International Journal of Information Security and Privacy
Volume 10 • Issue 2 • April-June 2016

Using Crowd Sourcing to Analyze Consumers' Response to Privacy Policies of Online Social Network and Financial Institutions at Micro Level

Shaikha Alduaij, University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA
Zhiyuan Chen, University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA
Aryya Gangopadhyay, University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA

ABSTRACT

As it becomes easy and inexpensive to store huge amount of data, concerns about privacy are increasing as well. Although service providers have privacy policies, research shows that users rarely read privacy policies. As a result, there has been little work done on how consumers respond to individual segments of privacy policies, which is important for organizations when designing privacy policies. In this study, the authors break down privacy policies of two well-known social network companies (Facebook, Twitter) and financial institution (Bank of America) into simple segments. They then use crowd sourcing to analyze consumers' response to these policy segments. The authors ask questions on users' awareness, expectations, familiarity, and privacy concerns of these policy segments. The relationships between various factors such as demographic factors, data type, data flow and consumers' privacy concerns were also investigated. The authors conclude with guidelines and suggestions for improvement and ways to increase users' awareness of privacy policies.

KEYWORDS

Awareness, Concerns, Crowdsourcing, Privacy Expectations, Privacy Policy

INTRODUCTION

The use of online services has become a necessity because it is involved in most individuals' daily activities, including those related to business, education, and communication. When using these services, users usually share their information for purposes such as registering a service, customizing their experience, or sharing their thoughts and interests with others. The collection and storage of this vast amount of information has raised users' concerns about how these practices will affect their information privacy. Natural questions arise in this context, such as how individuals' information will be stored, who will access it, how it will be used, and for what purposes.

To allay users' concerns, most of the service providers explain their practices with privacy policies. A privacy policy is a statement or legal document provided by the service provider to explain the handling of the users' gathered information by describing what information will be collected, how it will be used, with whom it will be shared, and the purpose of that sharing. Furthermore, it describes users' rights and options to change some of the practices.

It is important that a privacy policy is written and presented to user's clearly so they can understand how their information is being used. For example, Thelma Arnold, a user of the AOL search engine, did not know that AOL stored and shared her queries. Her identity was disclosed to the public by

DOI: 10.4018/IJISP.2016040104

Copyright © 2016, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

inferring her identity based on her queries. She stated in an interview, "My goodness, it's my whole personal life . . . I had no idea somebody was looking over my shoulder" (Barbaro & Zeller, 2006). Her unawareness of this practice could be because it was not clearly described in the AOL privacy policy or because she did not read the policy.

Making them clear and easy to understand would increase their readability and thus users' awareness of privacy practices. Some users accept the terms and conditions before using a service, and they are not aware of the risks that may be associated with the sharing of their personal data. The information users post or share online may be misused or used for purposes they are not aware of intentionally or accidentally. The risks or negative effect of using individual personal information can lead to embarrassment, decreased opportunity of employment, identity theft, cyber stalking, or phishing. These risks can happen by using information users themselves share without knowing how badly it can be misused. For example, some users may share on social media a photo for a ticket of an event they plan to attend. Any person who is allowed to view the picture may copy the barcode, print it, and use it (Ehling, 2013). Risks can also happen by using information stored in servers or shared with a third party. For example, an individual's identity may be associated with a disease when the information he searched for in a website is shared with third parties, disclosed to the public, or given to a data broker. This might affect users by loss of employment (Libert, 2015; Walters & Betz, 2012). Indeed, the availability of an individual's information can even lead to murder. Amy Boyer, a 20-year-old women from Nashua, was murdered after the criminal stalked her and gathered information about her work location using online sources (Donovan & Bernier, 2008).

Several research studies have tried to propose different solutions to improve privacy policies and increase users' privacy awareness. Some have introduced tools that help in privacy policy development and enforcement; others have introduced standardized presentations of privacy policies.

Research studies show that people think the privacy of their information is an important issue and they are concerned about it, but at the same time, they do not fully read the provided privacy policies for the services they use (Milne & Culnan, 2004). Some users believe that they do not have control over their information so they do not want to waste their time reading privacy policies if they are going to use the service anyway.

Recent research studies also found that privacy policies are difficult to understand and very time consuming to read (McDonald, Reeder, Kelley, & Faith, 2009). The results of the Consumer Action "do not track" 2013 survey showed that consumers are mostly unaware of several privacy issues due to the confusing nature of the privacy policies. For example, one study within this survey showed that 29% of the participants were not aware that their location data could be tracked by mobile phones. Moreover, 49% of the participants were unaware that is legal for a company to track their online activity. The extensive time required to read a privacy policy leads many people to avoid reading them. Studies estimated that if users read every unique service provider's privacy policy it would take them approximately 244 hours per year. Moreover, this lost time would cost approximately $781 billion per year (McDonald & Cranor, 2008).

There has been relatively little research studied how consumers respond to individual segments of privacy policies (e.g., whether they are comfortable with a specific segment of privacy policy), which is important for organizations to design better privacy policies. In this study, we used crowd sourcing to study consumers' responses to individual segments of the privacy policies of Facebook and Twitter, two well-known social network companies, and Bank of America, a major financial institution in the United States of America. The key contributions of this paper are:

- A model to break down each policy segment into an easy-to- understand format.
- An investigation of the relationship between various factors (including demographic factors and factors related to the privacy policy segments such as data type, data flow, user controls, as well as consumers' awareness and expectation) and consumers' privacy concerns.
- Guidelines and suggestions for privacy policy improvement and ways to increase users' awareness.

The rest of the paper is organized as follows: we first discuss recent work and then introduce the breaking down of privacy policies. We then describe a set of surveys conducted to study the response of consumers to segments of privacy policies of Facebook, Twitter, and Bank of America. We present the study's research questions and the study design and then describe the results of our surveys and discuss the results. Based on the results of our research, we propose guidelines and suggestion for privacy policies improvements and finally conclude the paper with a summary of our findings.

RELATED WORK

The related work can be divided into several categories: 1) research on limitations of privacy policies; 2) research on privacy concerns; 3) research on risks of disclosure of private information; 4) research on factors affecting users' privacy concerns. We discuss each category of related work.

LIMITATIONS OF PRIVACY POLICIES

Although privacy policies are meant to explain to Internet users everything related to the collection and sharing of their personal information when using the service of any individual provider, existing studies have found that these policies have many negative issues. To begin, the majority of privacy policies are not easily accessible, comprehensive, or manageable for the average user. They often fail to meet users' actual privacy concerns, and improvements have been arguably few and quite limited.

First, privacy policies are quite difficult to understand. McDonald et al. (2009) performed a comparative study of the formats of six different companies' privacy policies and found that users considered every studied format difficult to understand. Pollach (2007) discovered that privacy policies are written in unclear and misleading linguistic patterns. Additional studies have found that privacy policies require a high level of education to be fully understood. Jensen and Potts (2004) found that people without a high school education can only understand 6% of privacy policies, and according to Sumeeth, Singh, and Miller (2010), 20% of the policies they analyzed require a postgraduate education to be understood. Moreover, privacy policies often contain long sentences and domain-specific terms unfamiliar to typical users, making them difficult to read and comprehend (Jensen & Potts). Additionally, policies are often written in an unclear and confusing manner (McDonald et al.). For example, a company states in its online policy, "While Company does not currently support telemarketing, it is possible that in the future Company properties may contact you by voice telephone," to explain that that they may use information for telemarketing.

