with them. People with this communication style were shown to be more willing to share their location in location sharing social networks. Since

this measure has already been shown to have an impact on location sharing decisions, we include this in our survey as well.

Boundary Preservation: Developed by Page et al [23], this construct quantifies the perceptions of users regarding how the usage of apps would positively or negatively impact their relationships. Page et al [23] found that boundary preservation concerns were a major factor influencing concerns regarding location sharing, therefore we are including this construct in our survey.

Power Usage: Developed by Marathe et al [21], this scale quantifies the degree to which a user is a 'power user'. Power users are technologically adept and use their gadgets to the fullest potential. Kang et al[14] found that power users are less likely to share personal information on personalized mobile sites but become more revealing when they interact with non-personalized mobile content. Since our app is a highly personalized, we included this measure to see how it affects users' location sharing behavior.

Behavioral Scraped Variables

We collected the following scraped variables from user devices as part of this study:

Number of installed apps: The number of installed apps on the device.

Total dangerous permissions granted: The number of dangerous android permissions granted on the device. We included this measure as it indicates that the user is more willing to grant dangerous android permissions and is perhaps less concerned about privacy. According to

Google, dangerous permissions "cover areas where the app wants data or resources that involve the user's private information, or could potentially affect the user's stored data or the operation of other apps."[45]

Location ratio: The number of apps with location permission divided by the total number of apps. This measure serves to quantify the user's past behavior about location sharing. We include this as it provides a snapshot of the user's past location sharing behavior.

Perceived in Context of Past Behavior

We combined our behavioral scraped variables and perceived measures to devise the following perceived in context of past behavior variables:

Location comfort (percentage): The percentage of apps for which the users are comfortable giving their location. This variable captures the users' feelings regarding their past privacy behavior. The theory of planned behavior[1] suggests that user behavior is based on users' attitude towards that behavior. This variable captures the attitude towards past location sharing behavior.

Location revoke (percentage): The percentage of apps for which the users were uncomfortable giving location and would revoke the location permission. Similar to location comfort percentage, this captures the users' intention to revoke location for apps to which they are uncomfortable giving location. The theory of planned behavior states that behavioral intent is the basis for actual behavior[1].

CHAPTER TWO: METHODS

Overview

We recruited one hundred and fourteen participants for our study. The study involved downloading our Android app, and then going through the app screens till completion. Once the data was collected we used logistic regression to try to come up with the best possible model to predict 'LOCATION_GIVEN'.

Study Design

The goal of the study was to verify if existing perceived measures were suitable to predict actual user behavior i.e LOCATION_GIVEN in our case, and if the addition of new variables could help strengthen prediction models. As explained before, our study relies on three categories of variables: 1) Perceived, 2) Behavioral scraped, and 3) Perceived in context of past behavior
To collect all three kinds of data from the user in a seamless fashion, we built an Android app which incorporated a Qualtrics survey. The Qualtrics survey was used to collect the perceived measures. A background process in the app scraped additional data from the user's device. Furthermore, we asked 'in context' (described later on) to help us generate our *perceived in context of past behavior* variables. All measures were asked to all of our participants.

Operationalization of Measures

In this section, we describe how we operationalized each of the measures listed in the research framework section.

Behavior Intention

This measure had three questions all asked on 5-point Likert scale. A higher score indicated a greater behavioral intention to disclose information and use mobile apps in the next 3 months.

Items:

1) I am likely to disclose my personal information to use mobile apps in the next 3 months.
2) I predict I will use new mobile apps in the next 3 months.
3) I intend to use mobile apps in the next 3 months.

Perceived Surveillance

This measure had three questions, all asked on a 5-point Likert scale. A higher score indicated a higher perception of surveillance through mobile devices and apps.

Items:

1) I believe that the location of my mobile device is monitored at least part of the time.
2) I am concerned that mobile apps are collecting too much information about me.
3) I am concerned that mobile apps may monitor my activities on my mobile device.

Perceived Intrusion

This measure had three questions asked on a 5-point Likert scale. A higher score indicated a higher perception of intrusion.

Items:

1) I feel that as a result of my using mobile apps, others know about me more than I am comfortable with.
2) I believe that as a result of my using mobile apps, information about me that I consider private is now more readily available to others than I would want.
3) I feel that as a result of using mobile apps, information about me is out there that, if used, will invade my privacy.

Secondary Use of Personal Information

This measure had three questions asked on 5-point Likert scale. A high score indicates that the user is concerned that their information might be used for purposes other than it was collected for.

Items:

1) I am concerned that mobile apps may use my personal information for other purposes without notifying me or getting my authorization.
2) When I give personal information to use mobile apps, I am concerned that apps may use my information for other purposes.
3) I am concerned that mobile apps may share my personal information with other entities without getting my authorization.

FYI About Myself

This measure had six questions all on a 5-point Likert scale. A high score on this construct means that the user is high on the 'FYI About Myself' personality type.

Items:

1) I want others to know what I am up to without my having to bother to tell them.
2) Others should be able to find out about me when they feel they need to.
3) I want to know what others are up to without having to bother them by asking.
4) I would prefer to share about myself with everyone in case anyone wants to know.
5) Rather than wait for them to tell me, I would like a way to find out about others whenever I need.
6) It would be useful to me if others shared about themselves to everyone in case anyone wants to know.

Boundary Preservation

This measure had eight questions on a 5-point Likert scale. A high score on this construct means that the user feels strongly that using apps on mobile devices will harm their relationship with others.

Items:

1) Using apps on my phone will improve my relationships with others.
2) The apps on my phone expose information that will negatively affect my relationship with others.
3) My phone apps support new behaviors that will improve my relationships.
4) I am worried others will use apps in a way that is out of line with our relationship.

5) Using apps on my phone enhances my relationships with others by keeping us better informed.
6) I am concerned that using the apps on my phone will trigger changes in behavior that hurt my relationships.
7) I feel others will use apps in a way that pushes our relationship in a positive direction.
8) It is likely that using apps on my phone will negatively impact my relationships with others.

