A Study of Preferences for Sharing and Privacy

Judith S. Olson
University of Michigan
550 E. University
Ann Arbor, MI 48109-1092
jsolson@umich.edu

Jonathan Grudin
Microsoft Research
One Microsoft Way
Redmond, WA 98052-6399
jgrudin@microsoft.com

Eric Horvitz
Microsoft Research
One Microsoft Way
Redmond, WA 98052-6399
horvitz@microsoft.com

ABSTRACT
We describe studies of preferences about information sharing aimed at identifying fundamental concerns with privacy and at understanding how people might abstract the details of sharing into higher-level *classes* of recipients and information that are treated similarly. Thirty people specified what information they are willing to share with whom. Although people vary in their overall level of comfort in sharing, we identified key classes of recipients and information. Such abstractions highlight the promise of developing expressive controls for sharing and privacy.

Author Keywords
Information sharing, privacy, perceptions of trust

ACM Classification Keywords
H1.2 User/Machine Systems, *human factors;* H.5.2 User Interfaces, *User-centered design.* H.5.3 Group and organizational interfaces, *collaborative computing.* K.4.1. Public policy issues, *Privacy.*

INTRODUCTION
Technical advances have generated concerns about violations of privacy. People fear that they will not have control over who knows what about them. Willingness to share may vary with the type of information, who will see it, and reason the data is being sought [1, 2, 3, 8].

Surveys of privacy concerns generally focus on information disclosure to online retailers, not on sharing and privacy in workplaces or other settings (see [9] for a more extensive literature review). Information sharing in the workplace reduces duplication of effort. It is a key motivation for digitized content and networked computing. It is reasonable to expect the concerns to be different in the workplace than in the more public sector.

CHI 2005; April 2-7, 2005, Portland, Oregon, USA
ACM 1-59593-002-7/05/0004.

Our program of research asks several questions: What are people's concerns with sharing information? How do people differ, where are they in agreement, what kinds of people and kinds of information do they treat similarly and differently? Can we derive a small set of questions that provide an indication of someone's preference pattern? Could we provide people with shrewd guesses as to their access choices, which they could then modify with ease? Could we provide interfaces that allow people to make and maintain access control settings in context, when they best know how they feel?

OVERVIEW OF THE STUDY
We engaged in a two-phased study. We started with an exploratory phase: We first asked a set of people to relate various instances of when they shared something that they later regretted sharing. We identified all the pieces of information people regretted sharing and the kinds of people with whom they shared that information. In the second phase of the study, we chose 40 of these kinds of information that were shared or not shared with 19 types of people. We asked 30 people from varied backgrounds to rate *each* kind of information as to how comfortable they were in sharing it with *each* kind of person, no matter what their intent.

METHOD
General survey
To begin exploring the general issue of what kinds of items people are sometimes reluctant to share, we conducted a pilot survey asking respondents to provide examples "of a situation in which you or another person did not wish to share information. Include: 1) A description of the information and situation; 2) Why sharing would have been uncomfortable." The on-line survey was distributed to a few hundred usability engineers and researchers at a large software company and the students, faculty and staff at a computer-centric department at a major university. Both organizations were based in the United States.

We obtained 170 examples from 83 people. The responses provided a wide range of situations in which people had either bad experiences or simple qualms about

sharing information. For example, people described sharing early work drafts with people who then thought badly of them for sloppiness. Others shared home phone numbers only to be bombarded by telemarketers. People included family members as recipients of information (*e.g.*, sharing a report of an automobile accident with grandparents) who then thought badly of them, and trusted and competitive co-workers as well as the general public, a company website or one's personal website. They named information types like personal statistics (*e.g.*, age, Social Security number, salary, marital status) as well as more work-related objects (*e.g.*, working drafts, a complete list of finished work products, the history of their performance reviews), and health related information (*e.g.*, pregnancy status, and general health issues). They named information that is stable (*e.g.*, Social Security Number) and information that is dynamic (*e.g.* one's location), and some in between (*e.g.*, one's health status) [7]. Although we did not exhaust all possibilities, the degree of overlap among people suggested that we had captured a useful core set. These responses served as the basis for the items in the formal survey.