The second issue is that it takes a lot of time to read privacy policies. McDonald and Cranor (2008) indicated that privacy policies are often quite long. The average length is about 2,500 words, which is equivalent to a paper of five to six pages written in a 12-point font. The authors asked, "If website users were to read the privacy policy for each site they visit just once a year, what would their time be worth?" They calculated the average time of reading privacy policies in two ways. First, they analyzed the word count of the privacy policies for the 75 most frequently visited websites according to AOL search data. Second, they conducted a survey and calculated the average time spent by 212 participants to skim these privacy policies. Their results showed that it takes an individual 244 hours per year to read and 154 hours per year to skim every unique privacy policy for websites visited. Additionally, by estimating the value of time as 25% of average hourly salary for leisure and twice wages for time at work, the results indicate that the national cost to read these privacy policies would be $781 billion per year, and the cost to skim these policies would be $492 billion per year. Jensen and Potts (2004) added that the results of a survey done in a university setting for a stand-alone website requiring registration showed that due to its length very few participants visited the privacy policy page, specifically, only in 0.24% of 55,158 sessions was the policy viewed.

The third issue is that privacy policies do not consider consumers' common concerns about their information collection, transfer, and storage process and overall service provider's practices.

For example, Earp et al. (2005) found that privacy policies of 50 websites did not emphasize issues commonly raised in Internet users' privacy concerns such as revealing information about their identity or information related to their activities. Pollach (2007) also found that privacy policies poorly cover users' privacy concerns regarding data collection, data storage, data sharing, and unsolicited marketing communications. He conducted a survey to verify the coverage of these concerns in the privacy policies and found that 39.4% of the questions about these practices could not be answered due to the lack of sufficient information provided by the examined policies.

The fourth issue is low adoption of proposed solutions. A number of studies have introduced approaches and tools for privacy policy improvements. For example, P3P is a platform for privacy preferences introduced by the World Wide Web Consortium (W3C) to enable browsers to read a website's privacy policy and match it to users' predefined preferences. However, a large number of websites had errors when they adapted the P3P policy format (Leon, Cranor, McDonald, & McGuire, 2010). Privacy trustable seals is another suggested improvement for privacy policies. This program reviews and evaluates a service provider's policy, matches it to the website's privacy practices, and displays a trust mark on the website informing users of the evaluated privacy practices. This helps users save time by only requiring them to verify the seals displayed on the website instead of reading the full policy. Nevertheless, the results of Moore's (2005) study showed that users found it hard to recognize the graphic seals and distinguish between actual and fake seals, concluding that consumers may simply be responding to a graphic design, rather than to any attempt by the website to be certified as trustworthy. Additionally, only 34.3% of Pollach's (2007) survey questions relating to data handling of providers with at least one trust seal were answered correctly, which indicates that the coverage of privacy issues on privacy policies is another limitation of privacy trustable seals.

Our study suggests to alleviate some of the limitations, including the difficulty of understanding a policy and the long time needed to read a policy.

PRIVACY CONCERNS

Smith, Dinev, and Xu (2011) found that the efficiency of information storage and usage by service providers raised privacy concerns. Dinev and Hart (2004) showed that privacy concerns and a low willingness to provide personal information are caused by the conflict of the need for information to be collected and the resulting threat to users' privacy. Westin (2001) reported that 90% of Americans are worried that their personal information will be misused, and specifically 77% describe themselves as "very concerned."

Researchers have different findings when it comes to privacy controls. The survey Consumer action "Do not track" (2013) results showed that 87% of their participants strongly agreed that they have the right to make choices regarding their information privacy and want more control. According to Hazari & Brown (2013), the employed users of social networks websites are more likely to control their privacy preferences than the unemployed users because it can have a negative influence on their employment status and relationships. Young and Quan-Haase (2009) found that students made some effort to control their personal information in their profiles to protect themselves. For example, they provide limited information in their profiles to prevent other users from gaining comprehensive information about them. Moreover, Harris Interactive (2001) reported that almost 80% of Americans think that it is "very important" to control collected information. Although existing research showed that people need control over their information privacy and think this control is important, other research found that people provide much information without using the available controls or reading the privacy statements. Gross and Acquisti (2005) found that only 1.2% and 0.06% of their study participants changed default profile searchability and visibility, respectively. Hazari and Brown (2013) stated that users post a great deal of content to customize their experience using different services in a way that negatively affected their privacy. Further, some studies have shown that users do not read privacy notices or policies provided by the web services that they use (Chaianuchittrakul, 2013;

Lin et al., 2012). In this study we will investigate the provision of controls for each privacy policy segment, whether or not users make use of the provided controls, and whether using these controls lowers users' privacy concerns.

Although existing studies have specified users' privacy concerns, there has been relatively little work done on how consumer respond to the policies' individual segments such as sharing profile information with a third party for marketing purposes. In this study we use a crowd sourcing approach to analyze consumer responses to individual privacy policy segments. We overcome the readability issue of privacy policies by conducting the study at a micro level (i.e., segment level) because it is much easier for consumers to understand an individual policy segment than the full text of the policy.

FACTORS AFFECTING USERS' PRIVACY CONCERNS

There are multiple factors that affect an individual's privacy concerns and practices. Lin et al. (2012) found that users of mobile applications feel more comfortable and less concerned when they are notified of the reasons for which their information is collected by the application. For example, participants felt less concerned when they were informed that the application Dictionary collected their location for the purpose of providing words trending in their area.

The impact of demographic factors such as age and gender on individual concerns has been the focus of several studies with some different findings. In a study about the impact of gender on online privacy concerns for undergraduate students at a Southwestern American university, Yao et al. (2007) found that online concerns did not vary between gender groups. Nowak and Phelps's (1992) results also supported this finding, stating that users' concerns about threats to personal privacy did not correlate with gender. However, Janda and Fair (2004) found in their study of non-student adult Internet users that gender had a significant impact on privacy concerns showing that women have greater privacy concerns including fraud, privacy, security, hacking, and child protection. Their results also strongly indicated that older adults are more concerned about their online privacy. Tufekci (2008) stated that both age and gender have an impact on undergraduate student concerns and disclosure behavior when using social networking sites. The results showed that younger users tended to display more information and that there were significant gender differences based on the information displayed. For example, men tended to display their phone numbers in their account profiles more than women, while women tended to disclose information related to their religion and personal preferences such as favorite music and books more than men.

In our survey we analyzed factors affecting users' privacy concerns including demographics information, the service provider's service nature, and the type of information.

RISKS OF INFORMATION BREACHES

In this digitalized age people are using online services on a daily basis. They may use it for education purposes, travel reservations, social networking, doctors' appointments managements, and many other daily activities. Using these services involves the gathering of users' personal information including identity, financial, and health information. The availability of this valuable information improves users' vulnerability of privacy threads. Following are some examples of the risks users may face when their information is incidentally or intentionally disclosed.

