The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/1468-4527.htm

Phones, privacy, and predictions

A study of phone logged data to predict privacy attitudes of individuals

Isha Ghosh and Vivek Singh
Department of Library and Information Science,
Rutgers The State University of New Jersey, New Brunswick, New Jersey, USA

Received 31 March 2018
Revised 30 June 2018
Accepted 16 August 2018

Abstract

Purpose – Mobile phones have become one of the most favored devices to maintain social connections as well as logging digital information about personal lives. The privacy of the metadata being generated in this process has been a topic of intense debate over the last few years, but most of the debate has been focused on stonewalling such data. At the same time, such metadata is already being used to automatically infer a user's preferences for commercial products, media, or political agencies. The purpose of this paper is to understand the predictive power of phone usage features on individual privacy attitudes.

Design/methodology/approach – The present study uses a mixed-method approach, involving analysis of mobile phone metadata, self-reported survey on privacy attitudes and semi-structured interviews. This paper analyzes the interconnections between user's social and behavioral data as obtained via their phone with their self-reported privacy attitudes and interprets them based on the semi-structured interviews.

Findings – The findings from the study suggest that an analysis of mobile phone metadata reveals vital clues to a person's privacy attitudes. This study finds that multiple phone signals have significant predictive power on an individual's privacy attitudes. The results motivate a newer direction of automatically inferring a user's privacy attitudes by leveraging their phone usage information.

Practical implications – An ability to automatically infer a user's privacy attitudes could allow users to utilize their own phone metadata to get automatic recommendations for privacy settings appropriate for them. This study offers information scientists, government agencies and mobile app developers, an understanding of user privacy needs, helping them create apps that take these traits into account.

Originality/value – The primary value of this paper lies in providing a better understanding of the predictive power of phone usage features on individual privacy attitudes.

Keywords Mobile phones, Behavior modelling, Lifelogging, Privacy attitudes, Quantified self

Paper type Research paper

Introduction

In today's world, mobile phones have become one of the most favored devices to communicate with friends, family, as well as logging digital information about personal lives. As the usage of smartphones continues to grow, the volume of metadata being actively and passively captured by these mobile phones is also continuously increasing. Mobile phones are now increasingly equipped with sensors and many individuals are using the correspondingly generated data to keep track of their daily activities and social connections (Swan, 2013). For instance, simple communication logs can already provide significant insight into an individual's social network and social behavior. A variety of lifelogging devices with sensing technologies have been created and their applications provide users with the opportunity to track their lives accurately and automatically (Joho *et al.*, 2016). The idea of "Quantified Self" starts with tracking daily activities such as location, mood, health factors, sleep patterns, photos, phone calls and so on. These sensors are often used as a form of personal knowledge management to allow the individual to better manage their knowledge processes and interaction, collaboration and knowledge exchanges with others (Razmerita *et al.*, 2009). While there are newer technologies available to capture and monetize such data, the technology to maintain personal privacy while using lifelogging devices is lagging (Gurrin *et al.*, 2014; Sadeh *et al.*, 2009; Chin *et al.*, 2012). For example, multiple recent studies have pointed out that young adults today are careful

Online Information Review
© Emerald Publishing Limited
1468-4527
DOI 10.1108/OIR-03-2018-0112

about online information disclosure on social media sites (Acquisti and Gross, 2006; Acquisti and Grossklags, 2004; Buchanan *et al.*, 2007). However, the fact that smartphones constantly receive and send out data signals which contain personal information, even when users are not connected to online services, often slips under the radar (Karlson *et al.*, 2009; Schlegel *et al.*, 2011). Data generated from smartphones can provide insights into how often and which members of the network people choose to communicate with, changes in interactional patterns, as well as behavioral traits (Balmaceda *et al.*, 2014). This automatic and persistent log being collected over an extended time has resulted in a need for new tools to help people become more aware of their individual lifelogged data, and to offer them the means to manage their privacy settings according to their unique circumstances and needs.

While enormous amounts of personal data are being captured from user's smartphones to personalize user requirements ranging from next song, to next product, to next search result, little scholarly work has been done to understand how this mobile data (particularly call logs) can be used to build personalized privacy settings for users. This study is among the first attempts to understand the interconnections between long-term phone use information and an individual's privacy attitudes.

Based on a ten-week long field study involving phone metadata collection via a mobile phone-based logging app, a survey on privacy attitudes, and data gathered from qualitative interviews, this study[1] reports that an analysis of phone metadata may reveal vital clues to a person's privacy attitudes. Specifically, a predictive model based on phone usage metadata significantly outperforms a comparable personality features-based model in predicting individual privacy attitudes. Further, the qualitative insights obtained via follow up semi-structured interviews provide a window into the user's thought processes when engaging in specific phone behavior (e.g. returning missed calls) and how that relates to their privacy attitudes.

Related work and research questions

For this study, the presentation of related work focuses on research projects that discuss information sharing within and outside communities and their impact on privacy attitudes.

Understanding privacy attitudes

There have been several attempts to define privacy. In a systematic discussion of the different notions of privacy, Introna and Pouloudi (1999) developed a framework that explored the interrelations of interests and values for various stakeholders where privacy concerns have risen. "The central idea around privacy is the desire to keep personal information out of the hands of others, along with the ability to make connections and build networks" (Introna and Pouloudi, 1999). In this context, concern for privacy is a subjective measure – one that varies from individual to individual based on that person's own perceptions and values. In other words, different people have different levels of concern about their own privacy (Introna and Pouloudi, 1999).

However, a concern for privacy does not translate into similar behavior. Existing research (Barnes, 2006; Acquisti and Grossklags, 2004; Acquisti and Gross, 2006) has explored the dichotomy between concerns about privacy and actual behaviors exhibited by individuals. The existence of trade-off behaviors demonstrated by individuals as they balance the costs and benefits associated with online interactions, termed as the "privacy calculus" is also well documented (Wisniewski *et al.*, 2015; Krasnova *et al.*, 2010). From these instances, it is clear that a dichotomy exists between what people think and how they act. While there is some research to address the paradoxical nature of self-reported attitudes compared to demonstrated behaviors (Barnes, 2006; Acquisti and Grossklags, 2004), there is also an expectation that there should be at least some link between privacy behavior and attitudes (Kraus, 1995). A meta-analysis to systematically evaluate the associations between online privacy concerns, privacy literacy, online service use and adoption of privacy protective measures finds that privacy concerns are weakly correlated with the information shared on social networking sites (SNS), though not with frequency of SNS use (Baruh *et al.*, 2017). However, this relationship has not yet been investigated in the context of mobile phone-based interactions patterns.

Mobile phones are becoming increasingly ubiquitous throughout large portions of the world. As interactions get increasingly mediated via mobile phones, individuals communicating with each other via these devices, form their own social network (Beale, 2005). The disclosure of information within these networks can cause concern over the privacy of information being shared. Another effect of the increasing usage of mobile phones is the increase in data generated about call volumes, calling patterns and user location. Recent work suggests that users leave small implicit clues in their everyday phone behavior, which can predict multiple aspects of their daily lives including mental health, financial well-being and social status (Canzian and Musolesi, 2015; Singh *et al.*, 2013; Wang *et al.*, 2014). Along with individual usage behavior, the interpersonal networks formed via these phone interactions and their connections to a wide spectrum of offline interactional behavior has also been studied by researchers (Candia *et al.*, 2008). For instance, the patterns of calling activity have been found to be indicative of tie-strengths between individuals (Onnela *et al.*, 2007), and the frequency and duration of phone calls have also been found indicative of an individuals' emotional attachment to various members in their social network (Miritello *et al.*, 2013). The role of mobile phone as a means of achieving connectivity and therefore, as a conduit for emotional attachment is also established in research (Vincent, 2006).