Power Usage

This measure had twelve questions on a 5-point Likert scale. A high score indicates that the user is a power user i.e. adept at using technology to its fullest potential.

Items:

1) I think most technological gadgets are complicated to use.
2) I make good use of most of the features available in any technological device.
3) I have to have the latest available upgrades of technological devices that I use.
4) Use of information technology has almost replaced my use of paper.
5) I love exploring all the features that any technological gadget has to offer.
6) I often find myself using many technological devices simultaneously.
7) I prefer to ask friends how to use any new technological gadget instead of trying to figure it out myself.
8) Using any technological device comes easy to me.
9) I feel like information technology is a part of my daily life.
10) Using information technology gives me greater control over my work environment.

11) Using information technology makes it easier to do my work.

12) I would feel lost without information technology.

Number of Installed Apps

This is a behavioral scraped variable. This was collected by a background process in the app as the users went through the survey.

Total dangerous permissions granted

Similar to number of installed apps, this is another behavioral scraped variable. This was collected by a background process in the app as the users went through the survey. The collection of this variable did not require any special user permissions.

Location Ratio

This is a behavioral scraped variable. Data for this was collected by background process while the users were taking the first survey. LOCATION_RATIO is simply the total number of apps with location permission (ACCESS_FINE_LOCATION or ACCESS_COARSE_LOCATION) divided by the total number of apps on the device.

Location Comfort (percentage)

This is a *percieved in context of past behavior* variable. As stated before, the variable was generated by dividing the number of apps to which the user was comfortable giving location, by the total number of apps that had the location permission. A background process recorded the total number of apps to which the user had granted location permission. An app was considered

having the location permission if it had been granted either of the ACCESS_FINE_LOCATION or ACCESS_COARSE_LOCATION Android system permissions. Having a list of all the apps with location, we presented the users with a list asking them to select 'Yes' or 'No' for each of the apps to indicate if they were comfortable with those apps having the location permission.

Location revoke (percentage)

This is a *percieved in context of past behavior* variable similar to location comfort (percentage). After selecting 'Yes/No' to indicate location comfort, we presented the user with a screen with all the apps for which they had selected 'No' and asked them if they would revoke the location permission for that app. We then obtained this measure by dividing the total number of apps where the users would revoke the location permission by the total number of apps for which location was granted.

Data Collection and Recruitment

We obtained IRB approval to recruit one hundred and fourteen participants using Amazon Mechanical Turk [46]. Amazon Mturk allows us to offer payments for each 'human intelligence task' or HIT. To standardize our collected data, we only accepted users if they were from the US, above 18 years of age, and users of Android devices. Furthermore, we restricted our HITs to workers who had a HIT approval rate of greater than 95% and had at least 50 approved HITs. We employed an additional attention screen question in our Qualtrics survey (e.g., "Please select 'Disagree Somewhat' for this question."). We discarded the data for users who did not pass this attention screen.

Study Procedure

We recruited participants through Amazon MTurk. Upon accepting the HIIT, the user was directed to a web-based form of informed consent. If they agreed to participate in the study, we generated a 'Consent ID' and redirected the user to a Google Play link to install our custom-made application. We named the application 'UCF Permissions Study'. After successfully installing the app, the user was asked to input the unique 'Consent ID,' which served to verify if the user completed the consent form prior to downloading the app. At this stage, a unique user identifier (uuid) was generated, which tracked the user throughout the study. If the user fails to complete the consent then they were not given the app download link and were not paid for study completion.

Next, the user was presented a Qualtrics survey (with perceived measures) through the app, which was rendered via Android's web view. While the user was answering the survey, we ran a process in the background to collect application permission data from the device. (Note: The user was explicitly informed of this collection in the consent form before the study.) The application permission data included a list of all the apps on the user's device and the corresponding permissions granted to those apps. In order to minimize the number of permissions and to keep our data relevant, only the status of *Dangerous Android Permissions* which, according to Google, "cover areas where the app wants data or resources that involve the user's private information, or could potentially affect the user's stored data or the operation of other apps" [45] was recorded. The data record for the app contained a list of all the dangerous permissions granted to the app, and if the specific permission was present in the app's manifest and never asked, or if it was asked but denied. After the user completed the first survey, they were

presented a screen (Figure 1) with the following prompt: *"The following apps are installed on your phone. Please select whether or not each app should have access to your location."*

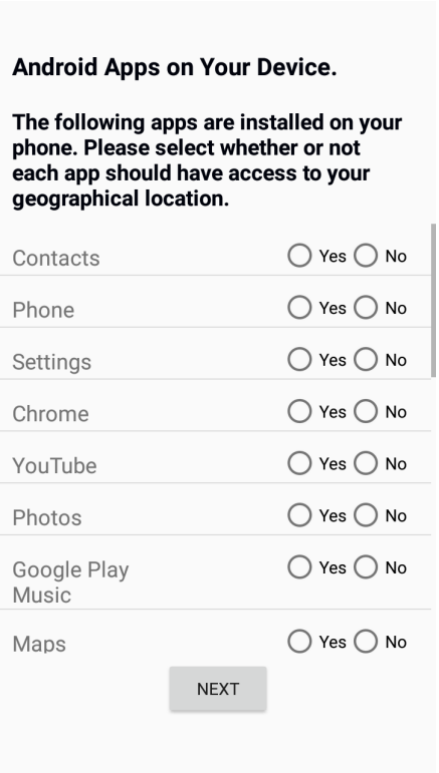

Figure 1: Location comfort question

Following the prompt, a list of apps which had the 'FINE_LOCATION' or 'COARSE_LOCATION' permission in their manifest were shown along with two radio buttons labeled 'Yes' and 'No' respectively. We did not distinguish between "FINE_LOCATION" and "COARSE_LOCATION" because the user experience of granting these permissions is exactly the same (the user does not know if the location given is fine or coarse). Having the permission in the manifest does not necessarily mean that the user has allowed the app to access their location as the app may not have explicitly requested this permission. We collect data for our

'LOCATION_COMFORT_PERCENTAGE' *perceived in context of past behavior* variable through this screen.

