# Private Lives: User Attitudes Towards Personal Information on the Web

**Carrie Gates**
**Tara Whalen**

Technical Report CS-2005-06

June 20, 2005

Faculty of Computer Science
6050 University Ave., Halifax, Nova Scotia, B3H 1W5, Canada

# Private Lives: User Attitudes Towards Personal Information on the Web

Carrie Gates
Faculty of Computer Science
Dalhousie University
Nova Scotia, Canada
gates@cs.dal.ca

Tara Whalen
Faculty of Computer Science
Dalhousie University
Nova Scotia, Canada
whalen@cs.dal.ca

## Abstract

*Research into privacy attitudes has generally concentrated on concerns over data collection and handling by businesses, government, and other formal organizations. However, there are many aspects of one's personal life that are not covered by this situation, such as casual snapshots and personal opinions. It is unclear what control people have over personal information that is released in this unregulated sphere, and what their concerns are about disclosure of private data. We conducted a set of studies investigating the types of personal information that people are comfortable having online in a public forum, which can be easily found through search engines. Additionally, we gathered information about people's experiences with how well they have controlled their own personal information as well as other people's data, and how satisfied they are with their ability to restrict what is revealed about them.*

*Our results show that people are most comfortable disclosing professional data (except salary), general personal data (vital statistics, leisure data, personal preferences, expressions), and photos. The items people were least comfortable with revealing online were personal contact details, affiliations, and government-released data (such as property assessments). Although we found some dominant concerns within our studies, we were unsurprised to find a great deal of variation among our respondents, and there were few pieces of data that* all *people would release. Some of our respondents were unsatisfied with only having social controls available to limit personal information publication, which suggests that mechanisms should be put in place to assist people in determining how much data about them is available on the web.*

## 1 Introduction

Online privacy research has traditionally focused on the formal, legal aspects of data collection and control, such as consumer information gathered by businesses or confidential patient data collected for medical purposes. However, outside of these narrowly-defined areas, there remain large regions of unregulated online territory in which people attempt to control revelation of their personal data. Certain aspects of the self, such as personal opinions or vacation photographs, do not fit neatly within a taxonomy of legally protected data. This is the sort of information released on home pages, blogs, and photo galleries: people decide which parts of themselves are included when they construct their online public persona. Determining which aspects to reveal publicly can be a difficult balancing act, given the accessible and persistent nature of online data. Complicating this problem is the fact that some elements are not released by oneself, but by other people. Consider the example of photographs: if a friend takes a group snapshot at a party, are they entitled to post that picture in their public photo gallery? What ability do people have to control how their personal information is publicized in online spaces?

Although some types of personal data are not protected by law, this does not mean that people are comfortable with having these items revealed. If people are not satisfied with their ability to control the things that are revealed about them, then it may be necessary to provide them with the means to better limit such exposure. Approaches might include policy changes that reflect the need to protect disclosure of personal data, or possibly technical tools that could assist users in maintaining a public online persona that they are comfortable with. Before such approaches can be considered, however, it is necessary to examine the degree to

which exposure is a problem, and the types of personal information that people are most concerned about controlling.

This paper contains the results of an exploratory study that investigates the types of information that people are comfortable having revealed in public online spaces. In particular, we look at those sources that are indexed by search engines, rather than controlled spaces (e.g., journals with protected access). Given the amount of personal data that can be correlated from various sources through simple searches, we felt this was an area that would be of concern to the majority of users. That is, we asked people what information they would release if they knew that it would be indexed (i.e., be searchable), exposing their data to a global audience. We conducted a short online survey to pinpoint the topics of most pressing concern, and refined these results to create a semi-structured interview. Our users described their comfort levels and concerns with several types of information, including contact details, photographs, and opinions; we looked at these aspects from both the perspective of self-revelation, and revelation by others. We also asked our participants to describe their experiences with controlling their personal data online, in order to gauge whether people were satisfied with the level of control they had over online revelation.

We describe the study that we performed in Section 2, and then provide detailed results in Section 3. We discuss the overall results and trends in Section 4, and then compare these results to related work in Section 5. We conclude with a high-level summary.

## 2 Description of Study

We conducted a two-phase study, the first part of which was an exploratory survey. We asked 16 librarians and library science students within our university to complete an online survey. Librarians were selected because of their role in providing access to information, tempered with a knowledge of privacy concerns. Details of this study are available in [8].

We used their responses to determine which types of personal data were of most pressing concern, which informed the design of our second study. We developed a more refined set of questions to determine comfort levels, and developed a set of categories of personal data to reflect various aspects of self-identity. We also concluded that richer responses were required in order to delve into the more complex aspects of privacy, and therefore we decided to follow up with a semi-structured interview rather than a simple survey.

The semi-structured interview had two main parts. The first part asked participants to rate how comfortable they would be with having specific types of information released online; we asked whether they themselves would release it, and how they felt about others releasing it. (We also asked them whether this type of data applied to them personally, in order to determine which responses were purely speculative, such as the case in which a person did not have work contact data.) For example, when asked about "age", people were asked to choose from the following set of responses:

```
Self-revelation (choose one):
o  I definitely would not make this available
o  I would prefer not to make this available
o  I might make this available
    Under what particular circumstances?

o  I would definitely make this available

Revelation by others (choose one):
o  I definitely would not want anyone else
   to reveal this online (workplace, family,
   friends)
o  I would prefer that others not make this
   available
o  I might allow someone to reveal this
    Under what particular circumstances?

o  I would allow anyone to reveal this
```

Participants were also prompted to discuss their particular concerns in several categories of personal data, and asked to relate any experiences they might have had online that affected their attitude.

The second part of the interview consisted of free-form questions on a more wide-ranging set of topics. In particular, we wished to hear about people's experience with controlling their personal data, and how they handled other people's personal data when they posted material online. We included a few imaginary scenarios to gauge how people might act when asked to make decisions about including or removing data.