Mobile phones, therefore, generate information on individual as well as interpersonal attitudes and behaviors. While some researchers have investigated the objective notion of privacy as "the right to be left alone" (Warren and Brandeis, 1890), recent literature has argued for the conceptualization of privacy as a function of managing disclosure in order to achieve desired levels of social interaction (Krasnova *et al.*, 2010). Altman's (1975) original theorization of privacy posits that individuals try to control disclosure by attempting to create an optimal function that balances the benefits of information sharing against the costs associated with such disclosure. Altman (1975) argues that information sharing and disclosure is a natural human phenomenon; however, individuals who indulge in information sharing also use practical mechanisms, like limiting the duration of conversation based on trustworthiness of the individual or confirming the number of people listening to the conversation, to adjust disclosures in relation to privacy attitudes, and concerns. This implies that an optimal state of privacy requires individuals to estimate the audience and the effects of their disclosure and then apply practical mechanisms at their disposal to maintain those connections. In the context of phone conversations, this interaction between disclosure and boundary regulation is made evident from metadata (call duration, number of accepted/rejected calls, etc.) and provides a link between individual privacy attitudes and interactional patterns as captured by different communication devices.

As seen from the above literature review, privacy attitudes often dictate interpersonal interactional patterns. Various facets of these interactions have been successfully captured using phone logs. Hence, this motivates a question as to whether the patterns of interactions as observed via mobile phones could be predictors of individual privacy attitudes:

RQ1. Do long-term logs of phone usage provide clues to an individual's privacy attitudes?

In order to gain an understanding of individual privacy attitudes, it was important to identify an instrument that would allow this study to gain a holistic understanding of privacy attitudes. Recent studies (Wang *et al.*, 2014; Wisniewski *et al.*, 2015; Acquisti and Gross, 2006) have undertaken significant research to develop scales to measure privacy attitudes as well as behavior. A large portion of this research deals with online interactions on specific SNS (Acquisti and Gross, 2006) or e-commerce sites (Wang *et al.*, 2014).

The current study goes beyond purely online settings and attempts to gain a sense of privacy attitudes in both online and offline behaviors.

While we recognize that a large portion of human interactions are conducted online, real world human interactions cannot be ignored. In order to get a holistic understanding of privacy attitudes, it is important to understand how individuals share information in offline settings and how these encounters can be predictors of their privacy attitudes. Based on this premise, we acknowledge the recent efforts toward building newer measures of privacy (Xu *et al.*, 2012; Buchanan *et al.*, 2007; Acquisti and Gross, 2006), but decide to use Westin's index, which remains one of the most comprehensive approaches toward obtaining a well-rounded understanding of privacy attitudes exhibited by individuals.

Social signals

In order to gain an accurate understanding of individuals' attitudes on privacy, it is important to understand the factors governing information sharing in social settings and the importance individuals' assign to disclosure when building relationships. Altman's (1975) theory of self-disclosure states that individuals selectively control access to information about themselves by regulating their social interactions. According to this theory, people develop privacy rules based on a set of criteria; for instance, they are motivated to conceal or reveal information based on cultural norms, gender, context, the risk-benefit ratio and other factors (Petronio, 2012). Research has shown that perceived risk is key to individuals' disclosure decisions. Individuals consider the risk to the discloser, the receiver, the relationship between them, and/or third parties (e.g. one's family) when making disclosure decisions (Petronio, 2012).

Among friends, the sharing of personal information creates a bond between communication partners (Petronio, 2012). This regulation and management of private information as experienced by friends is best explained by Petronio's (2012) communication privacy management (CPM) theory. According to this theory, people develop privacy rules based on a set of criteria; for instance, they are motivated to conceal or reveal information based on cultural norms, gender, context, the risk-benefit ratio and other factors (Petronio, 2012). The benefits accrued from being part of a large social network have been well-studied in sociology. Research has also linked membership in social networks to greater access to opportunities often termed as social capital (Bourdieu, 1986). While there are many ways to conceptualize social capital, Bourdieu (1986) describes it as the ability of individuals to access resources within their network. An increase in social capital has been linked to better public health, lower crime rates and more efficient financial markets (Adler and Kwon, 2002). Therefore, there are clear advantages to gaining membership within a network. However, the creation and maintenance of relationships within these networks over online and offline spheres require information disclosure which may be contrary to individual privacy concerns.

One way individuals can manage their privacy needs is by constructing and coordinating boundaries around sharing of private information. This concept of drawing and redrawing boundaries is extended in the work of Derlega and Chaikin (1977), which states that individuals often draw on personal attitudes and goals when constructing rules of sharing personal and shared information. CPM recognizes that individuals assess the risk-benefit ratio when deciding to reveal or conceal private information (Petronio, 2012). According to CPM, knowing whom to tell or trust is a central "rule" of interpersonal disclosure. Petronio (2012) argues that these rules fundamentally guide information sharing and are adapted over time. What an individual knows about a space for disclosure, shapes the rules which govern information sharing going forward. The CPM theory has so far, been studied and applied to both physical and online interactions (Petronio, 2012; Wisniewski *et al.*, 2015), however, as more and more communications occur over mobile phones, the network formed via these interactions cannot be ignored. Information shared and disclosed over these networks has unique implications for the CPM theory. For instance, the CPM theory states that, knowledge of context dynamically causes change in the discloser's privacy strategy, thus identifying the important relationship between situational knowledge and privacy attitudes. In the context of cell phone interactions, the ability to have a specific knowledge of whom the information is being disclosed to, may encourage individuals to disclose more freely, as they provide both knowledge of context and knowledge of the disclosure's range. This affordance combines aspects of face to face interactions (knowledge of the information receiver) as well as online (electronically mediated and lacking visual cues) interactions. This study contributes to the CPM theory and the related concepts of boundary negotiation, co-ownership, and privacy markers to understand the motivations and thought processes governing disclosure norms over mobile phone interactions. In order to examine whether these concepts apply to interactions over the mobile phone, this study proposes the following research question:

RQ2. How do information sharing practices vary across online, mobile phone and physical interactions?

Automatically inferring privacy needs from usage data

Social network sites offer a variety of tools that allow an individual to set disclosure rules. Previous research has explored various methods for improving the understanding of complex privacy settings in SNSs (Watson *et al.*, 2015). A recent study describes a privacy wizard for SNSs that describes a particular user's privacy preferences based on a limited amount of user input (Fang and LeFevre, 2010). However, this research does not take into account the unique affordance of smartphones that gather, store, and process multi-modal data over an extended period of time. This data is available and searchable to users in the form of usage logs, however, the potential of using this logged data as a predictor of individual privacy attitudes has not yet been investigated. At the same time, managing privacy on one's phone has emerged as a major research challenge in recent years (Almuhimedi *et al.*, 2014; Xu *et al.*, 2012). Based on the understanding that individuals' privacy attitudes are a vital building block in identifying their preferred privacy settings (Walrave *et al.*, 2015), this study tests the value of an individuals' phone metadata as a predictor of their privacy attitudes. To this end this study proposes the following research question:

Q2

RQ3. Can a machine learning algorithm be used to predict privacy attitudes based on logged phone metadata?

