Online Privacy Behavior Laboratory Study- Summary
Andrew Boyd, Tariq Wali, Anurag Govil, Jothi Pottukalam, Gopi Keshavan
Seidenberg School of CSIS, Pace University, White Plains, NY 10606, USA
[Team10@corp.aol.com]

Abstract
The study of online privacy behavior is a relatively new field, which suffers from a lack of empirical studies and needs to be examined in greater depth. This study defines various online privacy behaviors across a range of online contexts and has identified significant differences in user initiated disclosure of different types of personal information among various online contexts. For the purpose of this project, we define online context as a specific type of online interaction and have narrowed our range of online contexts to mainstream social networking sites like Facebook, professional networking sites like LinkedIn, eCommerce sites like Amazon and personal health sites like patientslikeme.com. Our study supports prior published observations that online users behave counterintuitively when disclosing personal information. This suggests that existing models of personal information disclosure are inadequate, and that a new model of online disclosure is required.

Our study adds a new finding to existing literature in this area; we have identified significant differences in disclosure behaviors among users within various online contexts.

Keywords: Online privacy, privacy behaviors, online context, social networking sites, safety and security.

1. Introduction
Social networking sites, by their very nature, encourage users to share personal and professional information. Within social science, computer sciences, and the popular media, there is growing concern over how social networking sites collect and use personal information and how this information is shared by users. In particular, revenue generating business models for social networking sites, content publishers, search engines and content aggregation sites represent a series of threats to personal privacy. When it comes to online sharing of information users do not fall neatly into clear-cut groups when looking at the type of information they are willing to disclose or the ways they restrict access to the information that they do share. Likewise when it comes to any of the social networking sites, those users who have the restricted access to their profiles are not always more forthcoming with the range of information they share but that doesn’t put them in the conservative users groups. It can also take into account the fact these users are also the ones who actually more likely to post fake information there, rather than the other way around. With that being said there could be various reasons behind this, as it could be referring to the basic placeholder information that is posted to the “front page” of their profile – which is visible to anyone, even when the rest of your profile is restricted – or perhaps they could be simply more comfortable joking around and being silly with friends. But still it goes to show that these users are not comfortable as they feel insecure while sharing their personal information. Many users don’t actively set privacy controls which make the profile public by default and can be easily accessible on internet on a simple search and available to potential marketers, advertisers, your employers, friends and even people with bad intentions.

The main difference between security and privacy is that, for privacy, the individual about whom the personal information is collected and handled effectively "owns" that information and, as such, should be able to control what happens to it. That is to say, the person should be able to control what personal information about him/her can be and is collected; ensure that it is correct and current; and decide who can look at it, share it and/or change it. Social networking websites, where people can meet, chat and message each other online are used by millions of people all over the world. However, users also have a role to play in protecting their own privacy. The rapid growth of Social networking sites coincides with increasing concerns about personal privacy. Users across a wide range of Social networking sites show a concern towards safeguarding their identity and private information from the unknown future usage on social networking sites. These virtual communities in which users can potentially share all kinds of information like name, age, sex , marital status, occupation, date of birth, phone numbers, nationality, political preferences, religious preferences personal preferences about music, cinema, and books have become widely popular and central to our everyday lives. No doubt this information is very critical and many users don’t understand the control over its publication or potential misuse.

Of the Social networking sites that have emerged in the past few years with millions of active users, Facebook
and LinkedIn have been especially significant with fastest growth and reach to a global audience. Facebook, in particular, has more than 300 million active users out of which 50% of the active users log on to the site in any given day [9]. The primary characteristics among users exhibit “friend” relationships with peers and friends around the world, quizzes and applications, groups and fan pages, users can message one another; upload photos, videos, and notes; join a wide variety of discussions, add applications; send electronic “gifts”; and compose elaborate personal profiles containing a wide variety of sensitive information.

