Journal of Computing and Information Technology - CIT 23, 2015, 2, 113–120
doi:10.2498/cit.1002535

Factors Influencing User's Attitude to Secondary Information Sharing and Usage

Johnson Iyilade, Rita Orji, Julita Vassileva

Department of Computer Science, University of Saskatchewan, Canada

The increasing availability of enormous data about users online, along with availability of sophisticated tools and technology to store, aggregate, and analyze data for secondary use has raised concerns about how to balance the opportunity for secondary use of data with the need to protect the user privacy that may result from harmful use. To develop a privacy protection mechanism that is useful and meets the expectations and needs of the user, it is important to understand users' attitude to privacy and secondary information sharing and usage of his/her data. While several studies have investigated factors influencing users' attitude to privacy in primary data collection context, none of the existing studies have provided an understanding of user perception and attitude to privacy in secondary context. To fill this gap, this work has identified five factors that are important in a secondary usage context and carried out a study on their influence on users' perception with respect to how their data is shared for secondary use. The main contribution of this paper is an understanding of factors influencing user decisions about privacy in secondary context, which can assist both technology designers and policy makers in the development of appropriate privacy protection that meets the needs and expectations of the user.

Keywords: privacy, secondary use, privacy attitudes, big data, user modeling, Structural Equation Modeling

1. Introduction

Recent advances in mobile, social and ubiquitous computing have made it possible for applications and devices to gather enormous data from/about the user in various contexts and through various means (e.g. volunteered by the user; from observation of user activities; or inferred from volunteered and observed data) (Poslad, 2009; Seng, 2007).

Although data collected directly from the user by an application or device (so called *primary data*) are currently fragmented and stored in various isolated applications, devices, and databases online, there is a growing trend towards developing tools and technologies that facilitate aggregation and reuse of these data for new purposes (so called *secondary sharing and use*) that might not be known during the time of primary collection (Iyilade and Vassileva, 2013a). For example, in user modeling community, many frameworks and tools have been developed in the past decade for collecting and processing users' data drawn from many sources in what is often referred to as *decentralized* (Vassileva et al., 2003) or *cross-system user modeling* (Carmagnola and Cena, 2011; Abel, et al., 2010). Similarly, recent technological advances in Big Data analytics have made it possible to easily aggregate large volume of user data from many sources for advance knowledge discovery (WEF, 2012).

Arguably, allowing applications to share user data for secondary use will benefit both the user and society at large. For example, the user can get better personalized services, since richer information about his/her activities and interests in various contexts will be available for personalization (Heckmann et al., 2005). In addition, the opportunity for aggregating user data from many sources in *Big Data analytics* is driving innovation in many areas such as healthcare, education, national security, law enforcement, fraud detection in credit card payment, urban planning, disaster recovery, and optimization of energy consumption (Podesta et al., 2014).

The above-highlighted benefits, however, are being overshadowed by the increasing concerns about the risks of secondary sharing to users' privacy. Since, in many cases, the data being shared are electronic logs of users' private and public lives, the data could be used for potentially harmful purposes such as surveillance or profiling the user for targeted discrimination with respect to employment, insurance and loans (Toch et al., 2012; Machanavajjhala, 2008).

To develop a privacy-enhancing technology that balances the opportunity for beneficial (secondary) use of data with the need to protect the user from potential privacy risks from misuse of his/her data, there is the need to first understand the factors influencing users' attitudes towards privacy and secondary sharing and use of their data. This will enable the design of technical tools and solutions that meet the expectations and needs of the user (Cranor et al., 1999).

In the past, several studies (e.g. Hann, et al. 2002; Tsai et al. 2011; Virpi et al. 2009; Knijnenburg and Kobsa, 2013; Acquisti et al. 2009) have been conducted to understand the factors influencing user privacy decisions, attitudes or behaviors, however, most of these studies have focused on privacy and information disclosure in the context of initial (primary) data collection, hence, factors that impact the users' information sharing attitudes in a *secondary data sharing and usage context have remained unexplored*. Furthermore, even where the factors analyzed in previous studies are relevant to secondary settings, the factors were investigated in isolation (each one as a single determinant of the users' attitude or decision). There is, therefore, the need to investigate the correlation of the factors that are influencing the users' decision to protect his/her privacy or allow secondary sharing of his/her data.

In this paper, we address this challenge by conducting a study with 822 participants on the factors influencing their attitudes toward secondary sharing and reuse of their online information. More specifically, we investigate the influence of the following five factors: *perceived benefits* (of secondary information sharing), *perceived risks*, perceived sense of *control*, attempt to *gain or protect online reputation* on users' secondary information sharing attitudes. As a secondary objective, we also investigate whether there are significant differences in the influence of these factors on users' attitude to secondary information sharing across younger (below 25 years) and older (above 35 years) age groups. The results of the study provide valuable insights into the privacy principles that are important to the users, as well as design guidelines for developing a privacy framework for secondary user data sharing and for tailoring default privacy policies to users of various age groups.

The rest of the paper is organized as follows: Section 2 gives a brief background into secondary user data sharing and its implications for privacy. It then discusses existing studies on user privacy attitudes and identifies factors that influence user privacy decisions in secondary context. Thereafter, in Section 3, we present our study design and methodology. After that, we present the results of our study and their interpretations in Section 4. We discuss the practical implications of our work for policy makers and for the design of privacy technologies and tools in Section 5. Finally, in Section 6, we draw conclusions, discuss limitation of our study and directions for future work.