To explore these research questions in greater detail, this study builds five hypotheses as explained in the next section.

Hypotheses generation

This research uses CPM theory as a framework to understand how people use mobile phone interactions to help maintain their overall privacy boundaries. Rather than create metaphorical boundaries like people do during everyday communication in face-to-face contexts, people can use phone settings, or "markers" in Petronio's (2012) terms, to help technologically enforce their boundaries. For example, a user can block another user's phone number or make softer changes like removing the other user from their contact list, hence causing the following calls from that person to show up as an "unknown number" and less likely to be picked up.

CPM theory predicts that individuals who have more privacy concerns will try to limit their interactions. A key element of this theory is the aspect of privacy ownership, which states that people believe they are the sole owners of their information and therefore have the right to grant or revoke access to others (Petronio, 2012). In the context of phone usage, we believe that making or receiving a call grants access to information about the individual and will therefore, have a relationship with individual privacy attitudes. Research examining the relationship between SNS usage (specifically Facebook) and privacy concerns has also found that as the frequency and intensity of Facebook usage increases the likelihood of updating privacy settings also increases (Stutzman and Kramer-Duffield, 2010). In the context of phone interactions, the number of calls made by individuals over the period of study and the time spent on these interactions can be used to determine their volume and depth of interaction. Based on previous research (Xu *et al.*, 2012; Young and Quan-Haase, 2009), this study attempts to understand the relationships between phone interactions and privacy attitudes. Analysis from this study is expected to show a negative relationship between number of calls and privacy attitudes:

H1. Higher call count is associated with a lower concern for privacy.

H2. Longer time spent on calls is associated with a lower concern for privacy.

As the number of interactions over phones increases, individuals performing these interactions form their own network and create strategies to perform information sharing and disclosure within this network. Navigating this environment has its own set of challenges that are different from managing privacy in face-to-face interactions. Phone conversations differ from face to face interactions in the lack of visual cues afforded by this medium, and therefore, users have to negotiate information boundaries while retaining network ties that govern these interactions.

As the number of mobile phone users' increase, more and more information will be shared over phone calls. Individuals receive a number of phone calls outside of their network or known "friends and families" in a given day. These calls could be from marketing agencies, credit card sellers or even phishing, where criminals try to gain sensitive personal information using fraudulent means. In such a scenario, it would appear that people who are highly concerned about their privacy would only respond to calls from within their network. Therefore, a higher call response rate (calling back missed calls) could indicate a lower concern for privacy. Similarly, a higher missed call rate could also indicate a lower concern for privacy:

H3. Higher call response rate is associated with a lower concern for privacy.

H4. Higher missed call rate is associated with a higher concern for privacy.

The creation and maintenance of relationships is one of the chief motivations for an individuals' use of SNS (Ellison, 2007; Acquisti and Gross, 2006). As individuals build larger networks, rules of disclosure are renegotiated and boundaries are coordinated with co-owners after an initial disclosure (Petronio, 2012). The structure of these online social networks allows for the same information to be shared between close friends, strangers or acquaintances (Acquisti and Gross, 2006). As newer contacts are included within an existing social network, there is more personal information shared within the network. Specifically, any effort made to expand one's social network requires one to inform others about herself and share certain aspects of one's life. Therefore, people who have high privacy concerns may not readily initiate newer contacts.

H5. Higher number of new contacts initiated is associated with a lower concern for privacy.

Method

The relationship between privacy attitudes and privacy behaviors is a complicated one. In the context of phone-based interactions, it was important to understand the "rules" or norms that govern information sharing and disclosure over phone interactions. It was also essential to gain an in-depth understanding of individuals' privacy attitudes and the effect of these attitudes on sharing personal as well as shared information. Hence, in order to better understand the relationship between phone metadata and privacy attitudes, this study uses a mixed-method approach combining phone log data and Likert scale style surveys with standardized open-ended interviews. In the current study we collect and analyze quantitative and qualitative data and use a triangulation strategy to add rigor, breadth and depth to this analysis (Denzin, 2012).

Given the complexities associated with privacy attitudes and their interconnections with different thought processes as well as conscious and unconscious phone usage patterns, this is an exploratory study focused on understanding the relationship between phone metadata and privacy attitudes exhibited by individuals. This work uses a field study in two stages to answer the research questions and the hypotheses proposed.

Stage 1

Participants were invited to the study site to read and sign the consent form and fill out an online survey. The survey consisted of Westin's (2003) Privacy Segmentation Index (PSI), demographic questions (e.g. gender, age), and personality traits (John and Srivastava, 1999) used later in the study as a comparison tool. While participants completed the survey, they were asked to install the study client (an app) on their mobile phones. This app collected call logs, SMS logs and location logs, over a ten-week study period. Phone metadata collected by the study client and the participant answers to the privacy attitude and behavior surveys were analyzed to test multiple hypotheses and build predictive models.

Stage 2

Participants who successfully completed ten weeks of the study were invited back to do a standardized open-ended interview to better understand their attitudes toward privacy. All participants answered identical standardized questions; however, follow-up questions were asked wherever a need for greater clarification arose (https://rutgers.qualtrics.com/SE/?SID=SV_23p12zLNwIH5Wp7). Asking open-ended questions allowed participants to fully express their viewpoints and experiences.

We summarize the approach taken to answer the identified research questions in Table I. *RQ1* is studied across the two stages of this study. A number of related hypotheses are

Research questions	Method	Instruments	Stage(s)
RQ1. Do long-term logs of phone usage provide clues to an individual's privacy attitudes?	1. Testing hypotheses connecting phone signals and a self-reported survey. 2. Interpreting associations found based on qualitative interviews.	1. Westin's Privacy Segmentation Index (PSI) (Westin, 2003) 2. Mobile phone logging app 3. Standardized open-ended interviews	1 and 2
RQ2. How do information sharing practices vary across online, mobile phone and physical interactions?	1. Qualitative analysis based on interviews.	1. Standardized open-ended interviews	2
RQ3. Can a machine learning algorithm be used to predict privacy attitudes based on logged phone metadata?	1. Machine Learning (J48 decision trees) algorithms using selected features 2. Interpreting the results based on qualitative interviews	1. Westin's Privacy Segmentation Index (PSI) (Westin, 2003) 2. Big 5 Personality Survey Data (John and Srivastava, 1999) 3. Mobile phone logging app 4. Standardized open-ended interviews	1 and 2

Table I. Variables used in the study

tested using phone-log data and self-reported survey data in stage 1. The interpretation of the results is facilitated by standardized open-ended interviews undertaken in stage 2 of the study. *RQ2* is studied based on the qualitative analysis of the empirical data obtained via standardized open-ended interviews in stage 2 (see https://rutgers.qualtrics.com/SE/?SID=SV_23p12zLNwIH5Wp7). Finally, *RQ3* is studied using the phone-log data and self-reported survey data collected in stage 1. The phone-log data are combined using machine learning (J48 decision tree) algorithm to create predictive models for self-reported privacy attitudes. The models based on phone data are specifically compared against a model created based on Big Five Personality traits of the users.