Next, users were presented with the following screen (Figure 2) and prompt: "*The apps below have access to your location, even though you stated that they should not. Please select the apps below for which you plan to revoke access to your location after completing this study."*

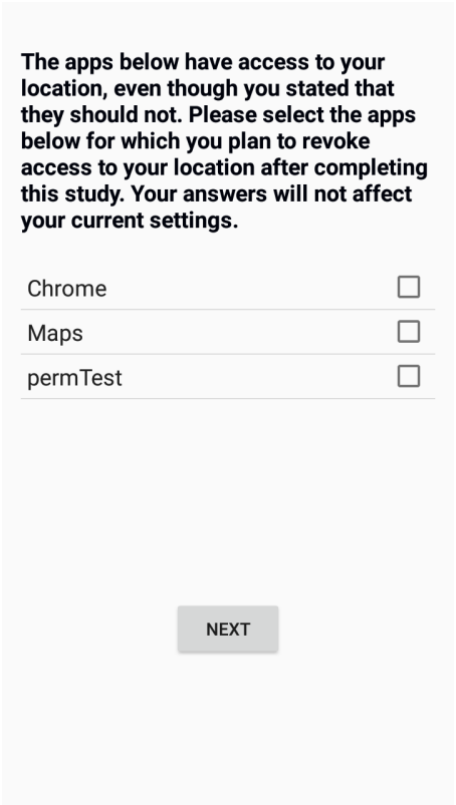

Figure 2: Revoke location question

Following the prompt, a list of app names with a checkbox is displayed. The apps are displayed if the following two conditions are satisfied: 1) The user explicitly granted the "COARSE_LOCATION" or the "FINE_LOCATION" permission to the app, and 2) The user indicated that they would be not comfortable with this app having their location in the previous

screen. The checkbox allows the user to indicate if they want to revoke the location permission to this application.

Next, the users were presented with a screen (Figure 3), which had the inverse prompt: "*The apps below do not have access to your location, even though you stated that they should. Please select the apps below for which you plan to allow access to your location after completing this study. Your answers will not affect your current settings.*"

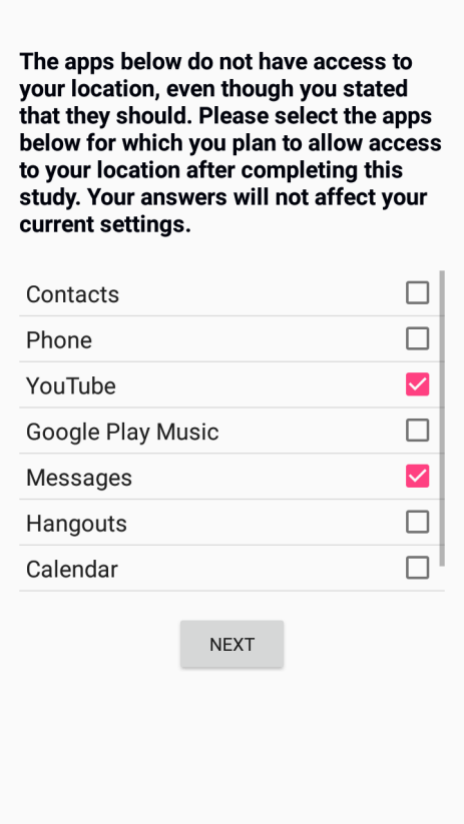

Figure 3: Allow location question

Following the prompt, a list of app names with a checkbox is displayed. The apps are displayed if the following two conditions are satisfied: 1) The user did not grant the "COARSE_LOCATION" or the "FINE_LOCATION" permission to the app or the app never asked for it, and 2) The user indicated that they would be comfortable with this app having their

location. The purpose of these two screens was to gauge user intention based on their initial attention and to understand how the user's reaction changes when they are explicitly told relevant information. The data needed for our REVOKE_LOCATION_PERCENTAGE *perceived in context of past behavior* variable is collected here.

Next, participants were presented a screen (Figure 4) that explained that they would be asked to grant the Location permission to the app. We explicitly told them that they could "Allow" or "Deny" this permission request.

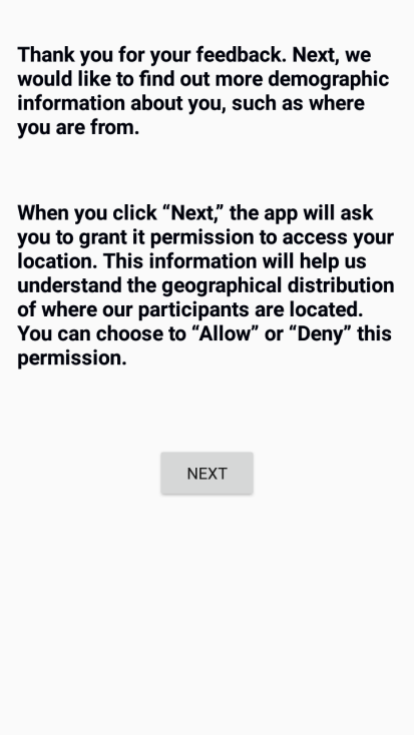

Figure 4: Pre location request

Next, we used the native permission dialog to ask users to grant (or deny) location sharing with the app. This measure represented our dependent variable and gave us vital data on the actual in app behavior of the user (LOCATION_GIVEN variable) allowing us to correlate it with data

collected from the surveys. When the user clicks next, they are shown the standard Android location prompt (Figure 5) asking if they would allow or deny the location request:

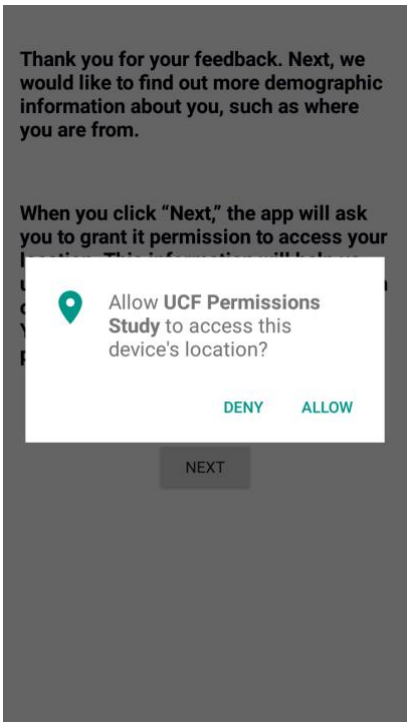

Figure 5: Location request screen

In a second Qualtrics survey, participants were asked to provide some basic demographic information. After successful completion of the survey, the user is given the completion code (Figure 6), which they can enter in Amazon Mechanical Turk to get paid.

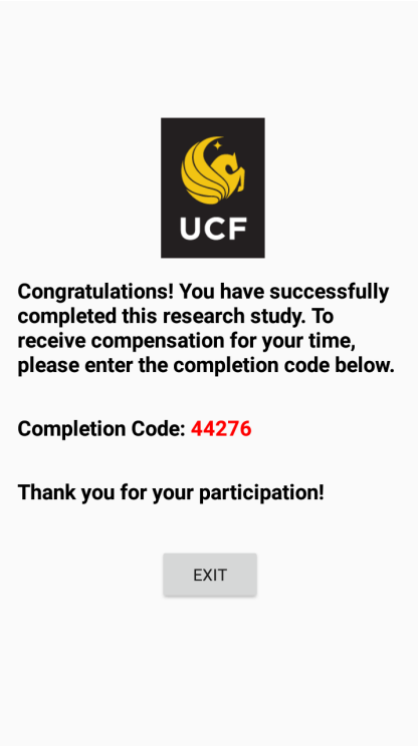

Figure 6: Completion screen

System Implementation

As noted before, the app was designed for the Android operating system using Android Studio. We used Qualtrics as our survey platform and displayed the survey using Android's web-view. On the backend, we used a Node.js API to serve requests and a MySQL database to store user data. All connections were through a secure HTTPS channel.

Data Analysis Approach

To prepare our data for analysis, we created composite variables (i.e., averaged across multiple items) for our constructs and calculated construct validity. We used a cutoff of 0.6 (Cronbach's alpha value) as a measure of reliability. According to the literature, a Cronbach's alpha value of 0.6-0.7 indicates adequate reliability [47]. The descriptive statistics for our data are shown in Table 1. Note that the Cronbach alpha value measures the internal consistency of an individual measure [48]. After preparing the data, we carried out a logistic regression analysis with different combinations of variables to find the best model for predicting LOCATION_GIVEN. We used logistic regression as we had a dichotomous dependent variable [24].

Table 1: Descriptive statistics of variables

Measure Name	Type	Mean	Median	SD	Skewness	Kurtosis	Cronbach's Alpha
Behavioral Intent	Perceived	4.09	4.33	0.786	-1.23	1.68	0.81
Perceived Surveillance	Perceived	4.16	4.33	0.711	-0.829	0.163	0.64
Perceived Intrusion	Perceived	3.83	4.00	0.999	-0.968	0.574	0.94
Boundary Preservation	Perceived	3.59	3.63	0.612	-0.645	1.841	0.72
Power Use	Perceived	4.23	4.25	0.524	-0.69	0.79	0.82
Secondary Use Of Personal Information	Perceived	4.13	4.00	0.967	-1.36	1.77	0.93
FYI About Myself	Perceived	2.59	2.58	0.990	0.02	-0.78	0.89
NUM_INSTALLED_APPS	Scraped	93.91	82.00	43.2	1.143	1.170	N/A
TOTAL_DANGEROUS_ PERMISSIONS_GRANTED	Scraped	230.44	218.50	91.5	0.87	1.185	N/A
LOCATION_RATIO	Scraped	26.51	25.52	8.59	0.294	-0.386	N/A
LOCATION_COMFORT _PERCENTAGE	Perceived In Context Of Past Behavior	33.50	25.00	27.9	0.784	-0.496	N/A
REVOKE_LOCATION _PERCENTAGE	Perceived In Context Of Past Behavior	26.88	13.27	30.7	1.073	-0.17	N/A

CHAPTER THREE: RESULTS

Descriptive Statistics

We recruited a total of 114 US based participants. 98% of the participants reported being in the US for longer than 5 years. All participants were above the age of 18. 55% of our participants were male and 45% were female. 40% of the participants identified their locality as urban, 42% as suburban and 18% as Rural. 58% of the participants had at least a 4-year college degree or more. A majority (82%) of our participants identified has White Caucasian. 58% were employed full time. Our participants had a diverse of occupations from Software Engineer to Food Management to Stay-at-home Mom and had a median income of $40,000 - $60,000.

In terms of usage statistics, participants had an average of 94 apps installed on their devices with a minimum of 29, maximum of 239, and standard deviation of 43.The top five most common non-system apps (applications which are not typically pre-installed on devices) were social media apps: 1) Instagram (43%), 2) Facebook (41%), 3) Snapchat (24%), 4) Hangouts (21%), and 5) Twitter (21%). On average participants granted 230 dangerous permissions to the various apps on their devices. 73% of participants granted us their location. Specifically, users had an average of 25 apps with access to their location (ACCESS_FINE_LOCATION or ACCESS_COARSE_LOCATION). Yet, on average, people were only comfortable with 33% of the location utilizing apps actually having their location. However, when asked if they would revoke the location for apps for which they were uncomfortable giving location to, only 27% said that they would do so.