2. Study Components

Our study focused on various incarnations of social networking sites and we defined them in within four contexts; social networking sites, professional networking sites, e-commerce sites, and health information management sites [14]. Background research for this study included general research on the scale and reach of sites, a brief overview of privacy settings for specific social networking sites, and a high-level online survey of user attitudes towards online privacy.

The second part of this study examined specific disclosures of personal information across the four contexts of social networking, career networking, shopping and healthcare sites.

3. Social Networking Site General Overview

Social networking sites are a significant extension and a great shift from how actually socializing took place in the real world and changed dramatically in the virtual world (online). Social networking sites have brought people on the level where everyone is connected and share information, exchange thoughts, stream media and post pictures etc and where you are likely to meet a person of your interests and tastes. According to US-CERT Social networking sites, sometimes referred to as "friend-of-a-friend" sites, build upon the concept of traditional social networks where you are connected to new people through people you already know. The purpose of some networking sites may be purely social, allowing users to establish friendships or romantic relationships, while others may focus on establishing business connections.

Although the features of social networking sites differ, they all allow users to provide information about themselves and offer a range of communication mechanisms (forums, chat rooms, email, and instant messenger) that enables you to connect with other users. On some sites, users can browse for people based on certain criteria, while other sites require that users be "introduced" to new people through shared connections. Many sites have communities or subgroups based on a particular interest [13]. The aim of this section of our study is to provide a concise and consolidated review of privacy issues and attitudes towards online social networking. Existing literature outlines perceptions, attitudes and behaviors of consumers’ in relation to their technology-related privacy concerns; this literature will be reviewed and a number of gaps in relation to technology-related privacy concerns will be outlined.

4. Network Market Dynamics & Site Popularity : Traffic Data

This section will discuss Facebook’s overall user base size across various dimensions of demography, accessibility and reach. The number of existing online social networks is impressive, although it is difficult to fairly assess their relative size and reach due to the problems with network size considering other partners and affiliations delineated in Fig.1. We feel that sites reach ability is a fairer indicator of a site’s popularity though this has complexities as well. We identified one of the publicly available Alexa traffic rankings [1]. While these are commonly used as a general indicator of the amount of traffic a site is receiving with daily reach in Table 1.

![Fig.1](image_url)

Table 1.
5. Selection Of Sites
We selected 2 social networking sites for our study. Our goal was both to conduct a representative survey of the major, general-purpose social networking sites, and include several representatives of other common social, professional and shopping networking niches for comparison.

6. General-Purpose Sites
Our operational definition of a general-purpose social networking service for our study is one which anybody is free to join, people commonly present their real-world identity, and the primary use of the site is interacting with other users (friends and families), communities, fan pages, groups etc.

7. Facebook
Facebook is one of the foremost social networking websites, with over 300 million active users which represents the highest ever user growth across any social networking site on internet. There are bound to be risks to privacy with this much detailed user information arranged uniformly and aggregated into one place.

Users may submit their sensitive data without being aware that it may be shared with advertisers and various other partner vendors. Third parties may build a database from the social networks data to sell. Intruders may steal passwords, or entire databases, from social networking sites. When a new user registers on Facebook he or she is provided with a blank profile template consisting of a number of predetermined response categories. These include “basic” information such as gender and hometown, contact information such as email, mobile phone number, campus room/residence, and “personal” information such as interests, activities, favorite movies/music/books, among others. New users may also upload a profile picture and join up to five networks. The default privacy setting for a new user (“X”) is that the entirety of this information is viewable by anyone in any of X’s networks. X’s profile is also viewable by anyone who is “friends” with X on Facebook. While anyone who is not connected to X via network or friendship cannot view X’s profile, by default they can still locate X using a global search engine as well as view a version of X’s profile consisting of their photograph, name, and network affiliation(s) and even friends and send messages.