Assessing Detailed Sharing Preferences

From this set of narrative situations, augmented with some from our own experiences and work we have done with prototyping various policies for sharing, we chose 40 types of information and 19 types of people. At the beginning of each session, participants filled out a questionnaire covering basic demographics and nine questions from a standard scale measuring basic trust of the world [4]. Then, we gave participants an empty table to fill in. They were asked to fill in each cell, indicating on a scale of 1 to 5 how comfortable they would be with sharing each particular type of information with each type of person without knowing what that person would do with it. They were to instantiate each type of person with someone in their current life, putting an "N/A" in the cells that were inappropriate, either because no such person existed (*e.g.*, "adult child"), or because that kind of information was not part of their life (*e.g.*, "desktop video conferencing number").

Given the size of this questionnaire, we asked people to strike out the rows and columns that did not apply, and to concentrate on the remaining cells one by one, either by rows or columns. People filled out the questionnaire with a mixture of going down columns and across rows, following their own strategy. We made this a paper survey in an effort to make it as easy and speedy as possible for the participants.

Our pilot testing showed that it required about 75 minutes to complete the grid. We recruited participants as if for a usability study to come to the lab for two hours, in groups of 1-6, to fill out the grid and then discuss sharing and privacy issues with us. In turn, they were given a standard gratuity for participants in two-hour user studies at our organization. No participant manifested or reported difficulty in filling out the form, and many reported finding the task interesting. The participants worked at mid-sized companies and used computers as part of their jobs; they were recruited from a participant panel. Thirty participants filled out the grid. Twenty-one of the 30 were males; 9 females. The median age was 35. Companies ranged from 20 to over 150,000 employees. Their occupations spanned a wide spectrum, including social worker, CIO, materials manager, real estate, and project manager. This sample was intended to survey people with experience with sharing information with team-mates, managers, family members, and others.

Two participants who missed some items (*e.g.*, leaving a row blank) were re-contacted and asked to complete the items. Of the 30 grids of nearly 800 cells each, we ended up with only about 20 blank cells.

Analysis

From the individual grids of ratings, we created several summaries. We computed the mean ratings over all 30 participants as well as the standard deviations of these values. Items that were left blank or marked as "N/A" were considered to be blanks. To facilitate visualization, the columns and rows were ordered left-to-right and top-down from lowest (least likely to be shared) to highest (most likely), and color coded to reveal the bands of opinions. We created a visualization for each participant and for the average for the whole group (see [9] for detailed renderings).

To see how information items and people with whom to share clustered, we separately performed a hierarchical cluster analysis on the rows and then on the columns. This analysis uses a Euclidean distance metric to assess the similarity of pairs of rows (or columns). The more alike the items are rated across the rows (or columns), the closer they are in the hierarchy. The hierarchy thus shows items that cluster, with those coming together near the leaves of the tree being more similar than those joining the cluster closer to the root [5]. Note that this does not illustrate whether or not items are shared, just how similarly they were treated by the participants.

After performing two cluster analyses on each participant's ratings, we did two cluster analyses on the *averages*, again one for the information and one for the people. A Principal Component Analysis determined the number of clusters. For *information*, the first three components covered 94% of the variance. For the *people*, the first three components covered 95% of the variance. As some of the clusters were large and others small, we expanded the number of clusters of information items to

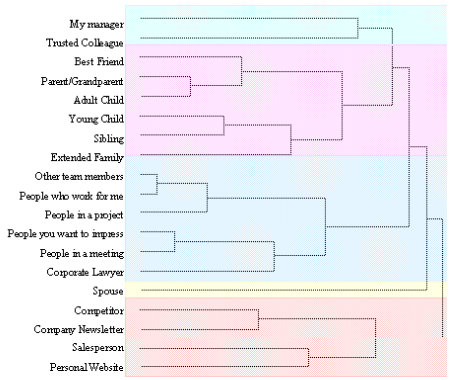

Figure 1. Clusters of people treated similarly.

be six, and the people these items were to be shared with to five. In addition, this clustering was informed by performing the cluster analyses using five methods for joining items to clusters: Average, single, complete, centroid, and Ward. A surprising number of clusters were the same in all solutions. We identified finer-grained clusters (three more in the information hierarchies and two more in the people hierarchy) from the differences in the solutions from the different methods. Later, we looked at the average variance within the clusters and found the small standard deviation (0.32) to be consistent with a discovery of stable categories.