Violation of any kind of private information will easily bring embarrassment. Publishing personal information relating to specific activities can even lead to more serious circumstances. The Ashley Madison hack captured public attention recently. A group of hackers stole personal information from the adult dating website, Ashley Madison, and they disclosed more than 30 million users' information including full names, street addresses, and e-mail addresses. Although the company announced that the stolen data do not contain financial information, subscribers were concerned that their reputations,

marriages, jobs, and lives could be at stake. The case got even worse when two suicides were linked to that information leak.

Publishing personal information by an individual, especially on social network websites, can decrease her opportunity to get a job (CareerBuilder.com, 2009). The results of the CareerBuilder survey reported that 45% of employers used social networking websites to assess job candidates. The results showed that employers eliminated the job's candidates for some of the following reasons:

1. Candidate posted provocative or inappropriate photographs or information.
2. Candidate posted content about their drinking or using drugs.
3. Candidate bad-mouthed their previous employer, coworkers, or clients.
4. Candidate showed poor communication skills.
5. Candidate made discriminatory comments.
6. Candidate lied about qualifications.

The availability of an individual's data in several websites severs could make her information vulnerable to identity theft. Gaining access to one source of information and linking it to other sources would help to get sufficient data about an individual. Identity theft involves gaining financial or medical advantages by using another person's name or identity. A survey report of identity theft in the United States showed that in 2005 consumers lost about $56 billion because of companies' data breaches (Javelin Strategy & Research, 2006). In addition to financial identity theft, a report by the Federal Trade Commission (2010) stated that from 2001–2006 more than 250,000 identity theft cases counted as medical identity theft. The victim of medical identity theft may be negatively affected by the falsifying of his medical history, which may lead to inappropriate medical diagnoses, loss of employment, and emotional consequences (Betz, 2012; Cullen, 2007; Federal Trade Commission, 2010; Identity Theft Resource Center, 2007; Schmidt & McCoy, 2008).

Cyber stalking is another issue on data breaches. The use of personal information for stalking an individual can lead to serious consequences including murder. The murder of Asia McGowan is an example of a cyber-stalking crime. The media reported that the killer, Anthony Powell, stalked the victim on Facebook and YouTube before he killed her (Reisman, 2009).

Phishing is a method to gain the victim's personal data usernames, passwords, and credit card details in the form of a "SPAM" attack. The information is usually obtained from users through electronic communication. An attacker with previous knowledge about an individual can easily gain additional information and use it for malicious reasons (Reilly, 2006). A survey of 1,335 US.net users reported that in 2004 US consumers lost about $500 million as a result of phishing scam attacks (Leyden, 2004).

Finally, users' health interests or health searches can be used by third parties and have a negative impact. First, it may lead to personal identification by associating an individual's name to a disease. This can happen when a person searches for a disease, symptom, or treatment on a website that shares user's queries with a third party. An attacker getting access to this information, incident leakage, or the selling of information by brokers can reveal the association between the user and a disease to the public. Second, this can lead to blind discrimination where the user's identity is not necessary disclosed, but she may be treated differently. For example, marketing companies may exclude an individual from receiving some offers based on perceived medical conditions (Libert, 2015).

Our study can help users understand and read policies which may lead users to make intelligent privacy choices to lower their privacy risks.

CROWDSOURCING

Amazon Mechanical Turk (AMT) is an online crowdsourcing service that allows a requester to recruit workers to perform specific tasks for a certain wage or "reward." There are many advantages of conducting online studies in general. However, AMT has the additional benefit of being chosen for research studies. Workers can look up tasks in which they are interested, and they can access them at any time and from anywhere. Also, the researcher (requester) gets the benefit of having access to diverse participants, as AMT workers tend to represent diverse backgrounds, languages, ethnicities, ages, and socioeconomic statuses. An additional advantage of AMT is the low cost at which studies can be conducted compared to using other online recruitment methods or paid laboratory subjects (Mason & Suri, 2012). For example, Göritz, Wolff, and Goldstein (2008) showed that response rates to online experiments can be low because of the inconvenience participants often face when a site uses a third-party payment mechanism. However, AMT has its own mechanism to pay workers. Moreover, AMT offers several mechanisms to insure response reliability. For example, it allows the requester to reject a specific task and not to pay a worker if the task is not performed as requested. Also, AMT keeps track of workers' rejected tasks, which gives them the impetus to do their best so that they can maintain their reputations. Qualifications are another mechanism that AMT implements to ensure reliable results. In some cases, the requester requires that he or she meet the workers in person before tasks are assigned and completed (Mason & Suri, 2012). Finally, with AMT, responses are not subject to experimental bias. This is because workers perform assigned tasks without interacting with experimenters, and because workers are usually selected based on their qualifications instead of their race or gender (Adams, 2013).

BREAKING DOWN PRIVACY POLICIES

Because privacy policies are usually written in a lengthy and unclear format, we first broke down privacy policies into easy-to-understand policy segments. We created these segments by analyzing the privacy policy text of different service providers represented in a table format to make it easier for us to design our study. The privacy policy segment is a collection of information that describes what information was collected or shared, the parties involved in the collection or sharing practice, the purpose of that practice, and the control over that practice.

A segment is divided into five fields: data, from, to, purpose, and user control. The "data" field includes the information collected, stored, or shared by the service provider. The "from" field describes the information provider such as other users, service providers, and third parties from whom information is collected. The "to" field includes the party who is receiving the collected or shared information including other users, service providers, and third parties. The "purpose" field describes the reasons for collecting, storing, or sharing stated information. Finally, the "user control" field states whether or not the user has the ability to control the collecting, storing, or sharing of the specified information.

To illustrate clearly the segments and their fields, Table 1 shows how we converted Bank of America's (BoA) privacy policy into segments with each segment allotted its own line. The first segment shows that personal information is collected from the customer and then passed to BoA for a purpose such as opening an account or performing transactions, applying for a loan or using a credit card, and seeking advice about investments. The table very clearly shows that users have no control over the sharing or collecting of these data.

Using the information in the table we categorized the values of the "data" field of all the privacy policies into four categories: identity information such as name and credit card account, activities information such as user's interaction with links and messages he sends, log information such as IP address and pages visited, and cookies. We also generalized the values of the "from" and "to" fields into three categories: users, service providers, and third parties. Based on this categorization

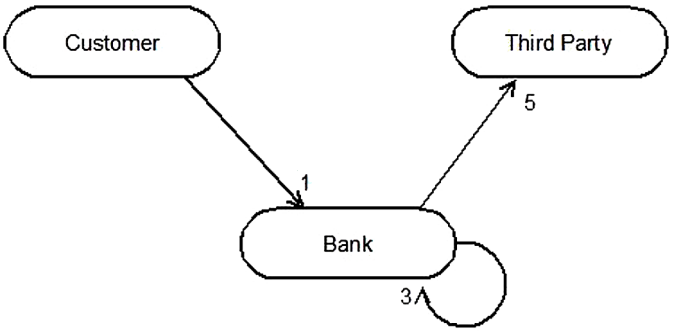
Table 1. Policy segments from Bank of America

	Data	From	To	Purpose	User Control
1	Personal information	Customer	Bank	open an account or perform transactions, apply for a loan or use your credit or debit card, seek advice about your investments	NO
2	Personal information	Bank	Bank	process your transactions, maintain your account(s)	NO
3	Personal information	Bank	Legal request	respond to court orders and legal investigations, or report to credit bureaus	NO
4	Personal information	Bank	Service providers	offer our products and services for customer	NO
5	Personal information	Bank	Financial companies	for joint marketing	NO
6	Information about transactions and experiences	Bank	Affiliates	for everyday business purposes	NO
7	Information about creditworthiness	Bank	Affiliates	for everyday business purposes	YES
8	Credit card accounts	Bank	Nonaffiliates	to market the customer	YES
9	Accounts and services endorsed by another organization	Bank	Nonaffiliates	to market the customer	YES

we created a data flow diagram where each generalized party category forms an element including user, service provider, and third party.