The semi-structured interview was open to all members of our university community who had at least five years' experience with the web or other online forums (such as newsgroups). Participants were recruited using mailing lists; 16 people responded to our recruitment email, and the majority of these came from within the Faculty of Computer Science. The interviews took approximately one hour, and were conducted over a period of two weeks. Participants filled out background questionnaires that gathered specific details about their online experience, including length

of time spent online weekly, the types of forums read and posted to (e.g., blogs, photo galleries), and how many years' experience they had online. We also collected basic demographic data such as age and sex. Interviews were digitally recorded, augmented with field notes.

## 3 Results

We examined nine categories of information: vital statistics, personal contact data, professional data, social and leisure data, affiliations, government data, personal preferences and expressions, photos, and social affiliations. We converted our questionnaire scale into a four-point numeric scale, with '1' representing "definitely not" and '4' representing "definitely." In this case, responses of 2 and below indicate a negative inclination to reveal the data (or have it revealed), whereas values of 3 and above indicate a positive inclination. "Might" (value '3') was considered positive because it shows a willingness to release data under the right circumstances. When calculated values are between 2 and 3 we have concluded that there is some reluctance to reveal. In these case, we have interpreted these values to indicate a lack of comfort, biasing the scores towards the negative end of the scale.

The average score and the standard deviation for each category is presented in Table 1, where the subjects had responded regarding posting this information *themselves* to a website or public forum.

| Category | Average | Standard Deviation |
|---|---|---|
| Vital Statistics | 3.17 | 1.06 |
| Personal Contact Data | 2.02 | 1.16 |
| Professional Data | 3.36 | 1.03 |
| Social & Leisure Data | 3.00 | 1.15 |
| Affiliations | 2.43 | 1.19 |
| Government Data | 2.09 | 0.96 |
| Preferences & Expressions | 3.29 | 0.87 |
| Photos | 3.17 | 1.08 |
| Social Affiliations | 2.38 | 1.20 |

**Table 1. Average scores and standard deviations for each category of data when the information is posted by the subject.**

In general, people were comfortable providing personal preferences and expressions, with a (high) average score of 3.29 and a (low) standard deviation of 0.87. People were also generally comfortable posting photos, with a score of 3.17 and a standard deviation of 1.08. They were also comfortable posting both personal data and professional data. The score for professional data increased even further, from 3.36 to 3.57, when the question about salary was removed from consideration. This also had a strong effect on the standard deviation, dropping it from 1.03 to 0.822.

If only professional contact data is considered (address, phone number and email), then the score increased even further to 3.65 with a standard deviation of 0.73, indicating that people are very comfortable providing this information on-line. In comparison, people were not at all comfortable providing the same information for personal contact, with a score of 2.02 and a standard deviation of 1.16. If personal email was not included in this category, then the overall rating decreased to an average of 1.73 with a standard deviation of 1.07, indicating that people did not feel at all comfortable posting their personal contact information.

Also, while people were comfortable posting their personal preferences and expressions (with a score of 3.29), they were far less comfortable posting their affiliations (e.g. political, religious), with a score of 2.43.

Table 2 shows the same averages and standard deviations as above, however these were calculated based on the results from how subjects felt about *others* posting this information about them.

| Category | Average | Standard Deviation |
|---|---|---|
| Vital Statistics | 3.02 | 1.12 |
| Personal Contact Data | 1.73 | 0.96 |
| Professional Data | 3.20 | 1.09 |
| Social & Leisure Data | 2.74 | 1.21 |
| Affiliations | 2.22 | 1.21 |
| Government Data | 2.36 | 1.14 |
| Preferences & Expressions | 2.98 | 1.08 |
| Photos | 3.13 | 1.03 |
| Social Affiliations | 2.39 | 1.12 |

**Table 2. Average scores and standard deviations for each category of data when the information is posted by others.**

In general, the average score for each category is lower when it comes to others posting information rather than the subjects themselves. The only exceptions to this were government data (2.09 to 2.36) and social affiliations (2.38 to 2.39), although subjects are not comfortable with either being posted. They are, however, comfortable with others posting their vital statistics (e.g. sex, marital status), professional data, and photos. On average, the score when moving from self-revelation to revelation by others dropped by 0.13.

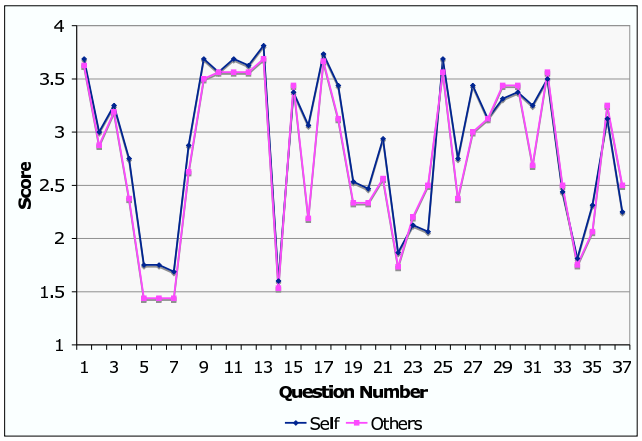


**Figure 1. The average score for all questions for both self-revelation and for revelation of information by others.**

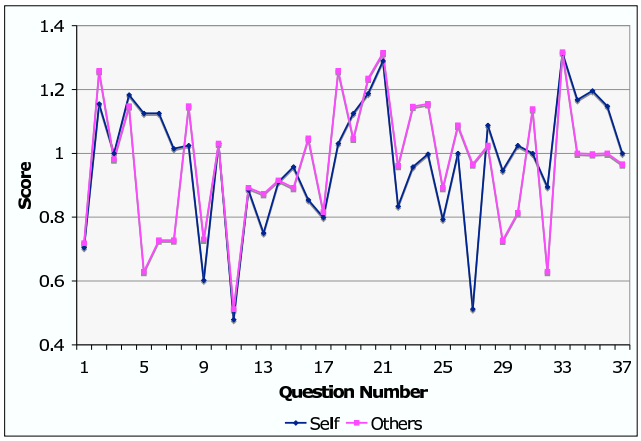


**Figure 2. The standard deviation for all questions for both self-revelation and for revelation of information by others.**

### 3.1 Interview Part I: Comfort Scales

Figure 1 shows the average score for each of the 37 questions in part one of the interview, with one line indicating the responses to each question when the subject was posting the information, and one for when others were posting the information. Figure 2 shows the standard deviations for the same set of information. The first four questions in this part dealt with vital statistics, the next four with personal contact data, then nine questions on professional data, two on social and leisure data, three on affiliations, two on government data, three on personal preferences and expressions, six on photos and four on social affiliations. As can be seen in the graphs, the comfort level subjects had with others posting their personal information closely tracks the comfort level they had with posting it themselves, with a correlation of 0.946. The standard deviation, showing the variance between users' responses, does not track nearly so closely, with a correlation of only 0.575. We examine each question asked in this part in more detail below.