8. Default Setting Facebook
After a deluge of academic, popular media, and governmental criticism, included a sternly worded directive from the Privacy Commissioner of Canada, Facebook is gradually moving towards an "opt-out" approach to privacy. This will mean that users' details are kept private until otherwise stated. Facebook’s historical information philosophy and current default settings are to share almost everything included by users in their Facebook account and partner applications. That means if users have never customized their privacy settings, their Facebook profile is shared with not only all their friends and groups, but also with everyone in every network they have joined [3].

9. Scenarios and approach
For the purpose of our study and to reach out to users we created groups on Facebook, MySpace, and LinkedIn. The execution of our plan primarily focused on Facebook and asked direct questions to users. We were able to recruit 170 members in three weeks. The task included sending invites to all Facebook friends of the team members and requested others to promote and join. We asked members to evaluate the likelihood of certain online scenarios and privacy behaviors and provide an understanding. While some criticized and other appreciated the medium but the answers were interesting [8].

The situations included the following.

Unknown Friend Requests: When you get a friend request from an unknown person with no mutual friends and just the basic view available, how do you react [8]?

Information you share: You’ve probably learned a long list of online privacy and security measures already: be it about password eavesdropping or impersonating or yet having an easy access to your sensitive personal information. The Internet is the world's biggest
information exchange: many more people could see your information than you intend, including your parents, your teachers, your employer, the police — and strangers, some of whom could be dangerous. Social networking sites have added a new factor to the "friends of friends" equation. By providing information about yourself and using blogs, chat rooms, email, or instant messaging, you can communicate, either within a limited community, or with the world at large. But while the sites can increase your circle of friends, they also can increase your exposure to people who have less-than-friendly intentions. You've heard the stories about people who were stalked by someone they met online, had their identity stolen, or had their computer hacked. 

Who controls the Information: With the current state with which social networking sites are reaching out to millions of users where an individual can be targeted based on Online Profiling, such as user preferences and personality, interests, profession, hobbies, sexual-orientation and various other characteristics:

- Defining the ambiguities towards trust of social networking sites.
- Discovering why knowledge, experience and comfort lead to increased disclosure of personal information.

Users on Facebook often post status messages depending on the mood and situation, and various other factors (likes, dislikes, interests, hobbies, work, sports etc) in Fig.2. The information exchanged is important enough for Facebook to study their users and categorize them in different segments for behavioral targeting.

This set up a base for our further research and allowed us build a series of hypotheses about whether users are likely to share more information if they are not asked for it across the previously mentioned 4 contexts of online engagement.

A note about Facebook: Facebook members expressed significantly greater trust in both Facebook and its members than that expressed by users of other sites and were more willing to share identifying information.

10. MySpace

MySpace was the pioneer in social networking sites when it started and over time lost its’ leadership role to Facebook, remaining the world’s second largest social networking site. MySpace is predominantly famous for Music, Videos, Music bands, groups / forums, themes and applications and profile customization (HTML).

11. Default Settings MySpace

By default, people on MySpace can see when you’re online. Your profile and photo is also set to be viewable by everyone [3].

Existing law should be extended to allow some privacy protection for things that people say and do in what would have previously been considered the public domain.

For the first time in history nearly anybody can disseminate information around the world. People do not need to be famous enough to be interviewed by the mainstream media. With the Internet, anybody can reach a global audience.

That openness is both good and bad. People can now spread their ideas everywhere without reliance on publishers, broadcasters or other traditional gatekeepers. But that transformation also creates profound threats to privacy and reputations.

5.3.2 Information Control Options

MySpace’s privacy options are very limited, but changing three key settings can provide you with some important privacy protection.

Online Now: By un-checking this box, other users won’t know when you’re actively on your account.

Profile Viewable By: By selecting “My Friends Only,” you limit the ability for strangers to find your personal information.

Photos: By un-checking this box, you prevent your photos from being emailed and shared by other users. You should be aware that MySpace does not have individual privacy controls for each section of information you provide. Whenever you decide to add
information to your MySpace profile, you should take extra care in deciding what you divulge.