We then represented policy segments using a data flow graph. Each node shows a generalized "from" or "to" field value. Each arrow shows a policy segment where data are passed from a "from" value to a "to" value. The number next to the arrow represents the number of such segments. Figure 1 shows the flow diagram based on the contents of Table 1.

Figure 1. Data flow graph for Bank of America



RESEARCH QUESTIONS

Our study survey served to answer the following questions:

1. Does users' demographic information affect their privacy concerns related to information collection and sharing?
2. Do users' frequency of usage and awareness or understanding of the service provider's privacy policy affect their privacy concerns?
3. Do users prefer additional control over the sharing and collecting of their data?
4. Are users aware of provided controls? Do they use them?
5. Which data flow or data type is most unexpected by users, and with which data flow or data type are users most uncomfortable?
6. Do users have difficulty identifying the reasons service providers collect or share certain types of information?
7. Does clarifying the purpose of collection or sharing information ease users' concerns about the privacy of their data?
8. Is it possible to predict the level of privacy concerns for a particular user?
9. What other factors affect users' privacy concerns?

STUDY DESIGN

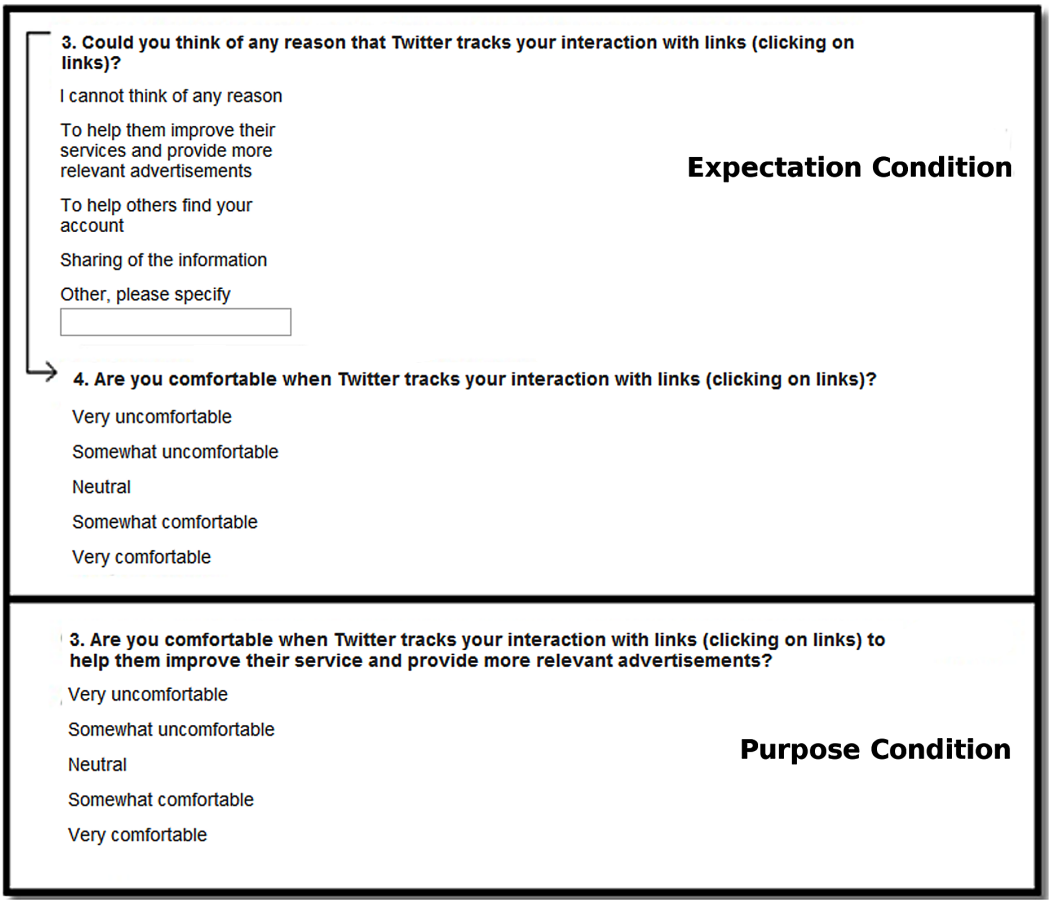
We selected privacy policies from three service providers: Facebook, Twitter, and BoA. We selected Facebook and Twitter based on their rating as the top two social networking websites according to Ebizmba.com. BoA was selected to represent financial institutions because it is rated one of the largest banks in the United States according to Relbanks.com.

We designed the surveys to gather information related to all of the segment fields. We used Quicksurveys.com to design six surveys, two survey versions for each of the three providers. The two survey versions are identical in terms of the sections and questions except that one version explains the purpose of the data collection to the participants while the other does not. Instead, in the other survey version, participants were asked to identify the reason for which their data were being collected or shared. Comparing the results of these three pairs of surveys allowed us to test the hypothesis that informing users of the purpose of data collection and sharing practices eases their privacy concerns.

The surveys were divided into the following sections:

- The first section gathers the participant's demographic information.
- The second section verifies that the participant has an account with the specific service provider and asks how frequently the participant accesses and uses the account.
- The third section assesses the participants' awareness or understanding of providers' privacy policy, to which they agreed before creating their accounts.
- The fourth section has two different sets of questions in each version of the survey. The first set of questions, called the "purpose condition," asks participants to indicate their level of comfort knowing that the service provider is collecting or sharing their information for a specified reason. For instance, Twitter tracks users' interactions with links to help them improve their service and deliver relevant advertisements. The second set of questions, called the "expectation condition," asks the participants to identify the reason a service provider is collecting or sharing specific information and then asks the participants' comfort level knowing this data will be collected or shared without explaining the purpose of the collection or sharing. A 5-point Likert scale ranging from very comfortable (+2) to very uncomfortable (-2) was used to measure the participants' comfort levels. Figure 2 shows some sample questions of the two conditions.

Figure 2. Sample questions



- The fifth section asks participants if they want to have control over their data collection and sharing practices.
- The sixth section asks participants if they are aware of the controls the service provider makes available to the user related to data collection and sharing practices, and, if so, whether the participants use the provided controls.

We published the surveys through the crowd sourcing tool Amazon's Mechanical Turk (AMT) because of crowd sourcing's low cost and its ability to attract a large number of participants from diverse backgrounds. We designed a human intelligence task (HIT) to link to each survey. Each HIT contained a short description of the survey, a link to the related survey, and a code to be entered after completion of the survey for verification and approval.