**Vital Statistics**: Within the category of vital statistics, we asked people how comfortable they would be having the following information released on a web page that was indexed by Google: sex, age, marital status, and family information. In general, people were comfortable revealing this information; however, there is only statistical significance for this observation for sex and marital status. For sex, only two people preferred not to reveal it, at $p < 0.005$[1], and 13 of the remaining 14 people said that they would definitely reveal this information. Similarly, only four people said that they would not reveal their marital status ($p < 0.05$). We observed gender-based differences in response for sex, where the three responses that were not "I would definitely post this" were from three of the four female participants, however our sample is too small to determine significance. In terms of age, we observed that in general people who were younger were more likely to definitely post their age, with six of the seven respondents 25 and younger falling into this category, versus only two of the nine respondents who were 26 and older. Subjects were least comfortable releasing information about their family (e.g., number of siblings), with seven of the 16 respondents stating that they would not release this information.

**Personal Contact Details:** We asked people how likely they were to release personal contact information, where we concentrated on home address, phone number, driving directions to their home, and personal email. In general, people were comfortable releasing their personal email address but not any other contact information. Only four people would post their home address ($p < 0.05$), four people would post their home phone ($p < 0.05$) and three people would post driving directions to their house ($p < 0.05$). Some people specified that they "might" post driving directions, but that they would only get you to the street, and not to the house. In comparison, only three people would *not* give out their personal email address ($p < 0.05$), although nine people said that they "might" (as opposed to definitely) give out their personal email, with spam identified as a major concern.

[1] We used a $\chi^2$ test of independence to analyse these results.

**Professional Data:** Professional information was divided into work address, phone, office hours, job title, salary, qualifications (e.g., degrees), resume and publications list. In general, people were comfortable releasing all information other than salary. Everyone stated that they would post their work email. Only one person said that the would prefer not to post their work address, while two stated that they would prefer not to post their work phone. One person said that they definitely would not post their job title, stating that if someone needed to know it, they could contact him and ask for it. All remaining fifteen subjects stated that they definitely would post their job title ($p < 0.005$). This same person stated that they definitely would not post their publication list, resume, and qualifications, and that they would prefer not to post their office hours. In contrast, everyone else stated that they would post their publication list ($p < 0.005$). Only one other person stated that they would not post their office hours (compared to 14 who would, for $p < 0.005$). Two others stated that they would prefer not to post their resume ($p < 0.05$). A number of people commented that they had a specific version of their resume for posting on-line, versus one that they would provide with a job application, with the main difference being a lack of personal contact information other than an email address. Of the 16 respondents, 13 said that they would be willing to post their qualifications, with one person commenting "Qualifications are good, happy things. People can know good, happy things."

In contrast to all other professional information, salary was considered to be a personal issue, with ten people stating that they definitely would not post this information on-line, while one person said that they would prefer not to post this information. Four people stated that they might post this information, with three of them specifying that they would post a range, but not their actual salary. The fourth person stated that they would only post this information if required by law. We excluded one person from consideration as he answered that he is a government employee whose salary is currently posted on-line, and he did not discuss his own personal comfort with this information being available. It appears that there might be age-based differences in the answers to this question. Other than the person excluded, there were eight people who identified themselves as being older than 25, with only one of them stating that he might reveal this information. In contrast, nearly half of those who identified themselves as 25 years and younger (three out of seven) stated that they might reveal this information.

**Social and Leisure Information:** Social and leisure data centered around information about personal items such as hobbies, club memberships and sports. When first asked if they would post this information on-line, only two people specified that they would not ($p < 0.005$). However, these results changed some when asked if they would still post this information if it might be considered to be controversial (e.g., Greenpeace) or had personal connotations (e.g., being a supporter of a prisoners' rights group might suggest that you or a family member had a criminal history). For organisations such as these, only seven people stated that they might post this information. One person who stated that they definitely would not post their hobbies, stated that they definitely would post their membership in a more controversial organization, as he viewed these organizations to be more of a business relationship instead of personal.

There appear to be experience-based differences in the responses to posting membership in controversial or personal organisations. Of the seven people who have had a personal web site for more than three years, five stated that they would post their membership in a controversial organisation, with four being definite about this response and only one stating that they "might". Only two people (of nine) who had a web site for three years or less said that they "might" post about membership in a controversial organization, both of whom stated that it would depend on the context or circumstances. The remaining seven would not post this information.

**Affiliations:** We also investigated the attitudes of people in terms of three types of affiliations: political parties, religion/church, and financial groups. Attitudes were roughly split regarding posting membership of political parties and church organizations. There were seven people who would post political affiliations, eight people who would not, and one who had no opinion. Seven of the eight people who would not post their political affiliation were older than 25 years. Also, seven of these eight had either never had a personal web site, or had one for three years or less. Only five people would not post their religious affiliations, while eleven would. Four of the five people who would not post their religious affiliation were older than 25 years, and none of the five had maintained a personal web site for more than three years. In contrast, eleven people would not post that they were associated with a financial or investment group, while four said that they "might", and one said that they did not know if they would. Three of the four who said that they might post were 25 years old or younger. This indicates that financial affiliations are considered to be more sensitive than political affiliations, with religious affiliations being the least sensitive.

**Government Data:** Subjects were then questioned about their feelings regarding information available on-line that has been put there by the government. In particular, we asked people their opinions on having property assessments and salary available[2]. For property assessment, six people said that they would make this available. Of these, two people commented that they would only make this available if required by law. Another two commented that they did not really care. One commented that there was no sense in worrying about something that was required by law. The final person commented that if he owned property on "Lake Moneybags" then he might want this information available, but otherwise not.