2. Background and Related Work

2.1. Privacy in Secondary Context

The concept of *privacy* has been defined first as the "right to be left alone" by Warren and Brandeis in 1890 (Spikermann and Lorrie, 2009). Recently, concerns for privacy have increased dramatically as computer technologies allow us to collect, store, aggregate, and analyze huge amounts of data inexpensively. As a result of the rapid influx of new technologies and devices in the last decade, existing privacy laws and regulatory frameworks are lagging behind in providing the needed balance between user privacy protection and the growing need for information sharing (Tene and Polonetsky, 2012). Existing laws are mostly based on the Organization for Economic Co-operation and Development (OECD) Privacy Principles agreed in 1980 (OECD, 1980) by many countries and on the principle of informed consent for data flow (Langheinrich, 2001). Generally, secondary data sharing and use for the purposes not explicitly stated at the point of data collection,

and for which the user has given explicit consent, have been considered a privacy hazard and as seek forbidden in the legislation of several countries (Tene and Polonetsky, 2012). In reality, however, technological progress, particularly advances in User Modeling, Personalization, and Big Data technologies, have made secondary data sharing and reuse a trend that could no longer be ignored. User data sharing and reuse are happening without the users' awareness and with limited means of control available to the user. Therefore, the challenge of privacy is shifting from focusing on limiting data collection to protecting the users from privacy risks due to misuse of their data for secondary purposes. To design privacy technologies that enable flow of data for beneficial use while protecting the user from potential privacy risks, it is important to understand user perception with respect to privacy in secondary context. This will ensure the solution is useful and meets the expectations of the user.

2.2. Related Work

For many years, researchers at the intersection of information disclosure and privacy have investigated various factors influencing users' privacy attitude or behavior in information disclosure decisions. These studies span many fields (e.g. Information Science, Economics, Law, Marketing and Computer Science) (Knijnenburg and Kobsa, 2012) and carried out to measure user privacy attitudes or decisions in various contexts (e.g. e-commerce websites (Hann et al., 2002), social networks (Tuunainen et al., 2009), recommender systems (Knijnenburg and Kobsa, 2012), location-based services (Cvrcek et al., 2006), etc.). To the best of our knowledge, none of the existing studies has been focused on users' attitude toward allowing secondary sharing of their data across applications and services. We, therefore, present a brief review of the related work and, thereafter, summarize it with the factors that we investigate in this study.

In Acquisti et al. (2009), the authors investigated users' privacy decision through the lens of behavioral economics and decision theories. In the work, privacy is viewed as an economic good where the users' decision for more or less privacy protection can be impacted by non-rational influences such as *endowment and order effects*. The *endowment effect* (Thaler, 1980) states that people would place more value on what they have than on what they do not have. In relation to privacy, this means people will demand more money to give up their privacy than they would be willing to pay to purchase more privacy. The *order effect*, on the other hand, measures how the order in which the request for data is presented influences the disclosure decision. Also, in Aperjis and Huberman (2012), providing *compensation* to the user was found to influence the users' privacy decision. Another study by Hann et al. (2002) attempts to quantify the monetary value that individuals attach to their personal information on websites. The study concluded that, among US participants, protection *against secondary use* is topmost concern of user and worth $39.83 to $49.78 while protection against *improper access* and *errors* is respectively valued at $29.18 to $36.47 and $15.46 to $19.32.

In e-commerce websites, Tsai et al. (2011) investigate whether a *prominent display of privacy information* will cause consumers to incorporate privacy considerations into their online purchasing decisions. The study found that the presence of *privacy indicator, statement or trust seal* in websites increases consumers purchasing interest. In addition, Egelman et al. (2009) found that the *timing* and *placement* of the online privacy indicator has an impact on user behavior.

Privacy decisions have also been investigated in the context of online social networks. Virpi et al. (2009) study the effect of *awareness of the privacy risks* of information disclosure on the users' decision to disclose or protect information on Facebook. In a sample of 210 Facebook users, the study found that users' awareness of the risks will make them limit the information they share on social network.

In recommender systems, some studies have found that presenting *justification* (Knijnenburg and Kobsa, 2012) for data collection and *satisfaction with the system* (Knijnenburg and Kobsa, 2013) influence the users' decision to disclose personal information.

Another study by Brandimarte et al. (2012) investigates the effect of *perceived control* over users' propensity to disclose private and sensitive information to websites. The study found

that users are more comfortable supplying personal data to websites when they feel in control, even if that control might be illusory. The focus of the study, however, is on the user having control of initial (primary) disclosure of data to websites and not the subsequent secondary sharing and usage. We believe, the objective risks nowadays arise not from control over initial disclosure, but from having some control over secondary usage of data.

Another factor that has emerged as an important influence on user information sharing attitude is the concern users have for their online reputation (Acquisti and Gross, 2006). A 2010 Pew report indicated that more than 70% of Internet users aged 18-29 say they have acted to limit what they share online in order to guard their reputation (Madden and Smith, 2010). It has also been found that users differ in their attitudes and behaviors around reputation and privacy. While some users are concerned about and manage their online reputation information (for instance, by customizing privacy settings and changing online behaviors) (Acquisti and Gross, 2006), others are largely unconcerned (Kumaraguru and Cranor, 2005).

Finally, a recent study in 2012 by IIC (2012) defines seven data contexts from user perspectives for data management that impact user information sharing decisions. The data contexts proposed include: (i) *Type of data* being requested (ii) *Type of Entity* requesting the data (iii) *Trust in the service provider* (iv) *Collection Method* (that is, whether the data is collected via passive or active means) (v) *Device context* (i.e if the data is being requested on a mobile device or PC) (vi) *Data Usage* (i.e what purpose is the data going to be used for) (vii) *Value Exchange* (i.e. if there is any commensurate value to the user for allowing use of data).