Simple Effects Analysis

We started our data analysis by running a 1-way ANOVA with 'LOCATION_GIVEN' as the grouping variable (0 = Did not grant location, 1 = Granted location). We did this to examine the simple effects of each of our independent variables on our dependent variable irrespective of one another. We also used the results from this simple effect analysis to guide the choice of variables for our logistic regression models.

Table 2: Simple Effects Analysis

Variable	F Score	Significance
Behavioral Intent	2.49	0.117
Power Usage	0.397	0.530
Boundary Preservation	3.84	0.521
FYI About Myself	**6.768**	**0.011**
Perceived Surveillance	**6.807**	**0.010**
Perceived Intrusion	3.087	0.082
Secondary Use of Personal Information	**3.989**	**0.048**
NUM_INSTALLED_APPS	1.137	0.289
TOTAL_DANGEROUS_PERMISSIONS_GRANTED	1.498	0.224
LOCATION_RATIO	0.621	0.432
LOCATION_COMFORT_PERCENTAGE	**9.626**	**0.002**
REVOKE_LOCATION_PERCENTAGE	**9.195**	**0.003**

Interestingly, behavioral intent was not a significant predictor of actual privacy behaviors. We can see from Table 3 that users who shared their location scored higher on FYI About Myself, confirming Page et al's [22] finding that people with the FYI About Myself personality type are more likely to share their location. Users who shared location scored higher on Perceived Survellience. Given that higher percieved survellience measures users' assessment of excessive data collection, this result is no surprise. Similarly concern for secondary use of personal information was much lower for people who shared location. Secondary use measures the user's concern about information being used for purposes other than what it was collected for, and this reflects in our results. Users who were more comfortable with their past location sharing decisions had a higher location comfort percentage score and were more likely to share their location. Users who were concerned about their past location sharing decisions but still did not want to revoke the location permission had a lower revoke location percentage score and were more likely to share their location.

Table 3: Comparison of means

Variable	**LOCATION_GIVEN = 0**	**LOCATION_GIVEN = 1**
FYI About Myself	2.20	2.70
Perceived Survellience	4.44	4.05
Secondary Use Of Personal Information	4.43	4.02
LOCATION_COMFORT_PERCENTAGE	20.43	38.16
REVOKE_LOCATION_PERCENTAGE	40.98	21.84

Logistic Regression Results

RQ1: Using perceived measures to predict LOCATION_GIVEN

The ANOVA suggests that perceived surveillance and FYI About Myself have a significant effect on LOCATION_GIVEN. We found that these two were indeed the most appropriate choices for a logistic regression model.

Table 4: Perceived Constructs Only (Nagelkerke R squared = 13.9%)

Variable	**B**	**Sig**	**Exp(B)**
FYI About Myself	0.482	0.049	1.619
Perceived Surveillance	-0.787	0.043	0.455
Constant	3.209	0.088	24.750

RQ1 asked if perceived measures can predict actual user location sharing behavior. This model suggests that while there FYI About Myself and Perceived Surveillance better than the rest, the overall predictive power is low as the R-squared value is only 13.9%.

RQ2: Using behavioral scraped data to predict LOCATION_GIVEN

The ANOVA analysis showed that behavioral scraped variables (INSTALLED_APPS, TOTAL_DANGEROUS_PERMISSIONS_GRANTED and LOCATION_RATIO) did not affect

LOCATION_GIVEN significantly. Running a logistic regression with these variables gives a similar result

Table 5: Behavioral Scraped Data (Nagelkerke R squared = 2.1%)

Variable	**B**	**Sig**	**Exp(B)**
NUM_INSTALLED_APPS	0.005	0.606	1.005
TOTAL_DANGEROUS_PERMISSIONS_GRANTED	0.000	0.972	1.000
LOCATION_RATIO	0.020	0.620	1.020
Constant	0.313	0.600	1.368

Surprisingly, actual behavioral scraped data proved to be much worse than perceived measures in predicting actual privacy behavior with the model explaining only 2.1% of the variance in LOCATION_GIVEN.

RQ3: Using perceived in context of past behavior variables to predict LOCATION_GIVEN

The ANOVA results indicate that perceived in context of past behavior variables (LOCATION_COMFORT_PERCENTAGE and REVOKE_LOCATION_PERCENTAGE) significantly impact LOCATION_GIVEN. Our logistic regression model confirms this:

Table 6: Percieved In Context of Past Behavior Only (Nagelkerke R squared = 16%)

Variable	B	Sig	Exp (B)
LOCATION_COMFORT_PERCENTAGE	**0.023**	**0.036**	**1.024**
REVOKE_LOCATION_PERCENTAGE	-0.012	0.094	0.988
Constant	0.742	0.119	2.101

This model does slightly better than perceived variables alone and explains 16% of the variance (**RQ3)**. Even though both the variables were significant in ANOVA analysis, in this model LOCATION_COMFORT_PERCENTAGE appears to explain more of the variance.

RQ4: Combined Model

In order to test our final research question, we decided to combine all of our variables to get the best model for 'LOCATION_GIVEN'.

We found that adding the *percieved in context of past behavior* variable LOCATION_COMFORT_PERCENTAGE to the perceived measures helps explain a lot more variance. This was the best logistic regression model that we could find to predict LOCATION_GIVEN:

Table 7: Combined Model (Nagelkerke R squared = 21.3%)

Variable	B	Sig	Exp(B)
LOCATION_COMFORT_PERCENTAGE	**0.025**	**0.019**	**1.025**
FYI About Myself	**0.489**	**0.051**	**1.631**
Perceived Surveillance	-0.583	0.150	0.558
Constant	1.615	0.421	5.025

RQ4 asked if past user behavior or perceived in context variables augment purely perceived measures in predicting user behavior. According to this model, *perceived in context of past behavior* variables can help strengthen prediction models for LOCATION_GIVEN. For a one point of increase in LOCATION_COMFORT_PERCENTAGE, the probability of sharing location increases by a factor of 1.025. For every one-point increase in FYI About Myself, the probability of sharing location increases by a factor of 1.631. The model which only used perceived variables had an R squared of 13.9%, where as this one improves it to 21.3%. A significant change in predictive power.