Survey instrument privacy attitudes

In order to gain an understanding of privacy attitudes displayed by individuals' Westin's (2003) PSI was adopted. This survey (Appendix 1) consisted of statements designed to measure levels of concern about personal information disclosed by individuals to companies or businesses and their concerns about whether their information was being protected or not. Each of these questions required responses to be made on a 4-point scale ranging from strongly disagree to strongly agree.

The following definitions, as suggested by Westin (2003), were used to classify individuals into one of three categories:

(1) privacy Fundamentalists are respondents who agreed (strongly or somewhat) with the first statement (Q.1) and disagreed (strongly or somewhat) with the second (Q.2) and third statements (Q.3).

(2) privacy Unconcerned are those respondents who disagreed with the first statement (Q.1) and agreed with the second (Q.2) and third statements (Q.3).

(3) privacy Pragmatists are all other respondents.

Personality traits: as a baseline to compare and interpret the predictive power of phone-based signals we also consider the personality traits information. This information was obtained based on a self-report survey (Big Five Inventory) defined by John and Srivastava (1999). Big Five Inventory consists of 44 multiple choice questions and is one of the most commonly used measures for personality traits.

Interviews

In stage 2 of this study, we invited participants who had completed Stage 1 back for qualitative interviews to understand their motivations and attitudes toward information disclosed over phone-based interactions (see https://rutgers.qualtrics.com/SE/?SID=SV_23p12zLNwIH5Wp7 for more details). During these interviews, we asked participants broad questions about the types of information they would be more likely to share (question 5). Based on participant responses, we asked for further details comparing online social networks, e-mail, phone numbers, and physical addresses. As these were open-ended interviews, we were able to frame additional questions based on participant responses. All interviews were recorded and transcribed by the researchers. These transcripts were then analyzed using selective coding as related to study variables as shown in Table II. The themes emerging from this analysis were used to understand the importance assigned to phone metadata and interpret results obtained from the predictive model.

Predictive modeling

We used a machine learning (J48 decision tree) algorithm to build predictive models for privacy attitudes based on phone data. The J48 decision tree algorithm estimates the target value (privacy attitudes) of a new sample based on various attribute values of the available

Table II. Variables used in the study

Variable name	Definition
Privacy concern (output variable)	Score as determined by responses to Westin's Privacy Segmentation Index. Scores for each question on a Likert Scale of 1–4 were added together to get a combined Privacy Concern Score
Call count	*n(Calls)* Total number of calls received or made by participants in the duration of the study
Call duration	∑(time spent on calls) Sum of time spent on all calls received or made in the duration of this stud
Call response rate	(Number of responded missed calls/Number of missed calls)×100 Where "responded missed calls" are those which were returned within 1 hour of the missed call
Missed call rate	(Number of missed calls/call count)×100 This is percentage of calls missed (not answered) by the participant
Number of new contacts in outgoing calls	This variable is defined as the number of calls made to new contacts, i.e. contacts that were not seen in the initial four weeks of the study but calls were initiated after the first four weeks

data (call response rate, missed call percentage and newer contacts). The results obtained based on phone use features are compared against those obtained by personality-based features for a comparison. The metrics used to quantify results are accuracy and Receiver Operating Characteristic (ROC – Area Under the Curve) (Chawla, 2005).

Participants

Participants for this study were recruited around a university in the Northeastern USA. Recruitment flyers were distributed in-person and via e-mail, and social media. All adults who owned a smart phone and were able to travel to the New Brunswick campus were invited to participate in the study. The participants were incentivized monetarily to participate in the study. A total of 59 participants completed the study. However, some of the participants did not complete all the surveys or did not include a specified identifier to link their inputs on different surveys and phone metadata. This resulted in a set of 53 participants with complete survey and phone log data. Of these 53 individuals, 32 (59 percent) were men and 18 (34 percent) were women (demographic data was unavailable for three participants). The majority of participants were undergraduates between the ages of 18–21 years. These 53 participants were also invited to follow up interview sessions out of whom 14 agreed to join.

Privacy of participant data

Privacy of user data was of utmost priority throughout this project. All data were secured and protected at standards applied to medical metadata. All data captured by the study clients was required to be no more detailed than those employed by typically installed apps like the Gmail and Instagram. The participants had the option to opt out of the studies at any time. The eventual goal is to design privacy apps that "learn" user attitudes and preferences over a short period of time (e.g. weeks) and can recommend privacy settings even after they stop receiving the data. Also, though outside the scope of the current work, the eventual privacy app coming out of this research project is intended to run under the Open source personal data source (OpenPDS) framework (De Montjoye *et al.*, 2014). OpenPDS keeps personal data in the cloud under the purview of an individual user rather than the third parties like Google or Facebook. All personnel involved in this study had undergone human subjects' training and the study was approved by the Institutional Review Board at the authors' home institution.

Results and discussion

A descriptive analysis of data from this study shows that while both males and females are concerned with data sharing, females tended to have a slightly higher concern for privacy with 60 percent females receiving a "very concerned" privacy score on Westin's Index while only 54 percent males received the same score. There were no significant variations in privacy attitudes across age or education. A visual representation of the demographic is presented in Figure 1.

A little over half (58 percent) the surveyed participants exhibited moderate (27 percent) to high (31 percent) concerns for privacy, with around 42 percent who had a low score or were "unconcerned" with sharing their data. This is very different from the results of the original survey conducted in 2003. Westin reported only 10 percent of the population was classified as "Unconcerned," with the majority of individuals displaying moderate (64 percent) to high concerns (26 percent) with the use or misuse of their personal information (Westin, 2003). A possible reason could be that the current sample is mostly single undergraduate students and is not representative of a larger population; however, with such a vast difference in results, it at the very least posits a question about the differences in information sharing attitudes of individuals in the early 2000s to the present day.

We also analyzed the effect of age, gender, and education on privacy scores and did not find any significant relationships. A similar test investigating the effect of personality traits on privacy scores also did not yield any significant relationships (Table AI).

Next, an analysis of the results obtained to answer the three research questions is presented:

RQ1. Do long-term logs of phone usage provide clues to an individual's privacy attitudes?

A study of these interconnections was undertaken by testing multiple hypotheses connecting an individual's privacy attitudes with phone-based signals as identified in the previous section. Each hypothesis was tested based on specific features identified as listed in Table II and the results obtained by undertaking Pearson's correlation analysis are shown in Table III.

As shown in Table III, *H3*–*H5* were found to be statistically significant (with Holm-Bonferroni correction) and *H1*, and *H2* were not significant. Though we found medium to large effect sizes, given the modest sample size ($n = 53$), we will be cautious in generalizing these results (Cohen, 1992). However, data gathered from qualitative interviews allows us to better interpret and gain an understanding of the underlying factors behind these relationships.