A critical look at the existing studies on factors influencing user information sharing attitudes reveals that many of the studies were carried out in various contexts (e.g. for websites, recommender systems, location-based services etc) and that most of them focus on factors influencing the users' disclosure decisions in primary data collection (i.e. by the application that collects the data for its own adaptation purpose) rather than sharing and reuse of already collected data for secondary purposes. However, factors influencing the users' attitude about secondary data sharing are different and likely to give different results when investigated in a secondary user information sharing context. This is because, unlike in primary context where data disclosure to a particular website or server involves direct interaction by the user with the system, secondary information sharing occurs, typically, in a peer-to-peer fashion (among collaborating applications, devices, web-services, sensors and agents) and involves little or no direct communication or interaction with the user. Hence, some of the subjective factors (such as the *presence of privacy indicator, statement or trust seal* (Hann et al., 2002) and *satisfaction with the system* (Knijnenburg and Kobsa, 2012)) that were found to influence users' information sharing decisions in websites may have little or no effect on the users' attitudes in secondary context since they are based on the "look and feel" of websites. Furthermore, even in cases where the factor is relevant to the users' information sharing attitude in secondary settings, in most of the previous studies the factors were investigated in isolation, as a single determinant of users' behavior (with the exception of Knijnenburg and Kobsa, 2012, 2013). To this end, there is the need for a unified approach that investigates the correlation of these factors and their relative influence on users' attitude in the context of secondary user data sharing and use.

Based on the insights from literature on various factors influencing users' data disclosure in primary context, we suggest five factors that encapsulate the relevant factors for measuring user attitudes in secondary context.

These are:

(i) ***Perceived benefits*** of secondary information sharing – The benefits considered are not just monetary but other benefits of secondary information sharing such as *use for public goods*, to get *personalized services*, etc.

(ii) ***Perceived risks*** associated with secondary user information sharing

(iii) Having ***control*** over who has access to the information, who it is shared with and for what purpose.

(iv) ***Gain reputation*** – would user be interested in allowing or not allowing secondary shar-

ing of information in order to gain reputation?

(v) ***Protect reputation*** – similarly, would the user perception about protecting their online reputation influence them to allow sharing or not allow secondary sharing of their information?

We seek to develop a model that measures the relative importance of these factors to the users' attitude to secondary information sharing and privacy decisions. Moreover, we are interested in the variation of the impact of these factors across various age groups.

3. Research Design and Method

In this section, we describe how we develop the research instrument, data collection, and validation methods.

3.1. Measurement Instrument

To collect data for our model, we use a scenario-based approach to elicit user response to various questions. We present a scenario that involves user information sharing for secondary purposes. The scenario is followed by questions measuring the relative importance of the factors and their influence on users' attitude towards allowing sharing of their information for secondary use. The survey was developed based on the outcome of the review in Section 2.3. We were specifically interested in determining the influences of the following five factors: (1) *perceived benefits*, (2) *perceived risk*, (3) *control*, (4) *gain reputation*, and (5) *protect reputation* on *user attitude* to allowing secondary sharing and use of their data. We presented the following scenario (adapted from IIC, 2012) for secondary user information sharing after initial collection:

"Nowadays, it is increasingly common for people to engage in various online activities on websites and mobile apps. It is also possible for people to wear shoes, clothes, or watches that have embedded sensors. For example, people engage in online shopping and buy clothes, sport goods, books, travel tickets. They purchase films, music and games; they compare prices of goods and services; they also use social networking and share sites such as Facebook, Twitter and LinkedIn to keep in touch with friends and family, conduct business, meet new friends, play games, and stay in touch with events around them. In addition, they use wearable sensors to track their physical exercises, food intake etc. The data/information you have shared online or that was collected about you by websites or mobile apps may sometimes be re-shared and reused for other secondary purposes beyond the original purpose for which it was collected".

This scenario is augmented by follow-up scenarios and scales for measuring individual factors. The scales include: (1) five questions for assessing the perceived benefits – e.g., *I will allow secondary sharing of my data with applications to get better services that are tailored to my preferences and needs*; (2) four questions for measuring perceived risk – e.g.; *I am concerned that my current online data may be misinterpreted resulting in discrimination, penalization, and even persecution*; (3) five questions for measuring control – e.g. *To what extent is your ability to control the kind of data shared about you important to you?*; (4) three questions for measuring gain of reputation – e.g., *I share personal information online to improve how I am perceived by my colleagues, friends, or peers*; and (5) five questions for measuring protection of reputation – e.g., *I do not share personal information online to prevent my employer from making wrong judgments about me*, and (6) six questions for measuring the users' attitude toward secondary information sharing – e.g., *I consider allowing sharing my data for secondary purposes as good*. All the factors apart from control were measured using a 5-level Likert scale ranging from "1 = Strongly disagree" to "5 = Strongly agree". The control factor was measured using 5-Likert scale ranging from "1 = Extremely unimportant" to "5 = Extremely important".

Prior to assessing participants' perception of individual factors, we ensured that the participants understood the individual factors by asking them a comprehension question. To achieve this, we included an open-ended question before each of the scales for measuring individual factors asking the participants to list the various risks, benefits, control, and reputation – related factors they associate with information sharing.

These questions, apart from making the participants reflect on the individual factors, ensured that our participants understood the factors well. Responses from participants who gave meaningless answers to the comprehension question were discarded. We further include open-ended questions at the end of each factor, allowing participants to provide additional comments about each factor. Finally, we included questions for assessing participants' demographic information.

3.2. Data Collection

To collect data for our study, we recruited participants using Amazon's Mechanical Turk (AMT). AMT has become an accepted method of recruiting study participants and several studies have successfully used AMT (For example: Buhrmester and Kwang, 2011; Mason and Suri, 2012; Heer and Bostock, 2010; Orji et al. 2014). We followed the recommendations by Mason and Suri (2012) for performing effective studies on AMT to overcome potential challenges associated with recruiting participants from AMT. Specifically, we used captcha to ensure that we retain only human participants in our survey. We used a mechanism provided by AMT that allows collection of responses from unique participants to ensure that participants could respond to our study only once. The study took an average of 15 minutes to complete. The responses from participants who completed a less than 10 minutes were discarded. We collected a total of 853 responses and retained a total of 822 valid responses, which were included in our analysis.