There was one person in our sample who was a public servant who made more than $50,000, and so had his salary available on-line. As this was required by law, he saw no reason to worry about it. For the remaining 15 people, this was a theoretical question. Given that this was required to be posted by the government, five people said that they "might" post the information, with comments ranging from preferring to see a range, to only doing this if it was the common practice, to only doing this if working in the public service. Only two of these five stated on the earlier question about releasing salary information that they would provide it. In contrast, ten people said that they either would or might let someone else (such as the government) provide this information.

**Preferences, Expressions, Creations:** Items such as personal preferences, personal expressions and creative works were then examined. Only one person stated that they would not reveal personal preferences (such as favourite books or tv shows), indicating that the majority of people do not object to having this information available ($p < 0.005$). Everyone felt the same way about other people revealing their personal preferences except one, who would prefer that others not reveal her personal information, simply because she did not like other people talking about her.

Personal expressions were defined as a person's opinions that they might post on a forum or in a blog. Ten people said that the either would or might make this available. Six of them commented that they would censor what they posted, or that they would limit access to their postings so that only friends could read it. Four of the ten stated that, while they were comfortable posting their own opinions, they would not want anyone else to do so on their behalf. There appear to be age-based differences in this category. There were four people who said that they definitely would post their personal opinions, all of whom were 25 years old and younger. Of the seven people who were 25 years and younger, only one would prefer not to post their personal opinions.

Creative works were defined as stories, photographs, paintings, or some other piece of art that had been created by the subject. Everyone stated that they either might or would post such material ($p < 0.01$). Only three of the subjects stated that they would prefer that others not post their material, even when attributed to them ($p < 0.05$).

**Photos:** Six different types of photographs were also discussed in terms of comfort level in posting: a casual snapshot, a group photo, an unattributed photo, a passport-like photo, a photo of a positive event, and an unflattering photo. In general, people were comfortable having their photo available on-line. The majority of people (twelve) were comfortable having a casual photo of them available on-line ($p < 0.05$). Eleven of these people were also comfortable having other people post a casual photo of them. People were also comfortable having a passport-like photo available of them, as well as a casual photo, with 14 people stating that they were comfortable having such a photo of them available on-line ($p < 0.005$). Twelve people were comfortable posting an unattributed photo of themselves, while thirteen people were comfortable with having someone else post such a photo ($p < 0.05$).

Thirteen people were comfortable posting a group photo where they were a member of the group ($p < 0.05$), and fourteen were comfortable having someone else post this same photo ($p < 0.005$). Similarly, fourteen people were comfortable posting a photo of a positive event (such as receiving an award), and fifteen were comfortable having someone else post such a photo. This number dropped to seven when the photo was not flattering (e.g., a blurry picture or a bad hair day). There was only one person who consistently did not want to have her photo published, with the only exception being that she recognized that sometimes group photos were required (e.g., for work).

There appear to be age-based differences in the category of photos. For all types of photos other than the unflattering ones, everyone who was 25 years old or younger were comfortable having their photo available on-line.

**Social Relationships:** The last category that this study investigated was social relationships, focusing on past partners, current partners, friends and fam-

[2]In Canada, this information is generally available, with some variations between provinces. In Nova Scotia, all public employees who make more than $50,000 per annum have their salaries posted on-line. Property assessments are also available, along with a map of the property. While they are not currently posted along with the owner's name, this is something that is done in other provinces.

ily. In general, people were comfortable divulging their friends, although this was also with restrictions, such as mentioning them by first name only. Eleven people said that they would or might post such information, while twelve said that they would be comfortable with others (such as their friends) posting this information. Eight people stated that they would or might post who was their current partner, with comments being made that only a first name might be posted, or that the partner would appear the same way that friends did, without being titled as "my girlfriend". One person stated that they might post who their current partner was as "a sign of commitment". Only four people stated that others might be able to post about their current relationship ($p < 0.05$). Again we observed some gender-based differences. For example, ten of the twelve males stated that they were comfortable posting who their friends were, while only one of the four females was. We also observed age-based differences, with six of the seven people who were under 25 year being comfortable posting who their friends were, versus five of the nine who were older than 25 years.

In general, people were more protective of information about their past partners and family. Only five people stated that they would or might post who their past partners were. Most people appeared to assume that past partners did not necessarily mean a legal commitment (e.g. marriage), but would include past girlfriends/boyfriends. There was one exception to this, who was currently divorced, and who stated that he saw no problem posting such information since it was a matter of public record. People were also protective of information about their family, with six people stating that they might post information about their family members. In contrast, nine people stated that they might allow others to post such information, often with the caveat that those who could were other family members, or that the site have access limited to just family and friends.

## 3.2 Interview Part II: General Discussion and Scenarios

### 3.2.1 General Discussion

Part II of the study consisted of open-ended questions and discussion, with the aim of determining more fully how people in general treated personal information online. It was found that everyone limited the information they posted in some form. Most often, people would not use full names of friends and they would be careful when discussing other people or groups of people (such as student societies). This was often the case even when they posted to a password-protected site, although they were more likely to post photos and journal entries here. One person stated that even when posting to online forums, she is careful to research first that the forum is not indexed by Google. We also asked whether people would post items anonymously, given the opportunity. Interestingly, opinions on posting anonymously were roughly split, with nine people stating that they would not post anonymously, and another six stating that they would use this option, particularly to post opinions.

When respondents were asked if they ever "Googled" anyone (looked up someone's name in Google), the majority responded yes, with only two people saying no. The main reason people gave for looking up other people was for contact information (seven people), for professional reasons (eight people), and out of "bored curiosity" (three people). Similarly, we asked if they had ever gone "ego-surfing", which is looking up your own name in Google. All but four people had done this, with two people commenting that they were now going to go home and try it out. Most people commented that they did this out of curiosity, and a couple of people wanted to know how much visibility they had and how they ranked on-line. One person had a problem with a "crazy ex", and so would always Google herself to see what he might be able to find about her. Another stated that she was pleasantly surprised to find some newspaper articles from when her father died. A final person commented that he was "impressed" that some material was still available.