12. Scenarios and approach
The study involved creating a group and a forum on MySpace where relevant information is posted on discussion boards for studying privacy behaviors and user responses to privacy topics with reference to MySpace. Users depict somewhat more or less same behaviors about privacy on MySpace and we arrived at a discussion which had some interesting responses. What are the possible motivations for making one’s profile private on MySpace?

A private profile would be the one where you apply privacy controls and user preferences on all fields asking for user information and which can possibly be shared. Most of the users stated they fear of their information being misused and factors like culture, education, personal beliefs, likes and dislikes, pride all add up to whether one can share their private info or not. For example, it is taboo in some parts of India to share the photos of their babies, though they are just newly born. Culture and local beliefs are also adding up, though we are in information age. Other factors can be age, caste, creed and personal education that an individual has and the overall privacy awareness.

- Hypothesis: Public sharing of private lives has led to a rethinking of our current conceptions of privacy.

13. LinkedIn and Amazon
With more than 60 million users representing 150 industries around the world, LinkedIn is a fast-growing professional networking site that allows members to create business contacts, search for jobs, and find potential clients. Individuals have the ability to create their own professional profile that can be viewed by others in their network, and also view the profiles of their own contacts [5]. Business-networking sites differ from general-purpose in that they specialize in maintaining professional contacts and searching for new jobs. Users typically share much less personal information, yet more professional information on these sites. They often implement specific features for specifying and managing business relationships and are frequently used for job-searching. We included LinkedIn, the most popular professional/business-networking site with the highest reach and accessibility. The nature of the study recognized LinkedIn as the best site to study professional networking online privacy behaviors.

14. Default Settings LinkedIn
By default, your LinkedIn public profile is set to display your full profile information. That means any information provided will be available publicly. More than likely you will notice the very first link returned by Google search on your first and last name is you on LinkedIn and many
users had no idea how the other person found them and didn’t appreciate the spam for that matter.

15. LinkedIn privacy behaviors
LinkedIn groups which according to the site are private place on the website where you can communicate by sharing information with your colleague and even with your department head. These groups are automatically populated with LinkedIn users who list their company X as their employee. With these seemingly private groups, employees with less understanding of how third-party communities work may share potentially proprietary information about customers, partners, product roadmaps or even financials. While employees should never share unapproved company information via an external social networking site, LinkedIn’s lack of management tools for an assigned, verified company representative is also disconcerting — especially since these groups seem to be generated largely in part by honor system [4]. Here is a caveat: According to LinkedIn’s help documents for Company Groups: “When someone leaves the company and updates their position on their LinkedIn profile, they are automatically removed from the company network since the company network is for current employees only. But then “If the person has forgotten to update their LinkedIn profile, you can flag their position to indicate that they no longer work at the company and that they should be removed from the company network. The system will take these flags into account and seek to reconfirm the user’s work email address if they try and access the company network.”The policy is based on assumption that someone has to mark the flag for ex-employee and then LinkedIn takes time to verify the flag, in the meantime employee has access to the private information in the group.

• Hypothesis: What constitute privacy behaviors and practices on Professional networking sites?

16. Amazon
Privacy in e-commerce has always been a contentious issue as users strive to protect their personal information being misused by others. However, the advent of the Internet and the increasing proliferation of technologies in both the marketplace and workplace have been matched by a heightened awareness amongst users that threats to their privacy exist and must therefore be addressed. Despite the empowering benefits of the Web, consumers are increasingly aware that the technology can also be used by online vendors to collect potentially sensitive information regarding them and that this information can be used without their express permission. For example, online transactions require customers’ to disclose considerably more personal and financial information than they would provide in offline. Marketers can use the trail of information that results from such Internet transactions - including information on the customer’s searches, comparisons, product and brand preferences, purchase and post-purchase information - to compose very precise customer profiles in their efforts to continuously learn about changing consumer needs. Websites can collect huge amounts of data from users. Retailers, for example, can track our every click, what we buy, how much we spend, which advertisements we see - even which ones we linger over with our mouse. Sites can easily access your entire web browser history, enabling them to try and guess your gender and other demographic information. They know how much we paid for our house, what magazines we subscribe to, which books we buy and what vacations we take. The company purchases just about every bit of data about us that can be bought, and then sells selections of it to anyone out to target us [6].