Before the main study, we conducted two pilot studies. The first pilot study was conducted on 47 participants (35 participants from AMT and 12 participants recruited from the Department of Computer Science, University of Saskatchewan, Canada) to test the validity of our study instruments and to compare the results. The preliminary evaluation shows similar results from the participants recruited from AMT and those from the university; however, it also revealed a need to re-word some of the study questions for understandability. We restructured the questions and conducted a second pilot study on another 11 randomly selected participants. The second pilot confirmed the suitability and understandability of our study instrument.

3.3. Participants' Demographic Information

A total of 822 participants were retained in our study and their demographic information is summarized in Table 1. The participants received $0.25 USD dollar compensation, which is within the range of the standard rates for other tasks recruited through AMT.

Total Participants = 822	
Gender	Females (340, 41%), Males (482, 59%)
Age	14-25 (322, 39%), 26-35 (341, 41%), 36-45 (91, 11%), Over 45 (68, 8%).
Education	Less than High School (4, 0.5%), High School Graduate (135, 16.4%), College Diploma (135, 16.4%), Undergraduate Degree (213, 25.9%), Graduate Degree (335, 40.8%).
Country	Canada (8, 1%), India (537, 65%), Macedonia (7, 1%), United States (206, 25%), Others (74, 8%).

Table 1. Participants' demographic information.

3.4. Data Analysis

The aim of this paper is to examine the factors influencing users' attitude toward allowing secondary information sharing. The paper also investigates whether significant differences exist across younger and older age groups with respect to their perception of various factors. This will inform design guidelines for tailoring privacy policies to various age groups. To achieve this, we used several well-known analytical tools and procedures. In this section, we present the details of the analysis. We also describe the results of the modeling process.

3.4.1. Measurement Validation

We determined the suitability of our data for factor analysis using the Kaiser-Meyer-Olkin (KMO) sampling adequacies and the Bartlett Test of Sphericity. Our results showed that the KMO was 0.86, well above the recommended value of 0.6; that the Bartlett Test of Sphericity was significantly significant ($\chi^2(378) =$

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.509	30.390	30.390	8.509	30.390	30.390
2	2.957	10.561	40.951	2.957	10.561	40.951
3	2.350	8.394	49.345	2.350	8.394	49.345
4	2.212	7.900	57.244	2.212	7.900	57.244
5	1.467	5.239	62.483	1.467	5.239	62.483
6	1.285	4.588	67.071	1.285	4.588	67.071

Table 2. Eigenvalue and total variance explained- factors with Eigenvalue less than 1 have been removed.

10965.103, $p < 0.0001$); and that all of the communalities were well above 0.3.

These results show that our data was suitable for factor analysis. We performed Exploratory Factor Analysis (EFA) – a statistical procedure that identifies the number of latent factors in a set of variables – using Principal Component Analysis (PCA) to determine the appropriate number of factors in our data. We first examined the scree plot of eigenvalue against the component number and considered factors with eigenvalue of at least 1. As shown in Table 2, there are six factors with an eigenvalue of at least 1 and the six factors explained a total cumulative variance of 67%. We further examined the six-factor solution using Varimax rotation (Brown, 2009). All the 28 items (except item number 5 measuring control with factor loading of .197) had factor loading greater than 0.30 and cross loading less than 0.30 and were therefore retained and included in our analysis. The 0.30 level is an accepted minimum loading because it indicates that the factor explained at least 10% of the variance in the corresponding variable (Tinsley and Tinsley, 1987). The PCA shows that the six factors – perceived benefit, perceived risk, control, gain reputation, protect reputation, and information sharing attitude – loaded into six different factors. We present the descriptive of each of the factors extracted from the PCA in Table 3.

3.4.2. The Measurement Model

After establishing that our data was suitable for factor analysis and determining the number of factors in the data using PCA, we employed the Partial Least Square (PLS) Structural Equation Modeling (SEM) (Kupek, 2006) to develop models showing the factors influencing users' information sharing attitude. SEM has been successfully applied in building models and estimating mediating factors to privacy and information disclosure decisions in recommender systems (Knijnenburg and Kobsa, 2013). PLS-SEM was chosen in our analysis because: One, it has less stringent requirements concerning data distribution assumptions (Henseler et al. 2009) and it is appropriate for complex predictive models (Baron and Kenny, 1986). Two,

Factors	# of questions	Mean (SD)	Cronbach's α
Perceived benefit	4	3.21 (0.85)	.794
Perceived risk	8	3.68 (0.83)	.721
Control	4	3.28 (0.79)	.755
Gain Reputation	4	3.07 (1.08)	.805
Protect Reputation	4	3.16 (0.93)	.841
Attitude Toward Information Sharing	6	4.45 (1.19)	.916

Table 3. Overview of the mean score and standard deviation factors for user information sharing.

it can accommodate small sample sizes, as opposed to covariant-based SEM. We used SmartPLS 2.0 (M3) (Ringle et al., 2012) in estimating our model.

Before estimating the structural path to examine the relationship between the variables we validated the measurement model using the criteria suggested by Chin (1998). PLS-SEM uses convergent validity, discriminate validity, and composite reliability to measure the suitability of any scale. We report here the common set of indices recommended for model validity and reliability in PLS. Using the criteria from Chin (1998) and Fornell and Larcker (1981), indicator reliability can be assumed because Cronbach's α and the composite reliability that analyzes the strength of each indicator's correlation with their variables are all higher than the threshold value of 0.7. Convergent and discriminate validity can be assumed as all constructs have an Average Variance Extracted (AVE) (which represents the variance extracted by the variables from its indicator items) above the recommended threshold of 0.5 and greater than the variance shared with other variables (Setterstrom et al., 2012). The measurement models yielded an acceptable value of all indices for PLS model validity and reliability.