We make a between-subject study design in which different participants participated in each condition. We posted 11 different HITs, two for each service provider and each condition. For the Twitter expectation condition, however, only one HIT was posted because it was our first published survey and we chose 50 participants for that HIT, which made it hard to manage and approve; thus, we decided to design the other HITs to accept only 25 participants. On average, each HIT took 12 days to be completed. Participants spent about 3 minutes and 42 seconds per HIT and were paid at the rate of $1 per HIT.

We collected a total of 323 responses with an average of 93% American participants. Tables 2 and 3 show the gender and age distribution of the participants. Twenty of the submitted responses were discarded due to the participants' rejection of the consent form or because they did not have an account with one of the study's service providers.

RESULTS

When analyzing the data, we focused on two main categories, both with specific subcategories. The first category, data type, has four subcategories of personal, activity, log, and cookies information. The second category, data flow, has three sub-categories of user-X, X-Thirdparty, and Thirdparty-X in which X is a service provider.

1. Does users' demographic information affect their concerns?

The average comfort level of our data shows that the older the participants were, the more concerned they were about the privacy of their information, with averages of (-0.6) and (-1.01) for age groups 18–25 and 40+, respectively. The results are statistically significant with a 0.006 p-value in the t-test. In addition, we calculated the average privacy concerns based on participant gender. The average concern of female participants was (-0.92) and of males was (-0.67), indicating that females are more concerned about the privacy of their information than males. These results are also statistically significant because the p-value is 0.008.

2. Do users' frequency of usage and awareness of the service privacy policy affect their privacy concerns?

We calculated the average of participants' comfort levels in regard to the frequency with which they use certain services. The averages fall between (0.75), for multiple times a day usage, and (0.93) for weekly usage. The results are shown in Figure 5. Based on t-test results, in which we can conclude that the results are not statistically significant, the p-value is 0.43.

Overall, the participants' awareness of the privacy policies of services they use is low; 65% of the participants were only partially aware of the privacy policies, and 23% of participants were not aware of privacy policies. The results show that participants who did not know of the existence of the privacy policies were more concerned about their privacy than the participants who were aware of the privacy policies. The average comfort level of participants who did not know service providers

Table 2. Age distribution of the participants

18-25	26-30	31-35	36-40	40+	Total
61	74	51	51	86	323
19%	23%	16%	16%	27%	100%

Table 3. Gender distribution of the participants

Females	Males	Total
183	140	323
57%	43%	100%

Figure 4. Average comfort level of the participants based on their age

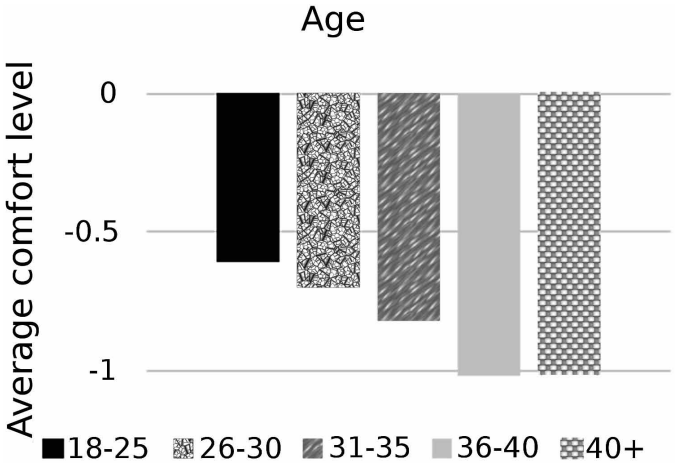
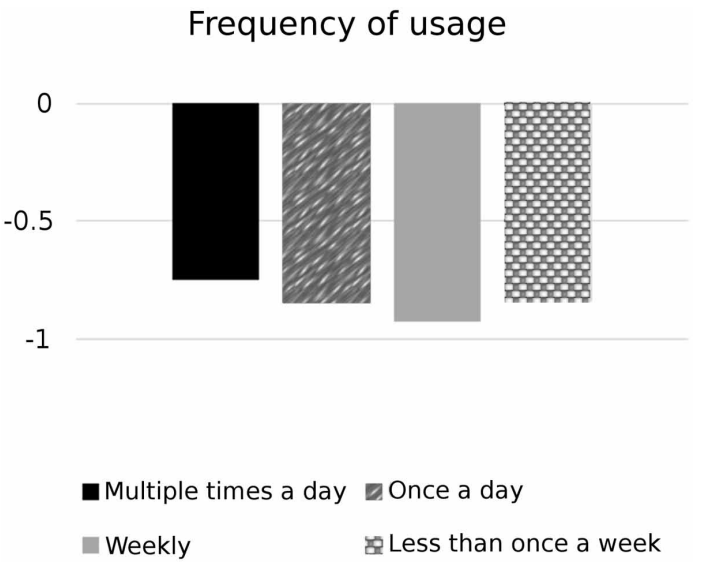


Figure 5. Average comfort level of the participants based on their frequency of usage



have privacy policies is (-1.8), and the average comfort level of participants who were fully aware of the privacy policies is (-0.53). The results are shown in Figure 6. The difference is statistically significant with a p-value of 0.0005, which indicates that awareness of the content privacy policies decreases participants' privacy concerns.

3. Do users prefer more controls?

The vast majority of participants prefer to control their own data collection and sharing practices. This is clearly seen by the fact that 92% of the participants thought that service providers should allow them to control the sharing or usage of their own information. Table 4 shows more details about users' control preferences.

Figure 6. Average comfort level of the participants based on their awareness of the privacy policy

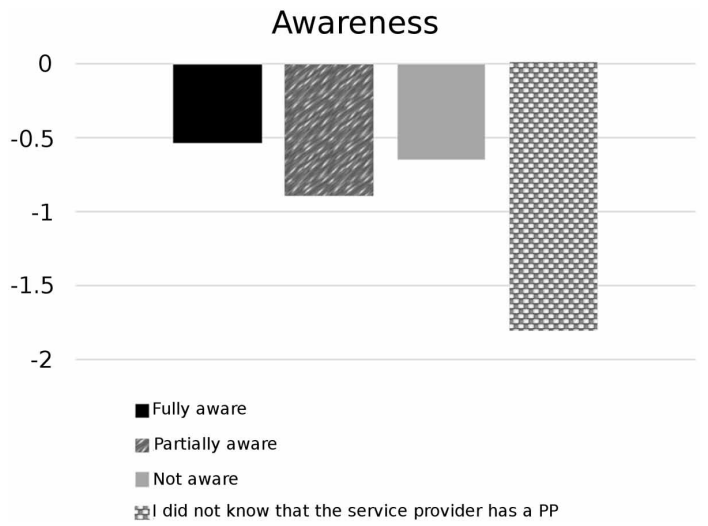


Table 4. Percentage of control preferences

Do you want more control on your privacy?			
	Yes	**No**	**Does not matter**
Facebook	90%	3%	7%
Twitter	92%	1%	7%
Bank of America America	93%	6%	1%
Average	92%	4%	4%

4. Are users aware of the provided controls? Do they use them?

About 62% of participants were not aware of the controls offered by the service providers. However, 63% of those who know that they have certain controls use them. The following tables show the detailed results for three companies.

