

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

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

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

## 1. INTRODUCTION

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.

This abstract describes 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.

## 2. DESCRIPTION OF STUDY

We conducted a semi-structured interview in 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. 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

The same set of responses was then provided for revelation of personal information by other people. 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 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.

## 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, where scores  $\geq 3.0$  indicate a willingness to release data.**

In general, people were comfortable providing personal preferences and expressions. People were also generally comfortable posting photos, as well as personal data (e.g., vital

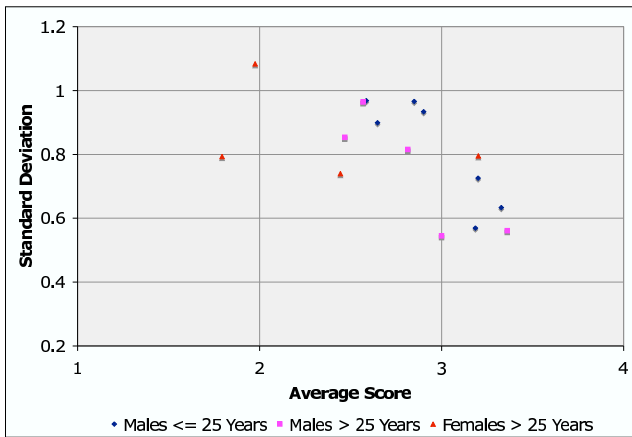


Figure 1: A scatter plot comparing the average score for each person to his/her standard deviation.

statistics) and professional data. The score for professional data increased even further, from 3.36 to 3.57, when the question about releasing salary information was removed from consideration. This also had a strong effect on the standard deviation, dropping it from 1.03 to 0.822. On the opposite end of the spectrum, people were not comfortable releasing their personal contact data (with the exception of their email address), government data (e.g., property values), and affiliations (e.g. political, religious, social).

We also found that 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.

#### 4. DISCUSSION

When the average scores were calculated for each person, the values ranged from 1.79 to 3.36. We found two people with average scores that were less than two, who seem to fall in the category of privacy fundamentalist [2]. Figure 1 shows a scatter plot that compares the average score to the standard deviation, demonstrating obvious clusters. Labeling the middle cluster as privacy pragmatists, and the right-most cluster as marginally concerned, our distribution shows the same trends as those found by Cranor *et al.* [1]. It is also interesting to note that the marginally concerned also demonstrate the lowest variance of the three groups. 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 vary in what information they will post and under what circumstances.

Figure 2 shows a scatter plot comparing the average score to standard deviation for each of the individual questions, rather than for each person. We note that there is a small cluster of questions that have a high average score and a low standard deviation, indicating that there is some information that the majority of people feel comfortable posting on-line. The other cluster forms an upside-down U-shape,

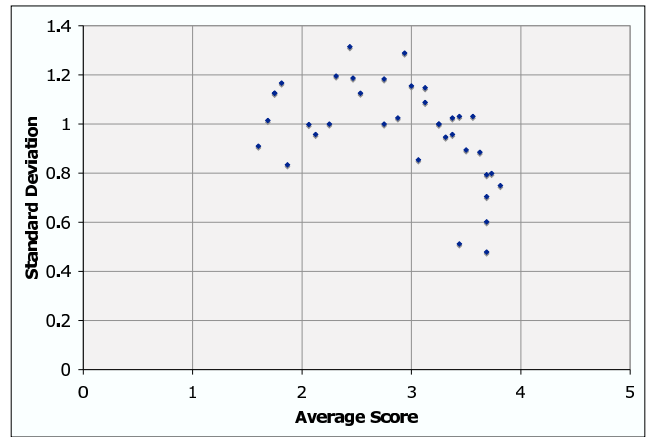


Figure 2: A scatter plot comparing the average score for each question to its standard deviation.

where those scores between two and three had the highest standard deviations (1.119 on average), and the scores less than two or greater than three had lower standard deviations (1.029 and 0.868 respectively). This indicates there may be classes of information where people are reasonably consistent about whether it could be published.

#### 5. CONCLUSIONS

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 (social or otherwise), and government-released data (such as property assessments).

We observed that subjects tended to respond consistently within different classes of information. For example, people were generally comfortable releasing professional data (such as their contact information at work, resume, and qualifications), and 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. 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.

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