Prior to comparing our models, we tested for measurement invariance across the data sample for the younger and older age groups. It is important to establish that the younger and older adults had similar interpretations of our study instrument's items. Establishing measurement invariance ensured that we have not measured different phenomena across the sub-groups. To assess measurement invariance, we used the component-based CFA in SmartPLS 2.0 (M3) (Ringle et al. 2012) to conduct factor analysis for each sub-group of data and retained items that had factor loadings of at least 0.5 (Hair et al. 2011) in all the sub-groups (and dropped items with loadings less than .5 for all groups) thereby establishing factor invariance. Items that were significantly different were dropped for the two sub-groups. This process established measurement invariance and ensured that our data were suitable for multi-group comparison (Setterstrom and Pearson, 2012).

3.4.3. Assessing Age as a Reliable Characteristics for Personalizing Privacy Policy

To examine the factors influencing users' attitude toward information sharing, we developed a model (using the 822 data entries retained in our survey) to show the relationship between the factors perceived benefit, perceived risk, control, gain reputation, and protect reputation and attitude towards information sharing. Again, to test for the moderating effect of age, we decided to compare two distinct age groups: 14-25 years (younger) and over 35years (older); this eliminates the tendency of overlap. We developed additional models – one for each of the younger and older age groups.

To establish that age is a reliable characteristic for personalizing privacy policy, we assess significant structural differences between the models for the younger and older age groups using the pairwise comparison approach recommended by Chin (2013). We found significant differences across the age groups; therefore, we establish that age is a reliable characteristic for personalizing privacy policy. Again, following the pairwise comparison, we controlled any possible family-wise type I error (due to multiple comparisons) using the Bonferroni-Holm adjustment (Holm,1979).

4. Results and Interpretation

As noted previously, we created three models (summarized in Table 4) – one for all ages (i.e the general population) and two for each of the younger and older age groups. We further discuss the findings in the general discussion.

4.1. The Structural Model

The structural models determine the relationship between the factors (perceived benefit, perceived risk, control, gain reputation, and protect reputation) and attitude toward information sharing. To measure the strength of the relationship between variables in structural models, we calculate the level of the path coefficient (β) and the significance of the path coefficient (p) (Hair, et al., 2011). Path coefficients measure the influence of one variable on another. The

Factors / Population	Benefit	Risk	Control	Gain Reputation	Protect Reputation
All Ages (above 14)	.26	-.01	.12	.33	.10
14-25	.06	-.05	.39	.33	.12
Above 35	.25	-.13	-.06	.52	.12

Table 4. Standardized path coefficients (β). All bolded coefficients are significant at $p < .05$, whereas unbolded coefficients are non-significant.

individual path coefficients (β) and their corresponding level of significance (p) obtained from our models are summarized in Table 4.

4.2. Results Analysis

We present here an analysis of the model results in Table 4 for the five factors for the general population and across the younger and older age groups.

4.2.1. Perceived Benefits

An individual's perception about the benefits of secondary user information sharing is expected to have a positive influence on users' attitude towards secondary information sharing. As shown in Table 4, our results confirmed that the factor – *perceived benefits* of secondary sharing – has a significant influence on users' attitude towards limiting information sharing. However, when viewed across age groups, it was a very significant factor for the older (above 35 years) group and non-significant for the younger (less than 25 years) group. This means that the older populations are more attracted by benefits than the younger population. In addition, when participants were asked which of the benefits of secondary user information sharing appealed to them the most, participants seemed to be more motivated by personal benefits than by altruistic reasons. This is in line with a prior study conducted in 2012 by IIC (2012). In our study, a total of 439 (53.4%) are motivated by the convenience of not having to repeat the same information across applications and services they use; 381 (46.4%) for better personalized services; 307 (37.3%) for monetary compensation; while 276 (33.6%) for safety and other public good purposes.

4.2.2. Perceived Risk

It is expected that as an individual's perception of risk increases, their likelihood of allowing sharing and reuse of information decreases. Surprisingly, the results of our model show that perceived risk has no significant influence on the decision of the general user population to allow sharing of their information. However, when analyzed across the different age groups, we found perceived risk to have significant negative influence on the decision of older age group, while it does not have a significant influence on the decision of the younger population. This means that the older population attaches more importance to the risks involved in the sharing and use of their information than the younger age groups.

4.2.3. Control

We hypothesize that giving the user control over which application to share their data with, which data to be shared, for what purpose, and for how long, should increase their tendency to share information. Our model results confirm that control has a significant influence on the general user population's attitude However, when compared to perceived benefits and gaining reputation, control is not as significant. Across the different age demographics, control is the most significant factor on the younger age groups' attitude, while for the older age groups, it is not a significant factor. One possible reason why control is less important to the older group compared to the younger group may be the fact that it takes substantial time and efforts to exercise this control (Brandimarte et al., 2012).

The younger group is expectedly more tech-savvy and, therefore, feels more comfortable in using the privacy tools for setting policies and preferences than the older groups.

4.2.4. Gain Reputation

Even though not explicitly investigated in previous studies, it seems intuitive that the desire to gain reputation should have some positive influence on the users' information sharing behavior. Surprisingly, our model results in Table 4 show that gaining reputation is the most important factor of all the five factors influencing users' attitude in the general population. Similarly, across the different age demographics, gaining reputation was also found to be very significant for the two different age groups (younger and older). Yet, the results show that it is more important to the older group than the younger age group.

4.2.5. Protect Reputation

An attempt to protect reputation may result in either *increased information sharing* (if the user believes he/she has a reputation and feels that by sharing he or she will maintain their status within a group or community) or in *decreased sharing* (if the user feels that what is to be shared may negatively impact on their reputation in a group or community). The result from our model in Table 4 shows that protecting reputation has a positive influence on user attitude for the general population. That is, the user is interested in sharing more in order to protect his/her online information. Similar outcome was observed across the age groups; protection of reputation has equal level of significance for the younger and older age groups.

5. Discussion

The results of this study gave us some interesting insights into the users' attitude and their expectations with respect to privacy in secondary context. In this section we reflect on these results and describe their implications for policy makers as well as design guidelines for privacy-enhancing technology for secondary context.