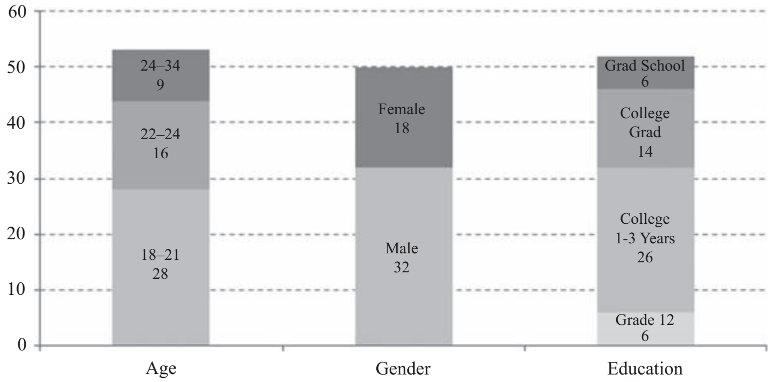

Figure 1.
Descriptive statistics of the sample population

Based on existing studies (Buchanan *et al.*, 2007; Acquisti and Gross, 2006) analyzing information sharing and disclosure, the hypothesis was that a higher call count and longer time spent on calls would be indicative of a low concern for privacy was proposed. However, the results obtained in this study showed that this was not the case. During qualitative interviews, conducted in stage 2, most participants confirmed that the information they disclose when speaking on the phone is related to their personal relationships with the people they are speaking with. One participant mentioned: "I mostly speak with people I know personally over the phone. If I don't know them, I may have a long conversation with them, but I wouldn't disclose any information." Therefore, individuals with higher privacy scores could also have a high call count and may engage in longer conversations without disclosing any personal or risky information. This is in-line with Baruh *et al.*'s (2017) analysis of privacy concerns on SNS where the frequency of SNS use was not correlated with individual privacy concerns.

Rather, calls that individuals choose to ignore or "miss" (refer *H3* and *H4*) can provide a better understanding of their privacy concerns. When participants were asked about their habits, it was observed that most participants with pragmatist or fundamentalist attitudes stressed that they would "only pick up calls when it was a familiar number" and participants with low privacy scores said they "typically answered all calls." The CPM theory talks about individuals desiring to maintain "control over access to self." The choice of deciding to not answer calls from unfamiliar numbers is an indication of individuals' using phone interactions to reinforce boundaries. Previous studies have analyzed information sharing and disclosure in online social networks (Dienlin and Metzger, 2016; Yao *et al.*, 2007) however, results from the current study show that an analysis of information sharing behavior over phones maybe a significant predictor of privacy attitudes.

There is a significant relationship between the number of new contacts (*H5*) included in an individual's network and their privacy attitudes. This implies that as an individual's network grows the amount of information disclosed increases. This is also in-line with the notions of co-ownership discussed in the CPM theory. According to the theory, original owners of information may choose to share information with others, in which case they would become co-owners of this information and have a "fiduciary responsibility" to protecting this information (Petronio, 2012). In the context of cell phone interactions, adding newer contacts results in the disclosure of metadata such as phone numbers as well as increases the probability of the call being accepted and therefore, increases chances of information being disclosed during cell phone interactions. During the qualitative interviews 10 out of 14 participants mentioned that they do not share numbers very readily, one participant also stated that "I don't want people to contact me directly, so I only give my number to people I know very well."

While similar results were found for online social networks (Acquisti and Gross, 2006; Tufekci, 2008), the present study shows that networks formed over phone calls, can be an important factor in determining an individual's information sharing patterns. This also implies that privacy attitudes are influenced by a number of factors and may not be

Q12

	Key variables	Hypothesis testing	p-value	Pearson's correlation coefficient
H1	Call count	Not Significant	0.917	−0.017
H2	Call duration	Not Significant	0.591	−0.085
H3	Call response rate	Significant	0.047	−0.308*
H4	Missed call rate	Significant	0.010	0.391**
H5	Number of new contacts	Significant	0.012	−0.383*

Table III. Results of hypotheses testing ($n = 53$) when controlling for age, gender, race, and personality traits

apparent by an examination of the simplest features (like total number of phone calls), but an analysis of more nuanced features, like the ratio of phone calls received to phone calls accepted, may reveal a more interesting pattern:

RQ2. How do information sharing practices vary across online, mobile phone and physical interactions?

Research has studied the benefits of exchanging information in order to communicate and build networks (Ellison, 2007). However, with the increase of SNSs, individuals' have a number of platforms to choose from when communicating with others. *RQ2* was framed with the intention of being able to gain a deeper understanding of how individuals perceive these communication platforms and the significance they assign to phone-based interactions. In order to answer this question, we conducted interviews with some participants and asked them to describe their data sharing practices over different communication media, specifically, we asked them to compare interactions over e-mail, social media, phones and in-person. We used the open and selective coding methods to analyze this data. During the open coding process, the importance attached to communications over phone when compared to e-mail or social media was clearly demonstrated. We therefore performed selective coding method with *RQ2* as the core concept and examined how individuals with different privacy attitudes felt about phone-based interactions.

We found an observable difference between how individuals with high or medium privacy scores view phone interactions when compared to participants with low privacy scores. For instance, participants with high privacy scores spoke about mobile phone interactions as "a more personal and intimate way to get in touch."

Another participant with a medium privacy score mentioned: "I would be more comfortable sharing my Facebook or my e-mail because those are much more general, a phone number gives you direct access to me, so I would be a lot more hesitant about sharing that". Almost all participants used words like "intimate," "direct" and "personal" when speaking about their phone numbers and words like "professional," "more general," and "safer" when they speak about SNSs. Participants also mentioned "trust" as the chief factor in determining whether they would share their phone numbers. For instance, participants said they would be extremely comfortable in sharing their phone numbers with their strong ties like parents or siblings, their roommates, and close friends, but had differing opinions when asked about weak ties like extended family or acquaintances, friends of friends, and people they met in a professional context.

Participants also spoke about using the depth of interaction as a measure for what form of contact information they would be more comfortable sharing. All participants with high privacy score spoke about "giving out their Facebook information" to someone they met for the first time in a social context. Some participants then had various other SNS information they would share, while others would give out their e-mail, but most participants with mid and high privacy scores mentioned that they would not share their phone numbers unless they had met the person face to face a few times and were "comfortable" and "felt a personal connection" with them. Participants with low privacy scores, on the other hand, mentioned "my phone is always around, so I typically would rather share my phone number than have people email me."

The CPM theory speaks about a "rule-based management system" that guides disclosure decisions. It predicts that individuals who have more privacy concerns will engage in stricter rules and boundaries (Petronio, 2012). The statements mentioned above show a link between the significance assigned to phone numbers and privacy attitudes of individuals. During the interviews, there was a clear differentiation between the information individuals share over phones vs the information shared over social media.

As participants view mobile phones as a more "personal" form of communication, it also implies that studying their phone usage behaviors could yield valuable clues about their privacy attitudes:

RQ3. Can a machine learning algorithm be used to predict privacy attitudes based on logged phone metadata?

Multiple significant hypotheses as well as participant comments during qualitative interviews suggest the predictive potential of nuanced phone usage metadata toward privacy attitudes. For instance participants with high privacy scores described being protective about their phone numbers as well as careful about accepting phone calls. Hence, the three features found to be significant in the aforementioned analysis were used to build a combined predictive model for privacy attitudes. We use the J48 decision tree algorithm to perform this analysis. This is a predictive machine learning model that decides the target value (privacy attitudes) of a new sample based on various attribute values of the available data (call response rate, missed call percentage, and newer contacts).

Two different classifications for privacy attitudes were considered. First, is the conventional three category classification as suggested by Westin and second is a two-class categorization based on the median value split.

In the first scenario the classes were defined based on the recommendations as made by Westin as already described in Section 2.1. This resulted in a split as follows: privacy fundamentalists, 5; privacy pragmatists, 31; privacy unconcerned, 17; total, 53. Given the multiple (> 2) classes present the multiclass classifier as implemented in Weka 3.6, was used with J48 decision tree as its underlying method (Bhargava *et al.*, 2013). Further considering the relatively modest sample size, the leave-one-out cross-validation was used as a tradeoff between the learning ability and the generalizability of results (Medelyan and Witten, 2008).