• Hypothesis: In the Presence of prominent privacy information, users are likely to purchase more if not solicited and phished.

17. Hypothesis Development
The IUIPC model draws upon Social Contract theory to present a theoretical framework consisting of multidimensional first and second order elements, as well as a series of demographic covariates. For example, the IUIPC construct states that individual attitudes towards the collection and control of personal information and awareness of information privacy practices constitute a user’s IUIPC profile [10]. This individual IUIPC profile influences trusting beliefs and risk beliefs, which in turn have an impact upon behavioral intent. Demographic co-variants are also related to individual IUIPC profiles [10] [7]. A limited number of demographic co-variants were included in the initial version of IUIPC proposed by Malhotta et al, and this study expands that list to include social media specific co-variants. Our hypotheses posit that specific co-variants are correlated to specific privacy attitudes.

18. Methodology and Study Deployment
To ensure we had all the necessary information for the this study we created accounts in all the sites mentioned above and started sending invites to people asking them to join. At the same time we created groups and invited people followed by posting open ended questions and about online privacy and behavior and studied how people react to them. We drew qualitative analysis based on the comments, emails, postings, member suggestions, meetings and online literature reviews.
Our team developed a survey, which asked users what type of information they make available on social network sites within our four online contexts. The survey for our study was distributed via mass email from our individual email accounts, Facebook postings, and professional contacts of individual team members. Users of our social network site groups were drawn from friends, family, coworkers, students and classmates. We asked study participants questions that enabled us to determine their privacy attitudes and behaviors relating to our four online contexts.

The survey consisted of questions that include users' activities on social network sites like Facebook, MySpace, LinkedIn and Amazon. In terms of privacy, we also asked users about consumer privacy and security settings. The users responded to each scenario using a seven point Likert scale coded from ‘Strongly Agree’ to ‘Strongly Disagree’.

19. Study Design and execution
Privacy disclosure within the four online contexts was broken down into four areas:

- Personal Demographic Data
- Lifestyle Data
- Location Data
- Personal Health Information

To produce our hypotheses, those four areas were mapped against our four online contexts:

- Social Networking Sites
- Professional Networking Sites
- eCommerce Sites
- Personal Health Information Sites

Within each of these contexts, we hypothesized that some behaviors are significantly different from others, and we tested the four behaviors across each context in order to differentiate behaviors among contexts. We have charted out this data and formulated some hypotheses about how likely disclosure will be within these areas. Table 3 values indicate our hypotheses for probability of disclosure. ‘High’ means a hypothesis of high probability of personal information disclosure and ‘Low’ means a hypothesis of low probability of personal information disclosure.

<table>
<thead>
<tr>
<th></th>
<th>Social Networking</th>
<th>Professional Networking</th>
<th>eCommerce</th>
<th>Personal Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Demographic Data</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Lifestyle Data</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Address Data</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Personal Health Information</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

Table 3

Our grid suggests 16 hypotheses (4 x 4). As part of assembling the study design we generated 16 hypotheses defined in the following format.

- Hypothesis 1 (H1): Online users are highly likely to disclose personal demographic data within Social Networking contexts.

Through to:

- Hypothesis 16 (H16): Online users are highly likely to disclose Personal Health Information within Personal Health Information contexts.