5.1. Users are not Principally Opposed to Secondary User Information Sharing

While many regulators forbid secondary use of data, except when it is known at the point of collection or with the explicit consent from the user, our study reveals that the majority of users are not principally opposed to secondary sharing and use of their data if the transaction is transparent and they are aware and able to control with whom their data is shared. Lack of transparency has eroded users' trust in the system and the service provider. Hence, to ensure users' trust and guarantee innovative flow of data for beneficial secondary purposes, the users must not be in the dark with regard to what is going on with their data and for what new purposes it would be shared.

5.2. Provide an Intuitive Means for User to Control Secondary Sharing of Their Data

There is the need to provide a transparent and intuitive means (e.g. through a web or mobile interface) for users to control what part of their data is shared for secondary use and for what purpose. Majority of our respondents' desire to have control of how their data is shared. In addition, since the need for being in control is more important to the younger age group (below 25 years) than to the older (above 35 years) group, designers can personalize the privacy control interface depending on the age of the user by ensuring that it is very conspicuous to the younger age group where to set the preferences for sharing data with third-party applications, and they should be allowed full control of the settings with more open defaults. In contrast, for the older age group, a default policy setting tailored to their needs of mitigating risk is required. As older adults are not too keen on being in control, they are likely to keep the default settings, so they need more conservative default setting Furthermore, since the older age group care more about reputation, the defaults for sharing social data (e.g. shared links, photos, status updates), could be less restrictive than sharing sensor, location data, or specific application data.

5.3. Display Appropriate Benefits to Motivate People to Share Data for Secondary Purposes

Our study also shows that perceived benefits from secondary sharing of information outweigh the perceived risks with most participants. Therefore, the privacy enhancing technology should highlight the benefits of the shared information to the user so that user can be motivated to allow sharing for useful purposes. For example, when sharing user data for financial benefits in form of discount prices, the system can display a message such as: *"Sharing this information will ensure that we find the best shopping deal for you within your location"*.

5.4. Provide Means for Adequate Security

Any discussion about privacy almost always leads to security considerations (Langheinrich, 2001). Not surprisingly, as the growing concerns for privacy increases, so also does concern about the security of data. Users are generally concerned that there might be no adequate security protection for their data kept by the service provider. This concern relates to protection from hackers during storage and also when it is being transmitted to third parties. As security of user data, particularly during storage, becomes a major issue, one solution to addressing this challenge is moving from a centralized to a decentralized approach to user modeling (Iyilade and Vassileva, 2013b), where user data fragments are kept by the various applications that do the primary user data collection and only what is needed for specific adaptation purpose is shared with other applications. In this approach, when the user data storage is hacked, only a portion of user information is accessible to the hackers. Thus, decentralized storage of user data fragments provide less attractive target for hackers.

5.5. One-size-does-not-fit-all

As pointed out earlier, there is a noticeable variation in the impact of the factors investigated across different age groups. For example: while risk is a major concern to older age group, it is not a major concern to the younger age group. Also, while having control is of significant influence on the younger age group, it is not significant for the older age group. Hence, one-size-does-not-fit-all and privacy solutions need to be tailored to preset defaults for different age demographics based on factors that are of interest to each group e.g. for the older age group where benefits are very significant, providing a clear message on compensation and other benefits will be helpful.

6. Conclusion

Secondary user data sharing and usage have become a major challenge to privacy as we move to a connected world of millions of mobile and ubiquitous devices that gather enormous about the user in various contexts. Understanding factors influencing users' attitude to allowing secondary sharing and use of their data is important to designing solutions that meet the privacy expectations of the user. This paper presents a study of the following five factors: *(i) perceived benefit; (ii) perceived risk; (iii) control (iv)gain reputation, and (v) protect reputation*; and their relative influence on users' attitude towards allowing secondary sharing and use of online information. Using structural models, we determine the relationship between the factors and the users' attitude toward secondary information sharing. In addition, we explore the correlation of these factors for different age groups.

The results from our study show that most users desire to have control over how their data is shared and with whom it is shared. The main motivations to share information are to reap benefits in terms of personalization, saving time of entering user data and to increase and protect their reputation. Also, we found that risk is not a very significant factor for secondary sharing of user data. Yet the study results revealed significant differences in the relative importance of the above listed factors to users of different age groups. This finding suggests that it is possible to define default user privacy solutions for users of different age groups that are tailored to their main privacy concerns and motivations. This would allow an easier and more efficient process for personalizing privacy policies for individual users.

Still, it can be argued that the demographic of our participants is not a statistically diverse representative sample of global Internet users. Our study has at least one participant from about each of 43 countries, however, majority (about 90%) of respondents are from the USA and India. It would have been desirable to have a large sample of participants from countries such as China and from Europe, with different privacy laws, culture and social norms. Nonetheless, our respondents are heavy Internet users and mostly younger populations who have at least an undergraduate degree, and are enthusiastic about new technologies. As such, we believe that this demographic sample is important for understanding the future Internet user population. As future work, we plan to investigate our data to see if it is possible to find further interesting insights from it, for example, with respect to the influence of gender or cultural background on privacy attitudes and decisions.

References

[1] F. ABEL, N. HENZE, E. HERDER, D. KRAUSE, Linkage, Aggregation, Alignment and Enrichment of Public User Profiles with Mypes. In *Proc. of 6th Int. Conf. on Semantic Systems (I-SEMANTICS)*, Graz, Austria, Article No. 11, (2010). ISBN: 978-1-4503-0014-8.

[2] A. ACQUISTI, L. JOHN, G. LOEWENSTEIN, What is privacy worth? In *Twenty First Workshop on Information Systems and Economics (WISE)*, (2009). Online at: `http://www.heinz.cmu.edu/~acquisti/papers/acquisti-privacy-worth.pdf`. Last accessed: 22-July-2013.

[3] A. ACQUISTI, R. GROSS, Imagined communities: Awareness, Information Sharing, and Privacy on the Facebook. *Lecture Notes in Computer Science*, (2006). 4258/2006:36|58.