Table 5. The percentage of Facebook participants' awareness of existed controls and the percentage of sub-question of whether they use the existed controls

Are you aware of controls provided by Facebook?		
Yes		**No**
40%		60%
Do you use it?		
Yes	**No**	
69%	31%	

Table 6. The percentage of Twitter participants' awareness of existed controls and the percentage of sub-question of whether they use the existed controls

Are you aware of controls provided by Twitter?		
Yes		**No**
52%		48%
Do you use it?		
Yes	**No**	
57%	43%	

Table 7. The percentage of Twitter participants' awareness of existed controls and the percentage of sub-question of whether they use the existed controls

Are you aware of controls provided by Bank of America? ofofAmerica?		
Yes		**No**
21%		79%
Do you use it?		
Yes	**No**	
64%	36%	

5. Which data flow or data type are the most unexpected by users, and which data flow or data type are users most uncomfortable with?

Using the "exception condition" version of the survey, we analyzed the results based on the data type categories and data flow categories to investigate which category is least expected by the participants.

First, for each service provider, we calculated the percentage of the participants who expected the collection or the sharing of their information through each flow. We also averaged the self-reported comfort ratings, ranging from "very comfortable" at +2.0 to "very uncomfortable" at -2.0, with "neutral" at 0.

We observed a strong correlation (r=0.76) between the percentage of expectations and the average comfort ratings. For example, 28% of participants expected Twitter would share their information with a third party, and overall participants felt uncomfortable about this information sharing (-1.26). However, 94% of participants expected a third party would share their information with Facebook. Overall, the average comfort level of (-0.54) shows that participants felt somewhat comfortable about this information sharing. We can conclude that when participants were not expecting a collection or sharing of their data through particular flow, they were less comfortable with this practice. The summary of our findings is shown in Table 8.

Next, we analyzed the results for each service provider and data type category. We calculated the percentage of the participants who expected the collection or the sharing of a particular data type. We also averaged the participants' self-reported comfort ratings.

We observed a correlation of (r=0.5) between the percentage of expectations and the average comfort ratings; 95% of Facebook users expected that Facebook collects or shares information about their activity information, and overall results showed that Facebook users felt somewhat comfortable about Facebook's monitoring their activities (-0.54); 47% of BoA users expected that BoA collects or shares information about their activity information, and overall results showed that BoA users felt

Table 8. Percentage of expectation and average comfort level based on type of flow

Type of flow	Organization	% Expectation	Average comfort
user-X	Facebook	-	-
	Twitter	43%	-0.62
	BoA	-	-
X-ThirdParty	Facebook	42%	-0.89
	Twitter	28%	-1.26
	BoA	42%	-1.00
ThirdParty-X	Facebook	94%	-0.54
	Twitter	27%	-1.00
	BoA	-	-

uncomfortable about BoA's monitoring their activities (-1.2). The findings indicate that the expectation of collection or sharing of a specific data type was somewhat linked to participants' subjective feelings. The more they expect the collection of a specific information the more comfortable they feel, and the less they expect the collection of a specific information the less comfortable they feel. The summary of the findings is shown in Table 9.

6. Do users experience difficulty identifying the reason service providers collect and share information?

We found that participants experienced difficulty identifying why service providers collect or share their information, with an average of 43% guessing correctly.

We divided the purposes of privacy policy segments into three categories: (a) for major functionality, (b) for sharing and tagging, and (c) for target advertising or market analysis. Some

Table 9. Percentage of expectation and average comfort level based on type of data

Type of data	Organization	% Expectation	Average comfort
Personal	Facebook	52%	-0.81
	Twitter	34%	-1.05
	BoA	37%	-0.8
Activity	Facebook	95%	-0.54
	Twitter	61%	-0.41
	BoA	47%	-1.2
Log	Facebook	-	-
	Twitter	20%	-0.56
	BoA	-	-
Cookies	Facebook	10%	-1.18
	Twitter	-	-
	BoA	-	-

privacy purposes fell into more than one category. For example, the purpose of Facebook's collecting users' activities falls into both the major functionality and advertising categories.

We compared the participants' provided purposes for information collection and sharing against the actual purposes as stated in privacy policies for different types of data categories. The results as shown in Table 10 indicate that the majority of the participants could not correctly identify why the service provider collects or shares a specific type of data. While more participants were able to identify why their activity information was collected or shared compared to their personal, log, and cookies information, overall, less than 50% selected the correct reason for these service providers' practices.

We followed the same analysis we performed on type to data for different categories of data flow. The results show that only 48% of the participants on average selected the correct reason for information collection and sharing practices. We noticed that for third party-X type of flow an average of 59% participants chose the correct answers. This average is better than the other two categories of data flows in which only 40% of answers were correct for X-Third Party and only 48% of answers were correct for user-X. Table 11 shows all of the results.

Table 10. Type of data. The first column shows type of accessed data. The second column shows the ground truth of why the data are shared. The third column shows the percentage of participants stated the purpose correctly. The last column shows the percentages of participants who chose that they cannot think of a reason

Type of data	Information used for: [1] Major functionality [2] Tagging or sharing [3] advertising or market analysis	% of correct choice	% of do not know
Personal	[1]	47%	18.9%
	[2]	37%	26.4%
	[3]	37%	10.2%
Activity	[1] + [3]	78%	2.9%
	[3]	47%	20.4%
Log	[1]	20%	13%
Cookies	[1]	10%	10%

Table 11. Type of flow. The first column shows the type of flow. The second column shows the ground truth of why the data are shared in this type of flow. The third column shows the percentage of participants stated the purpose correctly. The last column shows the percentages of participants who chose that they cannot think of a reason

Type of flow	Information used for [1] Major functionality [2] Tagging or sharing [3] advertising or market analysis	% of correct choice	% of do not know
user-X	[1]	61%	13.0%
	[1]+[3]	34%	1.9%
X-ThirdParty	[1]	41%	32%
	[2]	37%	26.4%
	[3]	42%	15.3%
ThirdParty-X	[1]	94%	4%
	[1]+[3]	24%	25.9%

7. Does clarifying the purpose of the information collection or sharing ease users' concerns about the privacy of their data?

In this section, we study whether or not clarifying the purpose of the information collection or sharing eases users' privacy concerns. We compared the average comfort ratings from surveys using an expectation condition in which the purposes of information collection or sharing are not revealed to participants to the average comfort ratings from surveys using purpose condition in which the purpose is revealed. The results were analyzed based on data type and data flow categories.

The results show that the differences between the comfort ratings with and without stating the purpose of information collection and sharing were not statistically significant. For example, with regard to activity type data, the t(298) and p-value are 1.97 and 0.102, respectively. This and the other results can be found in Table 12.

Table 13 shows the results for the three different types of data flow. The difference between the comfort ratings was not statistically significant, as shown by the example of user-X data flow, in which the t(102) and p-value are 1.98 and 0.402, respectively.