19. Study Results

Our study participants were young (61% between 25-34) (20% between 15-24), highly educated (80% BA and or MA), highly connected and overwhelmingly South Asian (80% self-identified as South Asian or Indian). Additionally, study participants were highly aligned to social networking behaviors (75% agree that Social Network Size is an important factor in personal success). Our study revealed some very interesting results about social network use (92% of the users use social networking sites to meet new friends and/or re-establish contact with old friends who share their interests, 80% use these sites for dating, 88% used Professional Networking Sites for job searching, 85% used eCommerce sites for online shopping and 72% of the users used Personal Health Sites for managing and researching information about their personal health. In many ways, these survey respondents from India can be viewed as the model user group for a new frontier of online engagement.

Study Participants Fig.4
Internet Experience Fig.5

Social Network Information disclosure Fig.5
19. Limitations
In view of the scale, scope and further research of this study, there are some limitations that should be kept in mind that apply to all phases of our study. First, there can be an improvement in the selection of sites and the criteria to assess them. Section 3 of the paper explains the given account of our selection criteria. We have made a considerable effort in identifying all sites that fall into the operational definition of social networking contexts; and it is imperative that we have not considered many other social networking sites because of the time constraints. The user count defined in the sample size is particularly sensitive to cut-off levels because of the high mobility in social networking contexts. Due to the lack of access to many aspects of our study, and challenges in communication in expanding our sample forced us to compromise on the depth of analysis. This study has adopted the point of view that at a later point in time the sample could be expanded more efficiently in breadth than in depth, consider other criteria’s and market dynamics. It might be possible we missed an important evaluation criterion or metadata entry. Our choices were driven by research and analysis and lessons learnt from past studies, trends, and current privacy disclosures in social networking contexts. We also had limitations in attempting to evaluate some more qualitative elements, such as the usability of privacy controls or the readability of privacy policies.

20. Conclusion
Our results support the majority of the hypotheses of our study design. Our in-depth qualitative research study aimed to investigate different attitudes and privacy behaviors in various social networking contexts and their use among 127 users in different geographies. The study included teens, young and adults, users and non-users, and covered predominantly India followed by United States. The qualitative nature of this study means that our findings do not necessarily represent of all those who use or do not use social networking sites, the results help us understand and gain better insights around people’s use of social networking contexts, attitudes, behaviors and various privacy disclosures.

Our study reveals 99 out of 127 respondents disclosed highly sensitive personal information. The majority of personal health information site users engage with these sites for in a research model of use, rather than as a personal health information disclosure and engagement model of use. We also found that large number of users have privacy concerns when they are engaged in any networking context on the Internet when they shop online or when it comes to socializing with users from different nations, and many do not care to read “fine print” or privacy policies in large. Instead, they tend to rely on the trust and reputation of networking context and notice the presence of privacy policies more often than actually reading them. While we also learnt that users find it difficult to get the privacy information they want and rather choose to bypass reading privacy policies or even verifying the complete credentials of users, friends, groups etc and yet hope for the best. It is interesting to see that users have generally realistic views of the relative likelihood of certain situations that could occur once their information is online. Many of them realize that the contexts they use track what information they post, their activities, hobbies, their purchasing and searching habit patterns which pretty much translates to everything we do in everyday life.

Our study determined that social networking sites are used widely for dating, making new friends and to follow up on their favorite cinema, music, fan pages, authors, and are also used to support and maintain relationships, although not generally to create them. A considerable number of young people communicate with strangers online, and post information which may be considered ‘private’ in most circumstances. The ability to restrict security setting to limit the disclosure of personal information is known about but not widely practiced. The overarching conclusion of this survey is that context matters. Particularly in the area of personal health information, the more personal an issue an online user seeks to address, the higher is their willingness to disclose information. This conclusion, although it is consistent with our hypotheses and study results, is counterintuitive. Online users should be less, not more willing to disclose personal information as sensitivity increases. Further study of the reasons behind this behaviour is required; existing models of behaviour in the field inadequately explain the observed behaviour of our study participants.

References

[1] © Alexa Internet, Inc. An Amazon.com Company