The proposed phone features based approach was also compared with two other approaches. One is a baseline "Zero-R" approach, which simply classifies all data into the largest category. The second approach is based on using personality variables (Big Five inventory), which have been shown by multiple efforts to be related with privacy attitudes (Junglas *et al.*, 2008). The same classification method was applied the different approaches. Lastly, given the unequal size of the classes, this study also reports the ROC – Area Under the Curve statistic along with the accuracy scores. Multiple prior efforts have suggested ROC as a more interpretable metric for unequal classes (Chawla, 2005).

As shown in Table IV, the phone features based model performed better than both the compared approaches. Focusing on the ROC metric, the model yielded 36 percent better prediction than the baseline model. Contrary to the expectations, the results also suggest that personality-based metrics may not capture the right kind of signals to have predictive ability on privacy attitudes. This result was better understood during the qualitative interviews (Stage 2), when participants with high openness and extraversion scores reported

Table IV. Classification results for three-way classification as per westin's taxonomy and two-way classification (high vs low privacy concern)

	Three-way classification		Two-way classification	
	Accuracy	ROC	Accuracy	ROC
Baseline (zero-*R*)	0.58	0.50	0.57	0.50
Personality features	0.53	0.50	0.43	0.50
Phone usage features	0.66	0.68	0.74	0.69

that though they would make friends easily, they would not exchange phone numbers. For instance, a participant with a medium privacy score reported "I hangout with people in class or in social gatherings [….] but I would only give my Facebook not my phone number." Participants also reported being more comfortable with exchanging emails or social media information with first time contacts rather than phone numbers. Therefore, while individuals may have high openness or extraversion scores, this does not necessarily translate to a lack of privacy concerns.

To ameliorate some of the complexities associated with multiclass (> 2) classification, a two-way classification problem was also considered, where the classes were based on a median split. Given that multiple participants fell at the median score, this resulted in two roughly equal classes of sizes of 30 (below or equal to median) and 23, respectively. The classification was then run in Weka using J48 decision tree algorithm with leave-one-out cross validation. As shown in Table III, this resulted in accuracy of 74 percent at a two-class classification task and an ROC metric of 0.69, which indicates a 38 percent improvement over the baseline. Again, the relatively poor performance of personality-based features suggests that traditional personality type measures may not be suited to predict privacy attitudes. While the current work has focused on a small number of features the significant increase obtained in prediction ability using just three features suggests value in exploring this direction further.

Limitations

As with any study, there are limitations of this work. The experiences of our participants are shaped by their cultural and geographic context, and thus research in other contexts may uncover other insights. The current sample includes students from one university in the Northeastern region of USA. and has limited diversity in terms of age. As people age and are at different stages of life, their privacy attitudes and needs change (Kezer *et al.*, 2016). Our sample also consists exclusively of college students and given the nature of data collected, there is a possibility of sampling bias. Also, a self-reported survey was used to measure privacy attitudes rather than observing and recording the participants' behavioral patterns in terms with respect to information sharing and disclosure. We also acknowledge the size of our sample is relatively modest. Taking into account these limitations, we will be cautious in generalizing the results found to larger populations until they are confirmed in other contexts and methods.

Theoretical and practical implications

The focus of this work is to investigate the potential of phone metadata as a predictor of individual privacy attitudes. In order to do this, we use methodological triangulation techniques (Denzin, 2012) combining statistical tests with predictive algorithms and qualitative interviews. We present a detailed and balanced investigation of the relationship between individual phone use patterns and privacy attitudes.

This work builds upon two major theoretical concepts – social capital and communication privacy theory – connecting social behavior and privacy needs.

Social capital describes the ability of individuals or groups to access resources embedded in their social network (Bourdieu, 1986) and hence individuals with higher privacy needs may engage in lesser interaction using phone. While such results have been reported in face-to-face interaction and physical connections, results from the current study indicate that the same may not be true for phone-mediated communications. In fact, the non-physical nature of interaction (which differentiates mobile phone-mediated communication from physical interaction), and synchronous nature interaction (which differentiates phone-mediated communication from online social networks) creates distinct dimensions for mobile phone usage behavior which differ from both online and face to face interaction.

Next, the study examines Petronio's (2012) CPM theory, with a particular focus on the context of phone-based interactions. Previous work suggests that boundary coordination is a means by which individuals may satisfy the tensions they experience in wanting to protect risky information while simultaneously receiving the benefits of disclosure. While, a previous study by Wisniewski *et al.* (2015) uses the CPM theory to understand tagging behavior of individuals on Facebook, this study presents an analysis of phone usage behaviors and reinterprets the concepts of boundary markers, boundary turbulence, as described by the CPM theory to understand phone use behavior.

According to Petronio (2012), people establish boundary markers through verbal (e.g. lowering the voice) and non-verbal (e.g. closing the door) codes. In the asynchronous context of phone-mediated communication contact lists act as virtual boundary markers. In this study, multiple participants indicated that they treat calls from contacts and unknown numbers quite differently. At the same time these boundaries are not universal and different users treat them differently. This study interprets the differences found between the behaviors of privacy fundamentalists and privacy unconcerned individuals (Westin, 2003) in light of Brown's (2006) classification of boundary types. It finds that the privacy unconcerned individuals often utilized soft boundaries (i.e. let anybody interested an access to them by responding to more number of calls) while privacy fundamentalists were more likely to maintain rigid boundaries and rarely allowed others access to their personal space. This behavior is demonstrated in phone-based interactions by intentionally "missing" or not responding to calls.

Boundary turbulence refers to the problems faced by owners when boundaries are not coordinated as well as they should be to maintain the level of privacy or exposure desired by owners (Petronio, 2012). When the boundaries are unclear, the participants may come into conflict with one another. In this study, the rejection of certain calls and a refusal to call back seems to be a corrective action to handle such boundary turbulence. Receiving calls from unexpected sources causes the turbulences and many participants deal with it by simply not accepting the phone calls and not responding to them.

While mobile phone-mediated communication does allow individuals more opportunities to segment their interactions and hence are often preferred over face to face interactions (O'Sullivan, 2000), a possible avenue for refining CPM is to further elaborate how medium selection functions as a form of boundary coordination and a means of avoiding boundary turbulence.

From a practical perspective, this work has implications for information scientists, government agencies, mobile app developers, and billions of end users. In today's world mobile phones have become an extension of an individuals' identity. A surprising amount of personal information is stored on phones and this data has been used in a number of areas from using phone-based features to predict changes in health, and automatically identify symptomatic days (Madan *et al.*, 2011) to predicting spending behavior based on phone records (Singh *et al.*, 2013). However, this data is rarely used to tackle the growing problem of aiding privacy needs of end users. While app designers often struggle to provide the right wizards and interfaces to users to specify their privacy needs (Stern and Kumar, 2014; Fang and LeFevre, 2010), this work suggests that many of the user's privacy needs can be inferred automatically by analyzing phone use behavior.