We noted before that when calculating location comfort percentage we listed which had either of the "COARSE_LOCATION" or "FINE_LOCATION" permission. This was because the permission request dialogue is the same when asking for either of these permissions. To the user the dialogue appears as shown in figure 5 for both "COARSE_LOCATION" and "FINE_LOCATION". We also noted that calculating location comfort for apps which only had the "ACCESS_FINE_LOCATION" permission made a negligible difference to our final model:

Summary of Models

We summarize the models from worst to best in the table below:

Table 8: Summary Of Models

Model	**Variables**	**Nagelkerke R squared**
Behavioral Scraped Data Only	LOCATION_RATIO, NUM_INSTALLED_APPS, TOTAL_DANGEROUS_PERMISSIONS_GRANTED	2.1%
Percieved Measures Only	Percieved Survellience, FYI About Myself	13.9%
Perceived In Context of Past Behavior Only	LOCATION_COMFORT_PERCENTAGE, REVOKE_LOCATION_PERCENTAGE	16%
Combined Model (Percieved + Perceived In Context of Past Behavior)	FYI About Myself, Perceived Surveillence, LOCATION_COMFORT_PERCENTAGE	21.3%

We found that adding the location comfort percentage to perceived measures only model improved the model's predictive power. The significance of this difference was confirmed using

a likelihood ratio test which gave a significant p value of 0.01 which is less than the threshold of 0.05.

CHAPTER FOUR: DISCUSSION AND CONCLUSION

Summary of Main Findings

In this study, we asked the following four research questions:

1) Do perceived measures predict actual user location sharing behavior?
2) Does behavioral scraped data predict actual user location sharing behavior?
3) Can perceived in context variables predict actual user location sharing behavior?
4) Can past user behavior or perceived in context variables augment purely perceived measures in predicting user location sharing behavior?

For RQ1, we found that perceived measures were a weak predictor of actual user privacy behavior (LOCATION_GIVEN) in our case. Our model showed that FYI About Myself and Perceived Surveillance were the most important factors in predicting if the user will share their location. For RQ2, we found that behavioral scraped data proved to be even worse at predicting user privacy behavior than perceived measures. No significant predictors were found in our model and R squared value was very low. For RQ3, perceived in context of past behavior variables taken alone did a better job than purely perceived variables and behavioral data. LOCATION_COMFORT_PERCENTAGE was found to be a significant predictor variable for predicting LOCATION_GIVEN. Finally, the combined model of perceived measures and *perceived in context of past behavior* measures was the best predictor model for LOCATION_GIVEN. FYI About Myself and LOCATION_COMFORT_PERCENTAGE were found to be the best predictors in this model (RQ4). This strongly suggests that *perceived in*

context of past behavior variables might be able to augment traditional surveys and scraped behavioral data to produce stronger predictive models or user behavior.

Implications for Mobile Privacy Research

Similar to Xu et al. [36] and Page et al. [22], we found that the perception of surveillance, as well as the FYI About Myself personality type influences user's location sharing behavior. This can help researchers narrow down on constructs and devise shorter and more streamlined surveys when trying to predict if the user will share their location. Directly scraped behavioral data, including as the number of installed apps, the number of dangerous permissions granted, and the ratio of apps having location to total apps did not predict the user's location sharing behavior. This finding suggests that while some phone metadata i.e., Ghosh et al's study [11] which found that device metadata such as call duration and ignored calls can predict user concerns as stated on a survey, it might not be able to predict actual behavior in the context of mobile location sharing. Overall, the results of this study suggest that combining social science approaches with computer science can yield stronger models. In our literature review, most of the mobile privacy research that we encountered was strictly divided between social sciences [34, 7, 19, 11, 15, 16, 14, 21] or using computational methods [4, 13, 12, 3, 17, 25, 22,2, 8], with few studies on the intersection of the two disciplines [9, 35]. Our work suggests that privacy research should try to identify the relevant *perceived in context of past behavior* variables for the phenomenon or privacy behavior they are trying to study. In our case, we identified LOCATION_COMFORT_PERCENTAGE as the *percieved in context of past behavior* variable which is strongly correlated to user location sharing decisions. This suggests that privacy

behavior of the past is not very important in predicting future behavior but instead how the user felt later on about their past behavior is more important.

Implications for Design

Mobile apps have become central to our digital lives and are often the conduits for data leaks and irresponsible data collection. Our work shows that user perceptions of their past behavior impacts their future behavior more than the past behavior itself i.e users' LOCATION_COMFORT was better at predicting whether the user would share their location vs their actual history of sharing location or granting dangerous permissions. This suggests that app makers should design apps to be "context aware". Apps which require privacy choices should periodically ask users to reflect about how they feel regarding their past privacy choices inside the app in order to understand the user's comfort level and help them make more informed decisions about their privacy in the future. An example of taking this approach would be if an app does not need location for essential purposes, it can measure the user's location comfort and then disable/enable location by default for itself. In this way the user's expectation of privacy can be preserved.

Limitations and Future Research

The primary limitation of this study is that its participants were U.S. only. User privacy decisions and experiences can be shaped by their cultural environment and limiting the users only to the U.S. limits the ability to generalize these results. We also limited our participants to age 18 and above to simplify the IRB process. Future research in this area should consider expanding the sample to multiple countries and ages and perhaps try to identify more culture specific *perceived*

in context of past behavior variables. Users recruited from Amazon Mechanical Turk are likely to be more tech savvy than the general population as general everyday users are not expected to know about Mechanical Turk's existence. This may mean that the study might not apply as well to a more tech-illiterate population. Furthermore, the relatively small size of our data (114 participants) means that further research should be done with a much larger data set (1000 plus users) in order increase the power of this work. An interesting direction ahead could be to measure how LOCATION_COMFORT_PERCENTAGE changes for specific apps for users across their lifetime (total time for which they are installed on the phone). This could help app makers identify user behavior trends and take corrective action.