We also asked if anyone had ever had any information removed from sites where others had posted about them. The majority (eleven people) said no. However, another four people said yes. One person had asked for their name to be removed from someone else's web site; however, the site maintainer's superior denied the request, citing freedom of speech. Two other people had made requests under more disturbing circumstances. One had asked for her name to be removed from the online telephone directory so that she could not be found by an ex-partner who was stalking her. The second needed to have her named removed from a site that listed all people who held a certain (controversial) belief (e.g. was pro-choice, agreed with gay marriage, etc.). The purpose of the site was specifically to list such people so that they could be harassed. In order to do so, she needed to go to the hosting company of the site. In the last example, someone stated that she had gotten "pissy about creative content in terms of recipes", where she had made formal requests to people who had posted recipes that they had copied from her forum to either remove the recipe or attribute the source.

For contrast, we also asked if the subject had ever been asked to remove anything that they had posted. The majority of subjects (eleven people) have not had this experience. Two of the three who had said that they had been asked to remove information said that they had done so. In the first case, an old room-mate had wanted his name removed from the subject's website. Since there was a strained relationship between the two (and the subject did not particularly want to be associated with the room-mate at all), he agreed to remove the name. In the second case, a moderator had left a particular on-line forum under bad circumstances, and wanted all of her posts removed. The subject, who was the new moderator, removed anything of the previous moderator's where there was original content, but refused to remove any posting that consisted only of quoted or forwarded content. In the third case, the subject had refused to remove information regarding a spat between friends.

Given that people have a general concern with personal information being posted in a public on-line forum, we asked about what kind of control mechanisms should be in place, if any: social, technical, and/or legal. There was no consensus on this issue, with two people saying that the controls should be technical, four saying that they should be legal, two saying that they should be social, two saying that it had to consist of all three mechanisms, and another saying that social does not work but that she can not see a practical way to do this. Similar to this last person, another subject commented that people do not self-police very well, particularly in large groups. This same person felt that everybody should be able to publish whatever they want about themselves, but not other people. This seemed to be echoed by two others, who focused on control of this information. One commented that you should be able to protect anything that is wholly yours, while the other stated that he would like to allow some people to see more than others, and that he would like to know who has accessed his information.

#### 3.2.2 Scenarios

We provided subjects with four different scenarios, and asked how they would respond to each of them. The first scenario consisted of them posting a wedding photo of themselves with the best man. The best man emails and asks to have the photo removed. Nine people stated that they would take down the photo, with one commenting that he would email back and apologise for not having asked permission first. Two people said that they would move it to a private (e.g. password-protected) section of their website, if they thought the request was reasonable. The remaining four people commented that they would use image editing software to remove the best man from the photo, and then leave that posted.

The second scenario consisted of a friend who posted a story about a road trip taken with the subject, where the subject is mentioned by name several times (and where it was specified that there was nothing embarrassing about the account). Eleven people responded that they would be fine with this, although one commented that he would be slightly uneasy about it since he finds blogs odd. Two people stated that they would be comfortable if it was only their first name, but not their last. One commented that she would ask them to remove her name "if it was obvious it was [her]", while another commented that he would ask the person not to use his real name.

In the third scenario, a consultant at a job search agency is using Google to determine more information about potential employees, other than what was solely posted on their resume. Ten people stated that they thought this was fine, with comments such as "You can't stop them!" and "People want to know who they are hiring, so why not?" Two of the respondents commented that "if you posted something then it was so people will find it." However, some of the respondents also noted that they would hope that this was not the consultant's only method of finding information. Three people commented that this type of search was "not a good idea" as it could return inaccurate information, and one stated that it was not ethical.

The view that it would be nice to know who has accessed your personal information, as mentioned in the general discussion section, was also reflected in subject's discussion of the fourth scenario. This scenario asked about people reactions to finding out that a coworker was gossiping about them, providing information available through government sites (such as salary and property assessments). While responses generally varied, several people commented that if this information was going to be available on-line, that logs of who accessed it should also be kept. There were also two people who commented that it was "disturbing" that some of this information is publicly available.

## 4 Discussion

When the average scores were calculated for each person, the values ranged from 1.79 to 3.36. The frequency distribution of personal preference scores on a per person (versus per question) basis are provided in Figure 3. Scores were binned using bins of size 0.2. We found two people with average scores that were less