Conclusions and future work

The past few years have seen a great increase in the adoption of wearable devices and mobile technologies allowing for a constant and automatic gathering of personal data (Mann, 2004). The privacy of this personalized information has been debated intensely over the last few years. There is substantial concern about privacy in light of technological advances, greater sharing of information via social networks and mobile phones, and increased power of state and non-state actors to collect information about individuals and institutions. The governmental agencies' ability to access phone data records has reopened the conversation around the changing dynamics of privacy in today's digitized world (Petrie and Roth, 2015). Given the ubiquitous nature of phone interactions in today's world, it follows that the information contained within these devices can be a relevant starting point in understanding human behavior.

This work pivots the use of phone use metadata for the purpose of understanding and inferring an individual's privacy attitudes. In particular, we focus on call logs to ground the study and also build upon significant literature connecting communication, privacy, and personal traits (Balmaceda *et al.*, 2014). Based on a ten week field study involving an analysis of data collected via a phone-based logging app, privacy attitude questionnaire and follow up, this work reports that multiple phone signals have significant predictive power on an individual's privacy attitudes.

While the current work has focused on relatively simple set of features and tested a small number of hypotheses, the significant jump obtained in prediction ability points to the value in exploring this direction further. In particular, the direction of using more nuanced behavioral features, over a correspondingly larger sample size and degrees of freedom is part of future work from this study. With appropriate refinements and advancements, the proposed methodology could allow for automatic privacy attitude understanding for billions of mobile phone users.

Note

1. A previous version of this paper focusing only on the quantitative aspects appeared in conference proceedings (Ghosh and Singh, 2016). The current paper provides a more comprehensive treatment on the topic and includes significant additions in terms of additional research questions (*RQ2*) and more comprehensive treatment of the other research questions based on the qualitative interviews undertaken and the theoretical framework adopted.

References

Q3 Acquisti, A. and Gross, R. (2006), "Imagined communities: awareness, information sharing, and privacy on the Facebook", *International Workshop on Privacy Enhancing Technologies*, Springer, Berlin and Heidelberg, June, pp. 36-58.

Acquisti, A. and Grossklags, J. (2004), "Privacy attitudes and privacy behavior", *Economics of Information Security*, Springer, Boston, MA, pp. 165-178.

Adler, P.S. and Kwon, S.W. (2002), "Social capital: prospects for a new concept", *Academy of Management Review*, Vol. 27 No. 1, pp. 17-40.

Altman, I. (1975), *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding*, Brooks/Cole Pub. Co, Monterey, CA, p. 256.

Balmaceda, J.M., Schiaffino, S. and Godoy, D. (2014), "How do personality traits affect communication among users in online social networks?", *Online Information Review*, Vol. 38 No. 1, pp. 136-153.

Q4 Barnes, S.B. (2006), "A privacy paradox: social networking in the United States", *First Monday*, Vol. 11 No. 9.

Baruh, L., Secinti, E. and Cemalcilar, Z. (2017), "Online privacy concerns and privacy management: a meta-analytical review", *Journal of Communication*, Vol. 67 No. 1, pp. 26-53.

Beale, R. (2005), "Supporting social interaction with smart phones", *IEEE Pervasive Computing*, Vol. 4 No. 2, pp. 35-41.

Bhargava, N., Sharma, G., Bhargava, R. and Mathuria, M. (2013), "Decision tree analysis on j48 algorithm for data mining", *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3 No. 6.

Q5 Bourdieu, P. (1986), *The Forms of Capital Handbook of Theory and Research for the Sociology of Education*, *R.(1974). The Power Broker: Robert Moses and the Fall of New York*, pp. 241-258.

Brown, N.W. (2006), *Coping with Infuriating, Mean, Critical People: The Destructive Narcissistic Pattern: The Destructive Narcissistic Pattern*, Praeger Publishers/Greenwood Publishing Group, Westport, CT, p. 191.

Buchanan, T., Paine, C. and Joinson, A.N. (2007), "Internet privacy scales", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 2, pp. 157-165.

Candia, J., González, M.C., Wang, P., Schoenharl, T., Madey, G. and Barabási, A.L. (2008), "Uncovering Q6 individual and collective human dynamics from mobile phone records", *Journal of Physics A: Mathematical and Theoretical*, Vol. 41 No. 22, p. 224015.

Canzian, L. and Musolesi, M. (2015), "Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis", *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, September*, pp. 1293-1304.

Chin, E., Felt, A.P., Sekar, V. and Wagner, D. (2012), "Measuring user confidence in smartphone security and privacy", *Proceedings of the Eighth Symposium on Usable Privacy and Security, ACM, July*, p. 1.

Q7 Cohen, J. (1992), "A power primer", *Psychological Bulletin*, Vol. 112, pp. 155-159, doi: 10.1037/0033-2909.112.1.155.

De Montjoye, Y.A., Shmueli, E., Wang, S.S. and Pentland, A.S. (2014), "Openpds: protecting the privacy of metadata through safeanswers", *PloS One*, Vol. 9 No. 7, p. e98790.

Denzin, N.K. (2012), "Triangulation 2.0", *Journal of Mixed Methods Research*, Vol. 6 No. 2, pp. 80-88.

Derlega, V.J. and Chaikin, A.L. (1977), "Privacy and self-disclosure in social relationships", *Journal of Social Issues*, Vol. 33 No. 3, pp. 102-115.

Dienlin, T. and Metzger, M.J. (2016), "An extended privacy calculus model for SNSs: analyzing self-disclosure and self-withdrawal in a representative US sample", *Journal of Computer-Mediated Communication*, Vol. 21 No. 5, pp. 368-383.

Ellison, N.B. (2007), "Social network sites: definition, history, and scholarship", *Journal of Computer-Mediated Communication*, Vol. 13 No. 1, pp. 210-230.

Fang, L. and LeFevre, K. (2010), "Privacy wizards for social networking sites", *Proceedings of the 19th International Conference on World Wide Web, ACM*, *April*, pp. 351-360.

Q8 Gurrin, C., Albatal, R., Joho, H. and Ishii, K. (2014), "A privacy by design approach to lifelogging", in O'Hara, K., Nguyen, C. and Haynes, P. (Eds), *Digital Enlightenment Yearbook 2014*, IOS Press, June, pp. 49-73.

Introna, L. and Pouloudi, A. (1999), "Privacy in the information age: stakeholders, interests and values", *Journal of Business Ethics*, Vol. 22 No. 1, pp. 27-38.

John, O.P. and Srivastava, S. (1999), "The big-five trait taxonomy: history, measurement, and theoretical perspectives", in Pervin, L.A. and John, O.P. (Eds), *Handbook of Personality: Theory and Research*, Vol. 2, Guilford Press, New York, NY, pp. 102-138.

Joho, H., Gurrin, C., Heinström, J. and Matsubayashi, M. (2016), "Information practices meet lifelogging technologies: towards a successful multimethod research framework", *Proceedings of the Association for Information Science and Technology*, Vol. 53 No. 1, pp. 1-4.

Junglas, I.A., Johnson, N.A. and Spitzmüller, C. (2008), "Personality traits and concern for privacy: an empirical study in the context of location-based services", *European Journal of Information Systems*, Vol. 17 No. 4, pp. 387-402.

Karlson, A.K., Brush, A.J. and Schechter, S. (2009), "Can I borrow your phone?: understanding concerns when sharing mobile phones", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM*, *April*, pp. 1647-1650.

Kezer, M., Sevi, B., Cemalcilar, Z. and Baruh, L. (2016), "Age differences in privacy attitudes, literacy and privacy management on Facebook", *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, Vol. 10 No. 1.