[4] J. C. ANDERSON, G. W. GERBING, Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin*, **103** (1988), 411–423.

[5] C. APERJIS, B. A. HUBERMAN, A Market for Unbiased Private Data: Paying Individuals According to their Privacy Attitudes, (2012). Online at: `http://www.hpl.hp.com/research/scl/papers/data-market/datamarket.pdf`. Last accessed: January 10, 2013.

[6] R. M. BARON, D. A. KENNY The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, **51** (1986), 1173–1182.

[7] L. BRANDIMARTE, A. ACQUISTI, G. LOEWENSTEIN, Misplaced Confidences: Privacy and the Control Paradox. *Journal of Social Psychological and Personality Science*, August 2012, doi: 10.1177/1948550612455931. Online at: `http://www.heinz.cmu.edu/~acquisti/papers/acquisti-SPPS.pdf`.

[8] J. N. BROWN, Choosing the Right Type of Rotation in PCA and EFA. Shiken, JALT Testing & Evaluation SIG Newsletter, **13** (2009), 20–25.

[9] M. BUHRMESTER, T. D. KWANG, G. S., Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data. *Perspectives on Psychological Science*, 2011.

[10] F. CARMAGNOLA, F. CENA, C. GENA, User Model Interoperability: A survey. *User Model User-Adapted Interaction*, (2011), pp. 1–47, Springer, Netherland.

[11] W. W. CHIN, The Partial Least Squares Approach to Structural Equation Modeling, Modern Methods for Business Research, Lawrence Erlbaum Associates, Hillsdale, NJ, 1998.

[12] W. W. CHIN, Frequently Asked Questions – Partial Least Squares & PLS-Graph, 2013. Online at: `http://disc-nt.cba.uh.edu/chin/plsfaq.htm`. Last accessed: June 10, 2013.

[13] L. F. CRANOR, J. REAGLE, M. S. ACKERMAN, Beyond Concern: Understanding Net Users' Attitudes about Online Privacy. AT&T Labs Research Technical Report, 1999. Online at: `http://arxiv.org/html/cs/9904010/report.htm`. Last accessed: May 15, 2013.

[14] D. CVRCEK, M. KUMPOST, V. MATYAS, G. DANEZIS, A Study on the Value of Location Privacy. In *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society*, (2006) pp. 109–118. New York, NY, US.

[15] S. EGELMAN, J. TSAI, L. F. CRANOR, A. ACQUISTI, Timing is Everything: The Effects of Timing and Placement of Online Privacy Indicators. *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, (2009) pp. 319–328.

[16] C. FORNELL, D. F. LARCKER, Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, **18** (1981).

[17] J. F. HAIR, C. M. RINGLE, M. SARSTEDT PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, **19** (2011), 139–152.

[18] I. HANN, K. HUI, T. S. LEE, I. P. L. PNG, Online Information Privacy: Measuring the Cost-Benefit Trade-off. In *Proceedings of 23rd International Conference on Information Systems (ICIS 2002)*, (2002) pp. 1–10.

[19] D. HECKMANN, T. SCHWARTZ, B. BRANDHERM, A. KRÖNER, Decentralized user modeling with UserML and GUMO. In *Proceedings of theWorkshop on Decentralized, Agent Based and Social Approaches to User Modeling DASUM-05, at UM2005*, (P. DOLOG, J. VASSILEVA, eds.), July, (2005) pp. 61–66. Edinburgh, Scotland.

[20] J. HEER, M. BOSTOCK, Crowdsourcing Graphical Perception. *Proceedings of the 28th International Conference on Human Factors in Computing Systems – CHI '10*, (2010) pp. 203. ACM Press, New York, New York, USA.

[21] J. HENSELER, C. M. RINGLE, R. R. SINKOVICS, The Use of Partial Least Squares Path Modeling in International Marketing. *Advances in International Marketing*, **20** (2009), 277–319.

[22] S. HOLM, A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6 (1979), 65–70.

[23] IIC (2012). Personal Data Management: The User's Perspective. Research Technical Report. Online at: `http://www.iicom.org/open-access-resources/doc_details/226-personal-data-management-the-users-perspective`. Last accessed: May 10, 2013.

[24] J. IYILADE, J. VASSILEVA, A Framework for Privacy-Aware User Data Trading. In *Proceedings of User Modeling, Adaptation and Personalisation (UMAP) 2013*. June 10-14, (2013a) pp. 310–317. Rome, Italy.

[25] J. IYILADE, J. VASSILEVA, A Decentralized Architecture for Sharing and Reusing Lifelogs. In *Second Workshop on LifeLong User Modelling*, in conjunction with UMAP 2013. (2013b) pp. 4–10.

[26] B. P. KNIJNENBURG, A. KOBSA, Making Decisions about Privacy: Information Disclosure in Context-Aware Recommender Systems. *ACM Transactions on Interactive Intelligent Systems*, (2012). Online at: `http://www.usabart.nl/portfolio/paper-tiis2013.html`. Last accessed date: July 10, 2013.

[27] B. P. KNIJNENBURG, A. KOBSA, Helping Users with Information Disclosure Decisions: Potential for Adaptation. *Conference on Intelligent User Interfaces (IUI) 2013*, (2013). Available online at: `http://www.usabart.nl/portfolio/paper-iui2013.html`. Last accessed: July 10, 2013.

[28] P. KUMARAGURU, L. F. CRANOR, Privacy indexes: A Survey of Westin's Studies. *Technical Report*, CMU-ISRI-5-138, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 2005.

[29] E. KUPEK, Beyond logistic regression: structural equations modelling for binary variables and its application to investigating unobserved confounders. *BMC medical research methodology*, **6**(13), 2006.

[30] M. LANGHEINRICH, Privacy by Design – Principles of Privacy-aware Ubiquitous Systems. In *Ubicomp 2001: Ubiquitous Computing*, (2001) pp. 273–291. Springer Berlin Heidelberg.