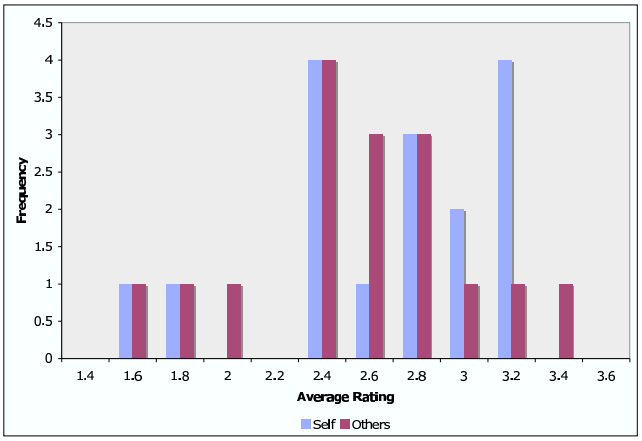


**Figure 3. The frequency distribution of people's comfort with posting personal information.**

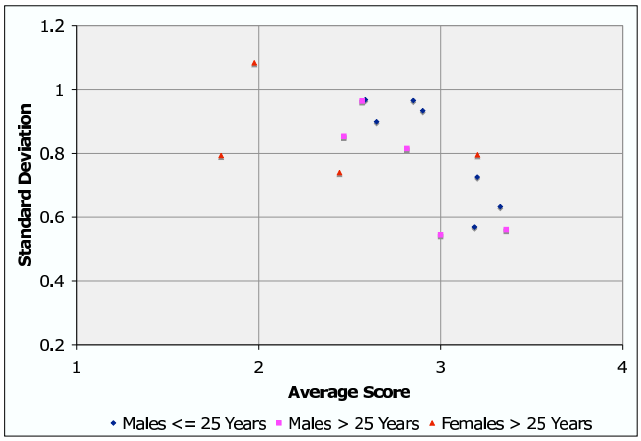


**Figure 4. A scatter plot comparing the average score for each person to his/her standard deviation.**

than two, who seem to fall in the category of privacy fundamentalist [3]. While it is less obvious from Figure 3 how to divide the remaining subjects into privacy pragmatists and the marginally concerned, clusters become a bit more obvious in Figure 4. This shows a scatter plot that compares the average score to the standard deviation. From this plot, it appears that the next cluster ranges from approximately 2.4 to 2.9, which forms the privacy pragmatists group. The last cluster has an average scoring ranging from 3 to 3.5, and forms the marginally concerned group. There are eight people in the pragmatist group, and six in the marginally concerned group. This distribution shows the same trends to those found by Cranor *et al.* [1], who classified 17% of their respondents as privacy fundamentalists (versus 12.5% for us), 56% as privacy pragmatists (versus 50% for us) and 27% as marginally concerned (versus 37.5% for us).

It is interesting to examine Figure 4, due to the clusters that it shows. The marginally concerned group not only exhibit the highest scores of the three groups (by definition of the group), but also show the lowest variance, with standard deviations ranging from 0.54 to 0.80. The privacy pragmatists, however, demonstrated overall more variability, with standard deviations ranging from 0.74 to 0.97. The two privacy fundamentalists had very different variances in their answers, with one person having a standard deviation in their responses of 0.79 and the other 1.08. This would indicate that the marginally concerned are the most consistent about allowing their personal information to be posted in a public forum, whereas the privacy pragmatists are more likely to a vary in what information they will post and under what circumstances.

There are both age and sex-based differences in categories. It has been noted that female Internet users tend to be more concerned about their personal privacy than males [4]. We observe the same phenomenon, although weak. The average score for the men was 2.91 while it was 2.35 for the women. Additionally, the only two privacy fundamentalists were both women. Interestingly, the standard deviation in scores was twice as high for women as for men (0.63 versus 0.31). This is reflected in the observation that one woman was a privacy pragmatist while one was marginally concerned (along with the two privacy fundamentalists), so that women spanned the entire spectrum of privacy categories.

The age-based differences in responses is weaker, but still present. In general, subjects who were 25 years and younger were more likely to publish their personal information, with an average score of 2.96. In comparison, those who were 26 years and older had a lower average score: 2.63. Again there was a considerable difference in the standard deviations of the two groups, with the older group having twice the variance of the younger group (0.53 versus 0.26). This indicates that those 25 years and younger were more likely to cluster around the boundary of privacy pragmatist and marginally concerned. The older generation was more likely to span the range of all three classifications.

Figure 5 shows a scatter plot comparing the average score to standard deviation for each of the individual questions, rather than for each person. Two interesting observations can be made from this graph. The first is that there is a small cluster of questions that

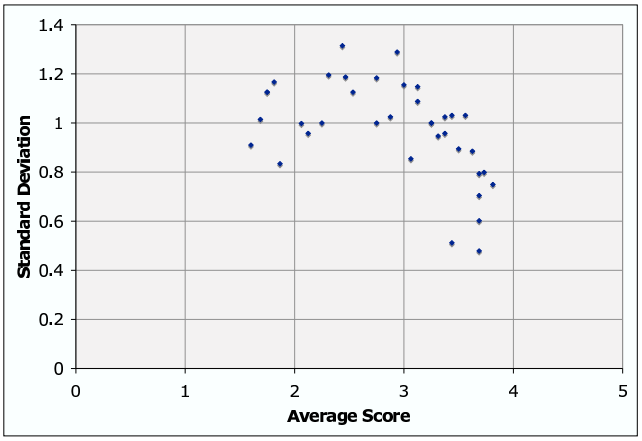


**Figure 5. A scatter plot comparing the average score for each question to its standard deviation.**

have a high average score and a low standard deviation. This indicates that there is some information that the majority of people feel comfortable posting on-line. There are seven types of information with an average score that is greater than 3.0 with a standard deviation of 0.8 or less: sex, work address, work email, job title, publications list, favourites (e.g. books, tv shows), and creative works. The other cluster forms an upside-down U-shape, where those scores between two and three had the highest standard deviations, and the scores less than two or greater than three had lower standard deviations. While all of the standard deviations in this group were greater than 0.8, the majority (10 out of 11) of those for scores between two and three were $\geq 1.0$, with the one exception having a standard deviation of 0.96. The average standard deviation for scores between 1 and 2 was 1.029, between 2 and 3 was 1.119, and for scores between 3 and 4 was 0.868.

In general, people were most comfortable providing two types of information: professional data and generic personal data. This latter category includes social and leisure data (such as hobbies and sports), preferences and expressions (such as opinions and creative works), photos and vital statistics (such as marital status). The scores in all of these categories were 3.00 or greater, with professional data having the highest score at 3.36.

The three types of information that people were most comfortable posting were their job title (average of 3.81 and a standard deviation of 0.75), their publication list (average of 3.73 and a standard deviation of 0.80), and their work email (average of 3.69 and a standard deviation of 0.48). All of these fall under the category of professional data.

In contrast, people were least comfortable providing personal contact information, affiliations (social or otherwise) and government data. The highest score in any of these areas was 2.43, with a standard deviation of 1.19. The three types of information that people were least likely to publish were their salary (average of 1.60 and a standard deviation 0.91), driving directions to their home (average of 1.68 and a standard deviation of 1.01) and home address and phone number (average of 1.75 and a standard deviation of 1.13). Most of these fell in the category of personal contact information, which had an overall score of 2.02 and a standard deviation of 1.16.

Preferences and expressions was the category with the smallest standard deviation (0.87) and the second highest score (3.29). When each individual question was analysed, the three with the least variance between subjects was work email (average of 3.69 and a standard deviation of 0.48), creative works (average of 3.44 and a standard deviation of 0.51) and work address (average of 3.69 and a standard deviation of 0.60). Creative works falls into the category of preferences and expressions. Interestingly, the other two questions fall into professional data, which had the third smallest standard deviation (1.03) , but the highest score (3.36).