8. Is it possible to predict the level of privacy concerns for a particular user?

In this section, we proposed a model using logistic regression to predict the level of concern for a particular user with specific demographic information, specific frequency of use, and specific level of awareness. We used R in our study as an analytical tool. We had one dependent variable, comfort level, and seven independent variables: service, dataflow, datatype, gender, age, frequency of use, and awareness level. The types of these variables are either canonical or ordinal. Table 14 shows each variable type. We started first by analyzing the data of all the 605 participants (instances) for the three service providers. The results showed a value of 0.19 for the R^2, indicating a weak model. To improve our model and remove the data variant in concerns, we divided our data into two clusters based on the comfort level variable. We ran k-means clustering with k=2 and got 383 instances that belong to Cluster1 in which participants were more concerned (i.e., lower comfort level) and 222 instances that belong to Cluster0 with participants who were less concerned. We then applied logistic

Table 12. Comparison of comfort ratings between the expectation condition (the 2nd column) and the purpose condition (3rd column) based on data type

Data type	Comfort rating w/purpose	Comfort rating w/o purpose	df	T	P
Activity	-0.51	-0.71	298	1.97	0.102
Personal	-1.02	-0.88	299	1.97	0.213
Log	-0.48	-0.56	102	1.98	0.720
Cookies	-1.08	-1.18	99	1.98	0.573

Table 13. Comparison of comfort ratings between the expectation condition (the 2nd column) and the purpose condition (3rd column) based on type of flow

Type of flow	Comfort rating w/purpose	Comfort rating w/o purpose	df	T	P
User-X	-0.47	-0.62	102	1.98	0.402
X-ThirdParty	-1.11	-1.06	294	1.97	0.632
ThirdParty-X	-0.68	-0.78	202	1.97	0.509

Table 14. Variables type

Variable	Type
Organization	Categorical
data_flow	Categorical
data_type	Categorical
Gender	Categorical
Age	Categorical
freq_usage	Categorical
Awareness	Categorical
Comfort level (dependent)	Ordinal (-2, -1, 0, 1, 2)

regression on each cluster. The results show an R^2 value of 0.28 for Cluster1 and 0.26 for Cluster0. We also applied different combinations of the data to obtain better results with higher R^2. Table 15 shows the R^2 values of each combination.

We also employed linear regression and ordinal regression, but neither method provided better R^2 results. Table 16 shows the R^2 values of each combination using the linear regression method. In most cases, the results were worse than applying logistic regression in terms of R^2 values. According to the R^2 values of the regression models, we found that it is difficult to predict a participant's privacy concerns given specific values by applying regression models.

In the next section we show which independent variables correlate to users' privacy concerns.

9. What other factors affect users' privacy concerns?

Table 15. R^2 values for the results of logistic regression modeling

	All participants	Cluster 0	Cluster 1
All three companies data	0.19	0.288	0.268
Data for Facebook only	0.179	0.33	0.272
Data for Twitter only	0.313	0.349	0.478
Data for Facebook and Twitter	0.134	0.485	0.206
Data for BOA only	0.457	0.069	0.438

Table 16. R^2 values for the results of linear regression modeling

	All participants	Cluster 0	Cluster 1
All three companies data	0.163	0.272	0.228
Data for Facebook only	0.124	0.448	0.175
Data for Twitter only	0.268	0.326	0.422
Data for Facebook and Twitter	0.155	0.307	0.233
Data for BOA only	0.379	0.064	0.329

To study other factors that affect users' privacy concerns, we used an R statistical tool to build a logistic regression model. Table 17 shows independent attributes along with p values and coefficients that have significant correlation with comfort levels. Note that all independent variables are categorical, so R changes them info factors so that each value of a categorical variable is seen as a binary variable. For example, each category of data flow will become a binary variable.

The results show that some factors related to data flow, data type, type of organization, and lack of awareness of privacy policy affect users' privacy concerns. For example, Facebook is positively correlated with comfort levels, which may be because people are willing to share their information on Facebook. Users are generally fine when their data are shared with the service provider. Users are concerned when cookies or personal data are being collected and shared. Users who are partially aware or unaware of the privacy policy also have greater privacy concerns.

The rest of the attributes' correlation results support the findings of previous research questions, question 1 and question 2. Table 17 shows these correlations with p-values <0.05. For example, older people or females are more concerned with privacy. Also, lack of awareness of privacy policies increases privacy concerns.

SOME GUIDELINES

Based on our findings, we came up with a list of guidelines to help organizations improve their privacy policies.

1. Improve privacy policy readability and describe privacy practices more clearly, especially regarding the use of cookies, personal information collection, and the sharing of users' information with third parties. This is suggested based on the finding that a large portion of users are partially unaware of some policy segment fields and as a result they tend to have higher levels of privacy concerns.
2. Make it more explicit to users that they have controls on certain information usage or sharing. This is based on the finding that most users are unaware of some provided controls; however, the users who know and understand these controls tend to use them.
3. Develop a visualization model for the privacy policies using the flow chart in Figure 3.
4. Another possible suggestion is to use tools that help people understand existing privacy policies. These tools use natural language processing techniques to parse automatically or semi-automatically existing privacy policies and present them in easy to understand formats such the table format we used in the chapter.

Table 17. Independent attributes that have significant correlation with comfort levels

Attribute	P-Value	Coefficient
organization=social page (Facebok)	<0.0001	1.09
data_flow=user-X	0.007	0.78
data_type=cookies	0.002	-1.1
data_type=personal	0.003	-0.67
gender=male	0.0005	0.58
age=36-40	0.001	-0.85
age=40+	0.001	-0.76
awareness=did-not-know-X-has-PP	0.004	-3.4
awareness=Partially-aware	0.002	-0.84

Figure 3. Average comfort level of the participants based on their gender

5. Develop a comparison model of privacy policies to help users better determine which policy is preferable to them based on the information it collects, the party with whom it shares information, and the controls it provides.

DISCUSSION

The purpose of this study is to analyze users' responses to privacy practices presented in the privacy policy of online social networks and financial institutions. We started by breaking down privacy policies into easy-to-understand segments and then used crowd sourcing to collect data.

The results of this study show that demographic factors do affect users' levels of concern related to the privacy of their information. It shows older users have greater concerns about their privacy. This finding supports Hazari and Brown's (2013) argument that privacy concern increases with age. Our study results indicate that females are more concerned about their information privacy, which supports the position of Nowak and Phelps (1992) but contradicts Hazari and Brown (2013), who stated that gender difference is not a significant factor in determining users' levels of privacy concern.

In addition, our study shows that an average of 92% of the participants preferred that companies allow users to control their data collection or sharing. The results also show that the majority of the users who know that they have control over some of their information will use the provided controls. This finding agrees with other findings (Consumer Action Organization, 2013; Young & Quan-Haase, 2009) that have stated that privacy control is a critical issue for users. However, surprising results of our study show that, when controls are provided, not all users are aware of them. We found an average of only 38% of participants were aware of provided controls. So we believe service providers need to make information on such controls more prominent and explicit.

Our results suggest that when users do not expect their data to be collected or shared, especially when involving third parties, log data, or cookies, they tend to have more privacy concerns. We suggest that privacy policies be written to make such practices more obvious to users. Nonetheless, while our original finding is consistent with Lin et al.'s (2012) results, our results show that providing the reason for the collection and use of particular information collected or used had no significant influence on users' privacy concerns, while their results show purpose has a significant impact on users' concerns. One possible reason could be that in our settings of online social networks and financial services, participants already know that their information is being collected or shared, and

they are expecting that this is done for specific reasons. In their settings of smart phone apps, users are unaware of what data are shared or collected in the first place.