Krasnova, H., Spiekermann, S., Koroleva, K. and Hildebrand, T. (2010), "Online social networks: why we disclose", *Journal of Information Technology*, Vol. 25 No. 2, pp. 109-125.

Q9

Kraus, S.J. (1995), "Attitudes and the prediction of behavior: a meta-analysis of the empirical literature", *Personality and Social Psychology Bulletin*, Vol. 21 No. 1, pp. 58-75.

Madan, A., Cebrian, M., Moturu, S., Farrahi, K. and Pentland, A. (2011), "Sensing the 'health state' of our society", *Institute of Electrical and Electronics Engineers*.

Mann, S. (2004), "Continuous lifelong capture of personal experience with EyeTap", *Proceedings of the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences, ACM*, *October*, pp. 1-21.

Medelyan, O. and Witten, I.H. (2008), "Domain-independent automatic keyphrase indexing with small training sets", *Journal of the Association for Information Science and Technology*, Vol. 59 No. 7, pp. 1026-1040.

Miritello, G., Moro, E., Lara, R., Martínez-López, R., Belchamber, J., Roberts, S.G. and Dunbar, R.I. (2013), "Time as a limited resource: communication strategy in mobile phone networks", *Social Networks*, Vol. 35 No. 1, pp. 89-95.

Onnela, J.P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J. and Barabási, A.L. (2007), "Structure and tie strengths in mobile communication networks", *Proceedings of the National Academy of Sciences*, Vol. 104 No. 18, pp. 7332-7336.

O'Sullivan, P.B. (2000), "What you don't know won't hurt me: impression management functions of communication channels in relationships", *Human Communication Research*, Vol. 26 No. 3, pp. 403-431.

Petrie, C. and Roth, V. (2015), "How badly do you want privacy?", *IEEE Internet Computing*, Vol. 19 No. 2.

Petronio, S. (2012), *Boundaries of Privacy: Dialectics of Disclosure*, State University of New York Press, Albany, NY, p. 268.

Razmerita, L., Kirchner, K. and Sudzina, F. (2009), "Personal knowledge management: the role of web 2.0 tools for managing knowledge at individual and organisational levels", *Online Information Review*, Vol. 33 No. 6, pp. 1021-1039.

Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M. and Rao, J. (2009), "Understanding and capturing people's privacy policies in a mobile social networking application", *Personal and Ubiquitous Computing*, Vol. 13 No. 6, pp. 401-412.

Schlegel, R., Kapadia, A. and Lee, A.J. (2011), "Eyeing your exposure: quantifying and controlling information sharing for improved privacy", *Proceedings of the Seventh Symposium on Usable Privacy and Security ACM, July*, p. 14.

Singh, V.K., Freeman, L., Lepri, B. and Pentland, A.S. (2013), "Predicting spending behavior using socio-mobile features", *Social Computing (SocialCom), 2013 International Conference on, IEEE, September*, pp. 174-179.

Stern, T. and Kumar, N. (2014), "Improving privacy settings control in online social networks with a wheel interface", *Journal of the Association for Information Science and Technology*, Vol. 65 No. 3, pp. 524-538.

Stutzman, F. and Kramer-Duffield, J. (2010), "Friends only: examining a privacy-enhancing behavior in Facebook", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1553-1562.

Swan, M. (2013), "The quantified self: fundamental disruption in big data science and biological discovery", *Big Data*, Vol. 1 No. 2, pp. 85-99.

Tufekci, Z. (2008), "Can you see me now? Audience and disclosure regulation in online social network sites", *Bulletin of Science, Technology & Society*, Vol. 28 No. 1, pp. 20-36.

Vincent, J. (2006), "Emotional attachment and mobile phones", *Knowledge, Technology & Policy*, Vol. 19 No. 1, pp. 39-44.

Wang, X., Hong, Z., Xu, Y.C., Zhang, C. and Ling, H. (2014), "Relevance judgments of mobile commercial information", *Journal of the Association for Information Science and Technology*, Vol. 65 No. 7, pp. 1335-1348.

Q10 Warren, S.D. and Brandeis, L.D. (1890), "The right to privacy", *Harvard Law Review*, pp. 193-220.

Watson, J., Lipford, H.R. and Besmer, A. (2015), "Mapping user preference to privacy default settings", *ACM Transactions on Computer-Human Interaction (TOCHI)*, Vol. 22 No. 6, p. 32.

Q11 Westin, A. (2003), "Consumer, privacy and survey research", August 17, p. 2004.

Wisniewski, P., Xu, H., Lipford, H. and Bello-Ogunu, E. (2015), "Facebook apps and tagging: the trade-off between personal privacy and engaging with friends", *Journal of the Association for Information Science and Technology*, Vol. 66 No. 9, pp. 1883-1896.

Xu, H., Gupta, S., Rosson, M.B. and Carroll, J.M. (2012), "Measuring mobile users' concerns for information privacy", *33rd International Conference on Information Systems*, *Orlando, FL*.

Yao, M.Z., Rice, R.E. and Wallis, K. (2007), "Predicting user concerns about online privacy", *Journal of the Association for Information Science and Technology*, Vol. 58 No. 5, pp. 710-722.

Young, A.L. and Quan-Haase, A. (2009), "Information revelation and internet privacy concerns on social network sites: a case study of Facebook", *Proceedings of the Fourth International Conference on Communities and Technologies, ACM*, *June*, pp. 265-274.

Further reading

Almuhimedi, H., Schaub, F., Sadeh, N., Adjerid, I., Acquisti, A., Gluck, J., Cranor, L.F. and Agarwal, Y. (2015), "Your location has been shared 5,398 times!: a field study on mobile app privacy nudging", *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM*, *April*, pp. 787-796.

Chawla, N.V. (2009), "Data mining for imbalanced datasets: an overview", *Data Mining and Knowledge Discovery Handbook*, Springer, Boston, MA, pp. 875-886.

Ghosh, I. and Singh, V.K. (2016), "Predicting privacy attitudes using phone metadata", *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Springer*, *Cham*, *June*, pp. 51-60.

Walrave, M., Vanwesenbeeck, I. and Heirman, W. (2012), "Connecting and protecting? Comparing predictors of self-disclosure and privacy settings use between adolescents and adults", *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, Vol. 6 No. 1.

Wang, Y., Leon, P.G., Scott, K., Chen, X., Acquisti, A. and Cranor, L.F. (2013), "Privacy nudges for social media: an exploratory Facebook study", *Proceedings of the 22nd International Conference on World Wide Web, ACM*, *May*, pp. 763-770.

Westin, A.F. and Ruebhausen, O.M. (2015), *Privacy and Freedom*, Ig Publishing, VT, p. 500.

Appendix 1

For each of the above statements the following options are provided:

(1) Strongly Disagree

(2) Somewhat Disagree

(3) Somewhat Agree

(4) Strongly Agree

(1) Consumers have lost all control over how personal information is collected and used by companies

(2) Most businesses handle the personal information they collect about consumers in a proper and confidential way

(3) Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today

Appendix 2

Table AI. Correlation scores of personality traits with privacy attitudes

		Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness
Score privacy	Pearson Correlation	−0.168	0.075	−0.046	0.005	0.016
	Sig. (two tailed)	0.228	0.592	0.743	0.974	0.909

Corresponding author
Isha Ghosh can be contacted at: isha.ghosh@rutgers.edu