The three sub-categories which demonstrated the greatest variance between subjects were unflattering photos (average of 2.44 and a standard deviation of 1.31), church or religious affiliations (average of 2.94 and a standard deviation of 1.29), and current partner (average of 2.31 and a standard deviation of 1.20). This indicates that there were greater differences of opinion on whether these topics would be posted. The scores also indicate that people are largely split on whether to post this information.

Of the 37 sub-categories, there were 28 sub-categories where people were either as comfortable or more comfortable releasing information themselves than having other people publish that same information. The nine categories where people were more comfortable having others publish the information were qualifications (average of 3.44 versus 3.38), property assessments (2.2 versus 2.13), public-service salaries (2.5 versus 2.06), group photographs (3.44 versus 3.31), unattributed photographs (3.44 versus 3.38), photographs of a positive event (3.56 versus 3.50), unflattering photographs (2.50 versus 2.44), information about friends (3.25 versus 3.13) and information about family (2.50 versus 2.25). These can be roughly grouped into three categories: something work might publish, something friends or family might publish, and something the government might publish.

Information published by the government deserves special mention. Although our study was primarily concerned with unregulated data, this was included to gauge people's reactions to information that is already publicly available, and to see how that might influence their responses. This category had an average score of 2.09 with a standard deviation of 0.96, indicating that people would generally not publish this information themselves. This score increased to 2.36 with a standard deviation of 1.14 when asked about others publishing this information. This indicates that, while people are more comfortable with the government publishing it than putting it on-line themselves, that opinion is still divided on the government making it available.

Investigating further, the average score for releasing salary information on-line was 1.6 for self-revelation, and 1.53 for others to reveal it. However, when it was specified that the salary came from public service, the scores increased to 2.06 and 2.5 respectively. This indicates that people consider their salary information to be very personal, but that they had mixed feelings about releasing it for those in the public service, with many people specifying that a range should be used.

We followed up on this information with a discussion question about what the government or workplace can publish. The responses here varied, with one person stating that most things in the government must be transparent, which would include salary information, and another stating that he was not as concerned with salary as he was with having addresses and photos available. Others were more pragmatic, stating that they would prefer that it was not put on-line, but that you often have to "suck it up". Further along the continuum, another person said that it was okay to put up information except for salary, while a second person said it was okay to put up information except for home address or phone. One person commented that the workplace and government should recognize that not everyone is comfortable putting some items on-line, and that doing so sometimes felt like bullying.

In general, people seemed most concerned with salaries and contact information being posted. This led some people to comment that the government should more carefully control the release of information. For example, one person commented that others should only have access to your salary if they need it, but that it shouldn't be posted for the general public as "this information is used in a negative way". Another person commented that Thailand had a good model, where you could access anything the government has unless it contained personal information; for example, property assessments would not be on-line, unless the property is owned by a public, political figure such as the prime minister.

### 4.1 Comparison to Prior Privacy Study

We had previously run a pilot study that examined privacy issues of on-line information, as well as its temporal nature. That study had 16 respondents (all from Dalhousie University's School of Library Science), who agreed to answer a short survey. Some of the questions from that pilot study are also present in this study.

In the previous study, respondents were asked to state if they would be uncomfortable, comfortable, or don't care, about publicly releasing different types of information. In general, the respondents felt comfortable releasing the work address and work phone, and did not care if they released their hobbies (clubs or sports), their resume, or their favourite books, movies or food. These results are consistent with the results from this study.

In the previous study, most respondents (out of 16) were uncomfortable releasing information about their home address (12 people), home phone (10 people), and driving directions to their house (12 people). They were also uncomfortable releasing their salary (12 people) or any memberships in financial organisations (10 people). This is consistent with our results, where the average scores ranged from 1.60 to 1.87 for these categories.

Agreement between the two studies was not as complete when asked about attitudes towards photos. For head-shots (passport-like photos), photos in which the respondent is not identified, and photos of the respondent receiving a reward, our study found that the respondents were largely comfortable releasing this information, with average scores ranging from 3.25 to 3.50. In contrast, our previous study saw less consistency in the results, with seven, five and four people, respectively, stating that they would be uncomfortable releasing this information (versus two, four and two, respectively, in our study).

There was also some disagreement between the two groups when examining affiliations. In the previous study, the majority of respondents were uncomfortable releasing information about their political or religious affiliations, with 12 and 10 people, respectively, stating that they would be uncomfortable. In comparison, the average scores in this study were 2.47 and 2.94 respectively, and both with a reasonably high standard deviation (1.19 and 1.29 respectively). This shows that the respondents in this study were more divided on the issue.

## 5 Comparison to Related Work

There have been a number of studies that focused on user attitudes towards personal data collected by online businesses, such as Miyazaki and Fernandez's 2001 survey [5], Fox *et al.*'s report [2], and ThePrivacyPlace.org's online survey [7]. As with our work, these studies present the concerns that people have over having certain type of data revealed; however, the context in which this personal data is provided is quite different from a public forum. In addition, corporations have specific legal obligations in handling protected data, which makes the limitations on disclosure much more stringent.

The work that is most closely related to ours was published recently as a technical report by Olson *et al* [6]. They explore what information people consider to be personal, and with whom they would share it. The information space to be explored was generated by asking the employees at a company and a university department to provide any incidents where they had provided information that they later regretted sharing. In particular, Olson *et al.* were interested in the type of information and the relationship with the person with whom it had been shared. Based on this, they generated a grid of people (relationships) versus types of information. They asked 30 people to fill in the grid with a value on a scale of one to five that indicated how comfortable they would be sharing this information with this person.

While the approach by Olson *et al.* [6] resulted in more specific sets of information that was more likely to be considered personal than our approach, there is some overlap. Additionally, a range of different types of people were identified by Olson *et al.*, while we focused on public sites. The group that most closely relates to our study was "Your personal website/blog", identified by Olson *et al.* When we compare the comfort level of all of our participants across all types of information, we find that people are generally comfortable revealing information in an on-line public forum (3.17 out of 4 with a standard deviation of 1.06). In contrast, Olson *et al.* found that participants were unlikely to post information to a website or blog (1.76 out of 5 with a standard deviation of 0.45). This difference is likely due to the types of information used by Olson *et al.* being generally more personal that ours.