RESULTS

We found that the overall average rating was 2.82 (from 1 to 5), with the average standard deviation at 1.46. We found that the participants do not want a transgression made public or their email to be widely shared, but are comfortable with people having their work email address and desk phone number (details are provided in [9]).

Some ratings were quite variable across people, while others showed low variance. We found zero variance in always sharing one's work email and work phone number with one's spouse and coworkers, always sharing one's home phone number with one's spouse and children (but not always with co-workers), and never giving a credit card number to the public. The highest variance (std>1.5) centered around various personal items being shared with co-workers, including sharing one's age with a competitor, one's pregnancy status with other team members and one's marital status in a company newsletter. Other high-variance items centered on sharing one's credit card number with one's parents or grandparents, and sharing one's pregnancy status with a sibling. Overall, the most disagreement came in rating one's personal statistics, with more disagreement about sharing them with coworkers than with family members. Similar high variance appeared in the ratings of work-related documents with family members, perhaps reflecting judgments of appropriateness (*i.e.*, they wouldn't care to

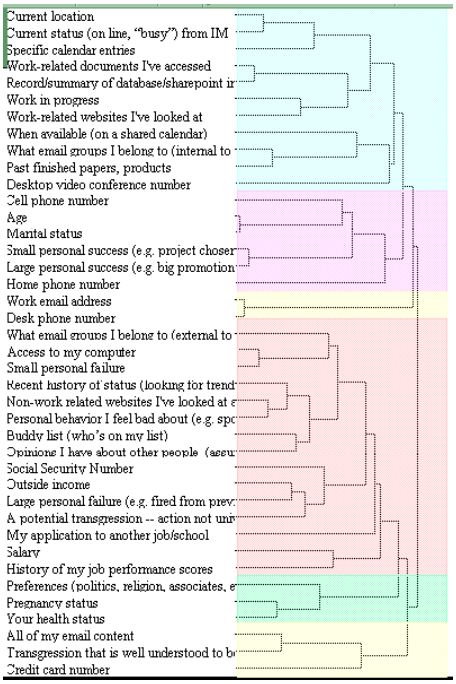

Figure 2. Clusters of information treated similarly.

see them rather than a desire to exclude).

People varied among themselves as to how willing they were to share things (see [9] for details). Like findings discussed elsewhere [13], we found some "privacy unconcerned" participants who prefer to share the most, many "privacy pragmatists," and some "privacy fundamentalist" who like to share the least.

Figures 1 and 2 displays the results of the hierarchical cluster analysis of average ratings for people and information respectively. We found that people cluster into the following: public (websites, telemarketers) and a competitor; coworkers, including the corporate lawyer; manager and a trusted co-worker; family; spouse. Of interest in this analysis is how far out the manager and trusted coworker join the work-life cluster, and how far out the spouse joins the family cluster, indicating that they are treated unlike the others. However, we found that managers and spouses are dissimilar. This result may be based in managers having access to some information (*e.g.*, the participants' salary) *ex officio*, whereas a spouse has information based on a trusted partnership.

The information items also clustered into crisp categories: Access to all your email content, your credit card number, and a transgression; Failures, opinions, salary and outside income, Social Security Number; Home and cell number, age and marital status, and successes; Pregnancy, health and preferences (religious, politics); Work related documents, websites, availability; Work email and desk phone number.

DISCUSSION

We found that peoples' willingness to share depends on who they are sharing the information with. Participants' information items clustered into a manageable set of categories, and most peoples' view of others they wish to share this information with is similarly clustered into a manageable set of categories.