[31] A. K. MACHANAVAJJHALA, Defining and Enforcing Privacy in Data Sharing. Unpublished PhD Thesis. Computer Science, Duke University, 2008.

[32] M. MADDEN, A. SMITH, Reputation Management and Social Media. Technical Report, Pew Internet & American Life Project, Washington, DC, 2010.

[33] W. MASON, S. SURI Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, **44** (2012), 1–23.

[34] OECD (1980). "Guidelines on the Protection of Privacy and Transborder Flows of Personal Data", Online at: `http://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm`. Last Accessed: July 10, 2013.

[35] R. ORJI, J. VASSILEVA & R. L. MANDRYK, Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Modeling and User-Adapted Interaction*, **24(5)**, 2014, 453–498.

[36] J. PODESTA, P. PRITZKER, E. MONIZ, J. HOLDREN, J. ZIENTS, Big Data: Seizing Opportunities, Preserving Values. A White House Special Report, (2014). Online at: `www.whitehouse.gov/sites/.../big_data_privacy_report_may_1_2014.pdf`. Last accessed: 20th May, 2014.

[37] S. POSLAD, Ubiquitous Computing: Smart Devices, Environments and Interactions. John Wiley & Sons Publishing, 2009. ISBN: 978-0-470-03560-3.

[38] C. M. RINGLE, S. WENDE, J. BECKER, Smartpls – Next Generation Path Modeling, 2012. Online at: `http://www.smartpls.de/forum /release.php`. Last Accessed: June 21, 2013.

[39] L. SENG, Context-aware Pervasive Systems: Architectures for a New Breed of Applications. Auerbach Publications, 2007. ISBN: 0-8493-7255-0.

[40] A. J. SETTERSTROM, J. M. PEARSON, H. ALEASSA, An Exploratory Examination of Antecedents to Software Piracy: A Cross-Cultural Comparison. 2012 *45th Hawaii International Conference on System Sciences*, (2012) pp. 5083–5092.

[41] S. SPIKERMANN, F. C. LORRIE, Engineering Privacy. *IEEE Transactions on Software Engineering*, **35**(1), 2009, 67–82.

[42] O. TENE, J. POLONETSKY, Privacy in the Age of Big Data: A Time for Big Decisions, 2012. 64 STAN. L. REV. ONLINE 63. Online at: `http://www.stanfordlawreview.org/online/privacy-paradox/big-data`. Last accessed: 20 July, 2013.

[43] R. THALER (1980). Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization*, **1**, 39–60.

[44] H. TINSLEY, D. TINSLEY Uses of Factor Analysis in Counseling Psychology Research. *Journal of Counselling Psychology*, **34** (1987), 414–424.

[45] E. TOCH, Y. WANG, L. F. CRANOR, Personalization and Privacy: A Survey of Privacy Risks and Remedies in Personalization-based Systems. *User Model User-Adapted Interactions*, **22** (2012), 203–220.

[46] J. Y. TSAI, S. EGELMAN, L. CRANOR, A. ACQUISTI, The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Journal of Information Systems Research*, **22**(2) (2011), 254–291.

[47] V. TUUNAINEN, O. PITKÄNEN, M. HOVI (2009). Users' Awareness of Privacy on Online Social Networking Sites – Case Facebook. *BLED 2009 Proceedings*, (2009) Paper 42. http://aisel.aisnet.org/bled2009/42

[48] J. VASSILEVA, G. MCCALLA, J. GREER, Multi-Agent Multi-User Modeling. User Modeling and User-Adapted Interaction. *Special Issue on User Modeling and Intelligent Agents*, **13**(1) (2003), 179–210.

[49] K. T. VIRPI, P. OLLI, H. MARJAANA, "Users' Awareness of Privacy on Online Social Networking Sites – Case Facebook" *BLED 2009 Proceedings*, (2009) Paper 42. Online at: http://aisel.aisnet.org/bled2009/42. Last Accessed: June 5, 2013.

[50] WEF (2012). Unlocking the Economic Value of Personal Data: Balancing Growth and Protection. World Economic Forum Report. Online at: http://www3.weforum.org/docs/WEF_IT_UnlockingValueData_BalancingGrowthProtection_SessionSummary.pdf. Last Accessed Accessed: 20 June 2013.

Received:
Accepted:

Contact address:
Johnson Iyilade
Department of Computer Science
University of Saskatchewan
Canada

Rita Orji
Department of Computer Science
University of Saskatchewan
Canada

Julita Vassileva
Department of Computer Science
University of Saskatchewan
Canada

JOHNSON IYILADE is a PhD candidate at the University of Saskatchewan, Canada, he received a PhD in Computer Science specializing in Web Service Provisioning from the University of Zululand, South Africa. He also holds M.Sc. in Computer Science and B.Sc. in Computer Science with Economics from Obafemi Awolowo University, Nigeria. His research interests include: Big Data analytics, data privacy and security, user modeling and personalization, and cloud services provisioning.

RITA ORJI is a visiting Senior Lecturer in the Computer Science Department at the Nnamdi Azikiwe University, Nigeria. She completed her Ph.D in Computer Science at the University of Saskatchewan, Canada. She holds a B.Sc. in Computer Science and a M.Sc. in Information Systems from Nnamdi Azikiwe University, Nigeria and Middle East Technical University, Turkey respectively. Her research examines how technological interventions can be designed to motivate desirable behavior change. She is particularly passionate about studying how behavior change support systems and persuasive technologies can be designed to help people move towards improved health and wellness. For the past five years, she has been researching various ways of tailoring/personalizing persuasive technological interventions to increase their relevance and effectiveness.

JULITA VASSILEVA is a professor of computer science at the University of Saskatchewan, Canada. Her research areas involve human issues in decentralized software environments: user modeling and personalization, persuasion technologies for behaviour change, designing mechanisms for encouraging participation and facilitating trust in decentralized software applications, such as online communities, social networks, multi-agent systems and peer-to-peer systems.