We also built a regression model to predict the level of privacy concerns for a specific user for a specific privacy policy segment field. This proved to be a weak model. A reason could be that our model did not capture other factors affecting users' concerns such as whether or not the user is concerned with privacy in general. This model indicates that, in addition to demographic factors, the type of organization also affects users' privacy concerns. For example, Facebook users are less concerned with privacy than those of the other two organizations. A possible reason is that Facebook users mainly use the service to post information about daily activities, thoughts, pictures, and interests and to share that information with others, so they felt more comfortable regarding the privacy of their information. However, this is not the case with BoA, as the main goal of using the service is not sharing information. The situation is similar for Twitter, whose users are more likely to use the service for posting news and comments with 140-character messages.

Other factors that affect privacy concerns include data type in which users are more concerned with sharing cookies and personal data and data flow in which users' concerns are higher when data are shared with third parties. Another remarkable result is that the lack of awareness of some policy practices is strongly correlated with increased privacy concerns.

CONCLUSION

People use online services in their basic daily practices, and they share their information in order to create profiles on service websites and exchange information with others. Service providers usually present a privacy policy that explains what information they collect, how they are going to store the information, and with whom they are going to share it. A privacy policy is the tool that aids users in identifying and understanding privacy practices when they are using an online service. It is important that a privacy policy is written in a clear and concise format so the users can be made aware of the privacy with which their information will be handled. They will be able to avoid privacy risks that can result in things like embarrassment and cyber-stalking, or more serious risks, such as identity theft, loss of employment, or even murder. An important step to improving users' awareness of their relationship with online privacy is to study their familiarity with current policies and then further develop the parts that are not clear to them by making it simple and easy to follow.

REFERENCES

Adams, H. R. (2013). *Protecting Intellectual Freedom and Privacy in Your School Library*. ABC-CLIO.

Barbaro, M., & Zeller, T. (2006). A Face Is Exposed for AOL Searcher No. 4417749. *New York Times*.

Benassi, P. (1999). TRUSTe: An online privacy seal program. *Communications of the ACM*, *42*(2), 56–59. doi:10.1145/293411.293461

Chaianuchittrakul, C. (2013). *Crowdsourcing Privacy Policy Analysis Evaluating the Comfort, Readability and Importance of Privacy Policies*. Carnegie Mellon University.

Consumer Action. (2013). Consumer action "do not track" survey results. Retrieved from http://www.consumer-action.org/downloads/english/Summary_DNT_survey.pdf

Culnan, M., & Williams, C. (2009). How ethics can enhance organizational privacy: Lessons from the ChoicePoint and TJX data breaches. *Management Information Systems Quarterly*, *33*(4), 673–688.

Dinev, T., & Hart, P. (2004). Internet Privacy, Social Awareness, And Internet Technical Literacy: An Exploratory Investigation. *Proceedings of BLED '04*.

Earp, J., Antón, A., Aiman-Smith, L., & Stufflebeam, W. (2005). Examining Internet privacy policies within the context of user privacy values. *IEEE Transactions on* Engineering Management, *52*(2), 227–237.

Eastlick, M. A., Lotz, S., & Warrington, P. (2006). Understanding online B-to-C relationships: An integrated model of privacy concerns, trust, and commitment. *Journal of Business Research*, *59*(8), 877–886. doi:10.1016/j.jbusres.2006.02.006

Göritz, A. S., Wolff, H. G., & Goldstein, D. G. (2008). Individual payments as a longer-term incentive in online panels. *Behavior Research Methods*, *40*(4), 1144–1149. doi:10.3758/BRM.40.4.1144 PMID:19001406

Gross, R., & Acquisti, A. (2005). Information Revelation and Privacy in Online Social Networks (The Facebook case).*Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society* (pp. 71–80). doi:10.1145/1102199.1102214

Harris Interactive. (2001). Retrieved from http://www.harrisinteractive.com

Hazari, S., & Brown, C. (2013). An Empirical Investigation of Privacy Awareness and Concerns on Social Networking Sites. *Journal of Information Privacy & Security*, *9*(4), 31–51. doi:10.1080/15536548.2013.10845689

Jensen, C., & Potts, C. (2004). Privacy policies as decision-making tools: an evaluation of online privacy notices.*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 471–478). doi:10.1145/985692.985752

Kelley, P., Cesca, L., Bresee, J., & Cranor, L. (2010). Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach.*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1573–1582). doi:10.1145/1753326.1753561

Leon, P., Cranor, L., McDonald, A., & McGuire, R. (2010). Token attempt: the misrepresentation of website privacy policies through the misuse of p3p compact policy tokens. *Proceedings of the 9th Annual ACM Workshop on Privacy in the Electronic Society* (pp. 93–104). doi:10.1145/1866919.1866932

Lin, J., Amini, S., Hong, J., Sadeh, N., Lindqvist, J., & Zhang, J. (2012). Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing.*Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 501–510). doi:10.1145/2370216.2370290

Marshall, N. (1974). Dimensions of privacy preferences. *Multivariate Behavioral Research*, *9*(3), 255–271. doi:10.1207/s15327906mbr0903_1 PMID:26805474

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23. doi:10.3758/s13428-011-0124-6 PMID:21717266

McDonald, A., & Cranor, L. (2008). The Cost of Reading Privacy Policies. *ISJLP*, *4*(3), 540–565.

McDonald, A., Reeder, R., Kelley, P., & Cranor, L.F, . (2009). *A Comparative Study of Online Privacy Policies and Formats*. Springer Berlin Heidelberg. doi:10.1145/1572532.1572586

Milberg, S., Burke, S., Smith, H. J., & Kallman, E. A. (1995). Values, personal information privacy, and regulatory approaches. *Communications of the ACM*, *38*(12), 65–74. doi:10.1145/219663.219683

Milne, G., & Culnan, M. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, *18*(3), 15–29. doi:10.1002/dir.20009

Nowak, G. J., & Phelps, J. (1992). Understanding privacy concerns. An assessment of consumers' information-related knowledge and beliefs. *Journal of Direct Marketing*, *6*(4), 28–39. doi:10.1002/dir.4000060407

Pollach, I. (2007). What's wrong with online privacy policies? *Communications of the ACM*, *50*(9), 103–108. doi:10.1145/1284621.1284627

Schoenbachler, D., & Gordon, G. (2002). Trust and customer willingness to provide information in database-driven relationship marketing. *Journal of Interactive Marketing*, *16*(3), 2–16. doi:10.1002/dir.10033

Sidel, R., & Johnson, A. R. (2012). Data breach sparks worry: Hack attack at card processor compromises potentially thousands of accounts. *The Wall Street Journal*.

Smith, S. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *Management Information Systems Quarterly*, *35*(4), 989–1016.

Sumeeth, M., & Singh, R. (2010). Are Online Privacy Policies Readable? *International Journal of Information Security and Privacy*, *4*(1), 93–116.

Tufekci, Z. (2008). Can you see me now? Audience and disclosure regulation in online social network sites. *Bulletin of Science, Technology & Society*, *28*(1), 20–36. doi:10.1177/0270467607311484

Westin, A. (2001). *Opinion surveys: What consumers have to say about information privacy*. Prepared Witness Testimony, The House Committee on Energy and Commerce.

Young, A., & 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* (pp. 265–274). doi:10.1145/1556460.1556499