Conclusion

The rapid growth of portable communications devices has meant that the boundaries of privacy have been tested in new ways. It is therefore crucial to understand and be able to predict user behavior in order to design experiences which respect the user's expectation of privacy and do not harm their well-being. This study contributes to the field of mobile privacy by shedding light on the kinds of percieved measures that can explain user location sharing behavior. We show that behavioral scraped data might not be the best indicator of future user behavior. Finally, we show that augmenting percieved measures with *percieved in context of past behavior* variables can help strengthen prediction models.

APPENDIX: IRB APPROVAL FOR HUMAN RESEARCH

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Pamela Wisniewski** and Co-PI **Muhammad Irtaza Safi**

Date: **July 02, 2018**

Dear Researcher:

On 07/02/2018 the IRB approved the following human participant research until 07/01/2019 inclusive:

Type of Review:	IRB Continuing Review Application Form Expedited Review Category #7; This approval includes a Waiver of Written Documentation of Consent
Project Title:	Examining Users' application permissions on Android Mobile Devices
Investigator:	Pamela Wisniewski
IRB Number:	SBE-17-13076
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 07/01/2019, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

This letter is signed by:

Signature applied by Renea C Carver on 07/02/2018 09:31:31 AM EDT

Designated Reviewer

LIST OF REFERENCES

1. Icek Ajzen. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50, 2: 179–211. https://doi.org/10.1016/0749-5978(91)90020-T
2. Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. 2015. 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), 787–796. https://doi.org/10.1145/2702123.2702210
3. B. Amos, H. Turner, and J. White. 2013. Applying machine learning classifiers to dynamic Android malware detection at scale. In *2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 1666–1671. https://doi.org/10.1109/IWCMC.2013.6583806
4. Steven Arzt, Siegfried Rasthofer, Christian Fritz, Eric Bodden, Alexandre Bartel, Jacques Klein, Yves Le Traon, Damien Octeau, and Patrick McDaniel. 2014. FlowDroid: Precise Context, Flow, Field, Object-sensitive and Lifecycle-aware Taint Analysis for Android Apps. In *Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation* (PLDI '14), 259–269. https://doi.org/10.1145/2594291.2594299
5. Louise Barkuus and Anind Dey. 2003. Location-Based Services for Mobile Telephony: a Study of Users' Privacy Concerns. In *In 9th Ifip Tc13 International Conference on Human-Computer Interaction, Interact 2003*.
6. Susanne Barth and Menno D. T. de Jong. 2017. The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic

literature review. *Telematics and Informatics* 34, 7: 1038–1058. https://doi.org/10.1016/j.tele.2017.04.013

7. David Bisson. The 10 Biggest Data Breaches of 2018... So Far. Retrieved October 7, 2018 from https://blog.barkly.com/biggest-data-breaches-2018-so-far
8. Lee J. Cronbach and Paul E. Meehl. 1955. Construct validity in psychological tests. *Psychological Bulletin* 52, 4: 281–302. https://doi.org/10.1037/h0040957
9. Tobias Dehling, Fangjian Gao, Stephan Schneider, and Ali Sunyaev. 2015. Exploring the Far Side of Mobile Health: Information Security and Privacy of Mobile Health Apps on iOS and Android. *JMIR mHealth and uHealth* 3, 1. https://doi.org/10.2196/mhealth.3672
10. Kassem Fawaz and Kang G. Shin. 2014. Location Privacy Protection for Smartphone Users. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (CCS '14), 239–250. https://doi.org/10.1145/2660267.2660270
11. Isha Ghosh and Vivek K. Singh. 2016. Predicting Privacy Attitudes Using Phone Metadata. In *Social, Cultural, and Behavioral Modeling* (Lecture Notes in Computer Science), 51–60.
12. Shion Guha and Stephen B. Wicker. 2015. Spatial Subterfuge: An Experience Sampling Study to Predict Deceptive Location Disclosures. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '15), 1131–1135. https://doi.org/10.1145/2750858.2804281
13. Cecilia Kang and Sheera Frenkel. 2018. Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users. *The New York Times*. Retrieved October 7, 2018 from https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html
14. Hyunjin Kang and Wonsun Shin. 2016. Do Smartphone Power Users Protect Mobile Privacy Better than Nonpower Users? Exploring Power Usage as a Factor in Mobile Privacy

Protection and Disclosure. *Cyberpsychology, Behavior, and Social Networking* 19, 3: 179–185. https://doi.org/10.1089/cyber.2015.0340

15. M. J. Keith, J. S. Babb, and P. B. Lowry. 2014. A Longitudinal Study of Information Privacy on Mobile Devices. In *2014 47th Hawaii International Conference on System Sciences*, 3149–3158. https://doi.org/10.1109/HICSS.2014.391
16. Ian Kerr and Jessica Earle. 2013. Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy. *Stanford Law Review Online* 66: 65.
17. Li Li, Alexandre Bartel, Tegawendé F. Bissyandé, Jacques Klein, Yves Le Traon, Steven Arzt, Siegfried Rasthofer, Eric Bodden, Damien Octeau, and Patrick McDaniel. 2015. IccTA: Detecting Inter-component Privacy Leaks in Android Apps. In *Proceedings of the 37th International Conference on Software Engineering - Volume 1* (ICSE '15), 280–291. Retrieved October 8, 2018 from http://dl.acm.org/citation.cfm?id=2818754.2818791
18. Yuanchun Li, Yao Guo, and Xiangqun Chen. 2016. PERUIM: Understanding Mobile Application Privacy with permission-UI Mapping. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp '16), 682–693. https://doi.org/10.1145/2971648.2971693
19. Jialiu Lin, Shahriyar Amini, Jason I. Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. 2012. 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), 501–510. https://doi.org/10.1145/2370216.2370290
20. Naresh K. Malhotra, Sung S. Kim, and James Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research* 15, 4: 336–355. https://doi.org/10.1287/isre.1040.0032