Some of the information categories in which there is overlap include salary, "preferences" (which appears to overlap with political and religious affiliations and friends), "past finished papers, products" (which is similar to a publications list), home phone number, age, marital status, "desk phone number" (assumed to be work phone number), and work email address. The only categories here that Olson *et al.* found people might post to a website or blog was finished papers (2.61 average out of 5), age (2.89), marital status (2.83) and work email address (2.56). In contrast, our sample was far more likely to publish this information, with scores of 3.73 (out of 4), 3.00, 3.25, and 3.69 respectively. We postulate that this might be due to a combination of the characteristics or our target audience (people who have five or more years of experience on-line, most of whom have web pages and who have posted to newsgroups and web forums) and the difference in allowing for people to place restrictions. That is, Olson *et al.* measured a strict level of comfort, whereas we had a different interpretation to our scale. Thus subjects in our study could say that they "might" put something on-line (a value of 3 out of 4), where they would follow this up with a comment that they would put a range instead of an actual age, or that they might password-protect a site.

Cranor *et al.* [1] investigated users privacy preferences in the context of providing this information to another organisation using the web via an on-line form. They found that respondents were generally comfortable providing their personal preferences (such as favourite television show or snack food), which is consistent with our results. Additionally they found that few people were comfortable providing their income, which is consistent with the reluctance we found to publicly disclose salary information. However, Cranor *et al.* found that a large number were usually or always comfortable revealing their email address (76%) and their age (69%). This is a bit higher than what would be expected given our data, where the average score for revealing a personal email address was 2.875 and for age was 3.00 (where a score of 3 represents that they "might" reveal this information).

Cranor *et al.* [1] also found that Internet users did not like receiving unsolicited communications. This is consistent with some of the comments we received when asking about the comfort level of users in posting their personal email address on the web, where a number of respondents indicated that their primary concern in doing this was spam.

## 6 Conclusions

This paper presents the results from a study on the control of personal information that is not legislated. It is a continuation of an earlier, survey-based study and consisted of interviews of 16 people. The interviews consisted of two parts, with the first focusing on determining comfort levels to posting different types of

personal information on-line. The second consisted of open-ended questions designed to explore the responses of interviewees more thoroughly.

We found that people can be grouped into roughly three categories, similar to those found in previous work (see [3] and [1]). Those who were the least concerned with privacy also exhibited the least variation in their answers, but would consistently publish information on-line. In contrast, the privacy pragmatists exhibited greater variation in their responses, and the privacy fundamentalists exhibited the greatest variation. It should be noted, however, that this is only an observation, as our sample size is too small to determine significance.

We also observed that subjects tended to respond consistently within different groupings of information. For example, we found that overall people were comfortable releasing professional data (such as their contact information at work, resume, and qualifications), and that they were equally comfortable releasing any one part of the data that was included under professional data, other than salary. This type of result implies that it should be possible to group related information.

It should be noted that this does not imply that a one-size-fits-all policy can be developed. For example, there were only two types of information (out of 37) where all respondents said that they either might or would definitely post the information: work email and creative works (such as poems or photographs). There was no information type that everyone agreed they would *not* post. We also observed some sex-related and age-related differences in some information being posted. In addition, we wish to stress that majority opinion cannot be used as a guide to appropriate behaviour. Although many of our respondents were comfortable having photos posted, for example, a privacy fundamentalist was strongly opposed to this action. Such minority views must be taken into account when considering privacy issues.

We hope that studies such as this will be used to influence privacy policy. For example, journal sites might use this to develop a set of standards for appropriate behaviour in their community. At the very least, we wish to begin the debate over how to best to negotiate and control personal disclosure in public online spaces.

Additionally, we hope that it can inform technical developments. With the proliferation of website publishing tools, we foresee the need to develop access control mechanisms that will help protect the user from accidentally publishing personal information. This study can help determine appropriate groupings of information and people to allow a more coarse-grained but light-weight access control. Such control can extend to monitoring users when they post to on-line forums or blogs to ensure that they are not violating any of their own policies. Such a system can act as a "reminder", without actually preventing publication.

## 7 Acknowledgments

The authors would like to thank the IBM Canada Centre for Advanced Studies in Toronto, the National Sciences and Engineering Research Council of Canada (NSERC), and the Killam Scholarship Fund (Dalhousie University).

## References

[1] L. F. Cranor, J. Reagle, and M. S. Ackerman. Beyond concern: Understanding net users' attitudes about online privacy. Research Technical Report TR 99.4.3, AT&T Labs, April 1999.

[2] S. Fox, L. Rainie, J. Horrigan, A. Lenhart, T. Spooner, and C. Carter. Trust and privacy online: Why Americans want to rewrite the rules. The Pew Internet and American Life Project. `http://www.pewinternet.org/pdfs/PIP_Trust_Privacy_Report.pdf`, 2000.

[3] Harris, Louis and Associates and A. F. Westin. *Harris-Equifax Consumer Privacy Survey 1991.* Equifax Inc., Atlanta, GA, 1991.

[4] Harris, Louis and Associates and A. F. Westin. *E-commerce and Privacy: What Net Users Want.* Privacy and American Business, Hackensack, NJ, 1998.

[5] A. Miyazaki and A. Fernandez. Consumer perceptions of privacy and security risks for online shopping. *Journal of Consumer Affairs*, 35(1):27–44, 2001.

[6] J. S. Olson, J. Grudin, and E. Horvitz. Toward understanding preferences for sharing and privacy. Research Technical Report MSR-TR-2004-138, Microsoft Research, December 2004.

[7] ThePrivacyPlace.org. 2002 Internet Privacy User Values Survey. `http://william.stufflebeam.cc/privacySurvey/results/resultsPage.php`, 2002.

[8] T. Whalen and C. Gates. Public perception of private information on search engines. Research Technical Report CS-2005-01, Dalhousie University, January 2005.