We believe that such findings can provide guidance to the design of access controls and interfaces, that could make specification easier for the end user. We foresee designs that allow for control of the grain size of definitions of groups of people and types of items. For example, there is promise in creating designs for specifying preferences that reveal, in a selective manner, finer-grained choices within a hierarchical scheme, allowing users to navigate to the level of precision they are most comfortable with. For example, a preference-specification tool could allow users to specify in general their permissions per category of person (*e.g.*, the public, high level people in your organization, co-workers, your family, your manager, your spouse, etc.), but make an exception for one particular person or a particular information type. Or, a automated privacy agent with such specifications could, with some content analysis, detect your email address, SSN, or personal facts in the document, automatically set the appropriate permissions, perhaps after engaging in a confirmatory dialog with the user. When people ask to access a file, the permission scheme could assess what type of person they are, and then either grant or deny access.

Others [4, 6, 10, 11, 12] have explored access control methods that can be populated with the abstractions of people and information that we have found. For example, rather than simply granting or denying access, additional sharing actions could be provided. For example, it may include a policy of *informing* the person requesting access, at the time of an attempted access, that there is an audit trail of accesses—and then logging the accesses for the owners' later review. Another potential policy, giving users moment-by-moment control, is to provide selected groups of people with a mechanism for easily requesting permission to access information of specified types. This would accommodate changes in willingness to share without requiring changes to global settings.

Beyond direct specification, there is opportunity to leverage the type of data we collected in our study within statistical recommender systems. Such systems can be viewed as performing dynamic cluster analysis of users based on a partial specification of preferences. Such systems could be deployed with the goal of providing guesses about sets of preferences with regards to sharing on a people and items bases, and then allow users to refine the guesses.

We believe that this research and follow-on studies will serve to inform designs for efficient languages and tools that allow users to specify, and refine over time, what they wish to share with whom.

REFERENCES

1. Ackerman, M. S. (2000). Developing for privacy: Civility frameworks and technical design. *Proc. Computers, Freedom, & Privacy,* 19-23.
2. Ackerman, M. S., Cranor, L. F., and Reagle, J. (1999). Privacy in e-commerce: Examining user scenarios and privacy preferences. *Proc. ACM Conference on Electronic Commerce,* 1-8.
3. Bellotti, V., & Sellen, A. (1993). Design for privacy in ubiquitous computing environments. *ECSCW-93,* 77-92.
4. Butler, J. K. (1991). Toward understanding and measuring conditions of trust: Evolution of a conditions of trust inventory. *Journal of Management, 17,* 643-663.
5. Corter, J. E. (1996). *Tree models of similarity and association.* Sage University Paper.
6. Ferraiolo, D., Duginin, J. A., & Kuhn, D. R. (1995). Role based access control (RBAC): Features and motivation. *11th Annual Computer Security Applications Conference.*
7. Lederer, S., Beckmann, C., Dey, A., & Mankoff, J. (2003). Managing personal information disclosure in ubiquitous computing environments. *Intel Research Berkeley Technical Report 03-015.*
8. Lederer, S., Dey, A. K., & Mankoff, J. (2003). Who wants to know what when? Privacy preference determinants in ubiquitous computing *CY* 2003 Shortpapers,* 724-725.
9. Olson, J.S., Grudin, J., & Horvitz, E. (2004). Toward understanding preferences for sharing and privacy. MSR Technical Report 2004-138. ftp://ftp.research.microsoft.com/pub/tr/TR-2004-138.pdf
10. Povey, D. (1999). Optimistic security: A new access control paradigm. *Proc. ACM New Security Paradigms Workshop,* 40-45.
11. Stevens, G. & Wulf, V. (2002). A new dimension in access control: Studying maintenance engineering across organizational boundaries. *Proc. CSCW 2002,* 196-205.
12. Stiemerling. O., & Wulf, V. (2000). Beyond 'yes or no'—Extending access control in groupware with awareness and negotiation. *Group Decision and Negotiation, 9,* 221-235.
13. Westin, A. F. (1998). *E-commerce & Privacy: What Net Users Want.* Hackensack, NJ: Privacy & American Business.