21. Marathe, S., Sundar, S.S., Nije Bijvank, M., van Vugt, H.C., Veldhuis, J., Communication Science, and Information Management & Software Engineering. 2007. Who are these power users anyway? Building a psychological profile. Retrieved October 7, 2018 from https://research.vu.nl/en/publications/b349ddc2-97fe-4a2a-85e9-e6e58b0f0478
22. Xinru Page, Bart Knijnenburg, and Alfred Kobsa. 2013. FYI: Communication style preferences underlie differences in location-sharing adoption and usage. In *UbiComp 2013 - Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 153–162. https://doi.org/10.1145/2493432.2493487
23. Xinru Page, Alfred Kobsa, and Bart P. Knijnenburg. 2012. Don't Disturb My Circles! Boundary Preservation Is at the Center of Location-Sharing Concerns. In *Sixth International AAAI Conference on Weblogs and Social Media*. Retrieved October 8, 2018 from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4679
24. Frank Schoonjans. Logistic regression. *MedCalc*. Retrieved October 28, 2018 from https://www.medcalc.org/manual/logistic_regression.php
25. Suranga Seneviratne, Aruna Seneviratne, Prasant Mohapatra, and Anirban Mahanti. 2014. Predicting User Traits from a Snapshot of Apps Installed on a Smartphone. *SIGMOBILE Mob. Comput. Commun. Rev.* 18, 2: 1–8. https://doi.org/10.1145/2636242.2636244
26. Shwadhin Sharma and Robert E. Crossler. 2014. Disclosing too much? Situational factors affecting information disclosure in social commerce environment. *Electronic Commerce Research and Applications* 13, 5: 305–319. https://doi.org/10.1016/j.elerap.2014.06.007
27. Fuming Shih, Ilaria Liccardi, and Daniel Weitzner. 2015. Privacy Tipping Points in Smartphones Privacy Preferences. In *Proceedings of the 33rd Annual ACM Conference on*

Human Factors in Computing Systems (CHI '15), 807–816. https://doi.org/10.1145/2702123.2702404

28. H. Jeff Smith, Sandra J. Milberg, and Sandra J. Burke. 1996. Information Privacy: Measuring Individuals' Concerns About Organizational Practices. *MIS Quarterly* 20: 167–196.
29. Daniel J. Solove. 2005. A Taxonomy of Privacy. *University of Pennsylvania Law Review* 154: 477.
30. S. Shyam Sundar and Sampada S. Marathe. 2010. Personalization versus Customization: the Importance of Agency, Privacy, and Power Usage. *Human Communication Research* 36, 3: 298–322. https://doi.org/10.1111/j.1468-2958.2010.01377.x
31. Haoyu Wang, Jason Hong, and Yao Guo. 2015. 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), 1107–1118. https://doi.org/10.1145/2750858.2805833
32. Heng Xu, Tamara Dinev, H. Smith, and Paul Hart. 2008. Examining the Formation of Individual's Privacy Concerns: Toward an Integrative View. *ICIS 2008 Proceedings*. Retrieved from https://aisel.aisnet.org/icis2008/6
33. Heng Xu and Hock-Hai Teo. 2004. Alleviating Consumers' Privacy Concerns in Location-Based Services: A Psychological Control Perspective. In *ICIS*.
34. 2017. The world's most valuable resource is no longer oil, but data. *The Economist*. Retrieved October 7, 2018 from https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

35. 2018. Global social media research summary 2018. *Smart Insights*. Retrieved October 6, 2018 from https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/
36. Measuring mobile users' concerns for information privacy | Request PDF. *ResearchGate*. Retrieved October 7, 2018 from https://www.researchgate.net/publication/290022632_Measuring_mobile_users'_concerns_for_information_privacy
37. Cell phone sales worldwide 2007-2017. *Statista*. Retrieved October 6, 2018 from https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/
38. Mobile app revenues 2015-2020 | Statistic. *Statista*. Retrieved October 6, 2018 from https://www.statista.com/statistics/269025/worldwide-mobile-app-revenue-forecast/
39. Facebook Reports Second Quarter 2018 Results. Retrieved October 6, 2018 from https://investor.fb.com/investor-news/press-release-details/2018/Facebook-Reports-Second-Quarter-2018-Results/default.aspx
40. All of Facebook's Ad Targeting Options (in One Epic Infographic) | WordStream. Retrieved October 7, 2018 from https://www.wordstream.com/blog/ws/2016/06/27/facebook-ad-targeting-options-infographic
41. Facebook - revenue and net income 2017. *Statista*. Retrieved October 6, 2018 from https://www.statista.com/statistics/277229/facebooks-annual-revenue-and-net-income/
42. New Survey Finds Deep Consumer Anxiety over Data Privacy and Security. *IBM News Room*. Retrieved October 7, 2018 from https://newsroom.ibm.com/2018-04-16-New-Survey-Finds-Deep-Consumer-Anxiety-over-Data-Privacy-and-Security

43. (PDF) Communication Privacy Management Theory: Exploring Coordination and Ownership Between Friends. *ResearchGate*. Retrieved October 26, 2018 from https://www.researchgate.net/publication/262830701_Communication_Privacy_Management_Theory_Exploring_Coordination_and_Ownership_Between_Friends
44. Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings | USENIX. Retrieved October 9, 2018 from https://www.usenix.org/conference/soups2014/proceedings/presentation/lin
45. Permissions overview. *Android Developers*. Retrieved September 17, 2018 from https://developer.android.com/guide/topics/permissions/overview
46. Amazon Mechanical Turk. Retrieved September 17, 2018 from https://www.mturk.com/
47. Spss Explained. Retrieved October 13, 2018 from https://dl.acm.org/citation.cfm?id=983788
48. Using and Interpreting Cronbach's Alpha | University of Virginia Library Research Data Services + Sciences. Retrieved October 30, 2018 from https://data.library.virginia.edu/using-and-interpreting-cronbachs-alpha/