Why is Evaluating Usability of Privacy Designs So Hard? Lessons Learned from a User Study of PRISM

Sameer Patil, Alfred Kobsa
Department of Informatics
University of California, Irvine
Irvine CA 92697 USA
{patil, kobsa}@uci.edu

ABSTRACT

Privacy is a thorny issue that affects all information systems designed for interpersonal interaction and awareness. Theoretical insights regarding privacy and user experience in a variety of systems have produced numerous design principles and guidelines for building systems sensitive to privacy issues. In order to truly improve support for privacy management, the usability of systems that implement these principles is critical. Yet, usability evaluation of privacy designs is a relatively unexplored area. In this paper, we describe our experience of conducting a longitudinal user study to evaluate the effectiveness of the privacy management enhancements offered by PRISM, a plugin for Instant Messaging (IM). Although the study was unable to achieve its intended objective, the lessons learned highlight the difficulties faced by evaluators of privacy management mechanisms. We hope that our experience will benefit future evaluations of privacy management mechanisms, and will initiate discussions on overcoming some of these challenges.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: Evaluation/methodology

General Terms

Human Factors, Design

Keywords

Privacy, User studies, Instant messaging, IM, Methodology

1. INTRODUCTION AND RELATED WORK

Recent years have seen a proliferation of technologies for interpersonal awareness and interactions. Media spaces, chat rooms, Instant Messaging (IM), blogs, and social networking have enabled new forms of communication, and allowed fostering of greater awareness even when people are physically separated. At the same time, these systems have impacted privacy [1]. Making sure that privacy concerns do not overshadow the utility of these systems is an important challenge faced by the designers of such systems [11].

Prior research has drawn upon theoretical insights regarding the concept of privacy (e.g., [18]) as well as practical experiences with usage of specific systems (e.g., [6, 8, 17]). These endeavors generated numerous principles and guidelines (e.g., [2, 10, 13, 14]) to aid designers in increasing the "privacy-sensitivity" of awareness and communication systems. However, successfully translating these suggestions into working systems also requires that privacy management mechanisms be *usable*, i.e. – lightweight, convenient, and seamless. If privacy management mechanisms are not usable, they are utilized insufficiently and/or inappropriately [9, 15], even when designed according to the guidelines.

In fact, a community of researchers has formed recently around the theme of "usability of privacy management," and the related issue of security. Toward this end, it is important to understand how to evaluate the usability of privacy management mechanisms, and how to measure or rank their effectiveness. Usability evaluation of privacy designs is still a relatively unexplored area.

Traditional usability evaluation typically involves a single user interacting with a system to perform a given set of carefully designed tasks. Usability insights are obtained from observing the user interaction, listening to the user's "think aloud" explanations, and recording metrics such as completion rate, time taken for completion, etc. However, such evaluation is unsuitable when it comes to privacy considerations in systems built for interpersonal interactions, for several reasons:

- The systems are comprised of not just the technology, but they also include the people who interact with each other. These people will therefore need to be included in the study.
- Privacy is a highly context-dependent concept, and several privacy studies encountered measurement errors when the context was not properly considered. In the case of interpersonal privacy, the context for privacy management is provided not just by the situation of the user alone but also by the activities of those that he or she interacts with using the system. Designing tasks that depend on actions of third parties that are not part of the study is impossible, at worst, and impractical, at best.
- Privacy practices often co-evolve as a community of users gathers more experience with a system and/or develops a set of social norms around its usage. A study over a short period of time will not be able to capture these developments.

Therefore, it seems that a study that observes users' privacy-related behaviors *in vivo*, over an extended period of time is the best way

to alleviate the above concerns. However, when we conducted such a study to evaluate the usability of a plugin that enhances privacy management in Instant Messaging (IM), we encountered unexpected problems. In this paper, we describe the lessons learned from our experience. We believe that these lessons will inform future usability evaluations of privacy management systems. We also wish to initiate discussions on effective strategies for overcoming some of these obstacles.

2. DESCRIPTION OF THE STUDY

We studied users of IM with the goal of understanding their privacy attitudes and practices [12, 19, 20, 21]. Based on the findings from these studies, we generated several design ideas for improving privacy management in IM. We implemented many of these ideas in the form of a plugin – called PRISM (PRIvacy-Sensitive Messaging) – for the cross-platform instant messaging client GAIM (now Pidgin: http://www.pidgin.im) [22].

We then conducted a longitudinal user study to evaluate the extent to which PRISM met its objective of improving IM privacy management. The study involved a quarter-long, upper-division undergraduate course at a large public university in the U.S. The course required the students to engage in a team project that lasted through the quarter. Each team comprised 4-5 students. The teams worked on projects defined and managed by external "customers" who came from both within as well as outside the university. The subjects of the study were the students, the customers, the instructor, and the teaching assistant.

The first two weeks of the quarter were utilized for instructions and setup. After filling out a pre-study questionnaire, all subjects were required to install GAIM. They were also asked to create a separate IM account for class purposes. The students were asked to add their project partners, the instructors and their customers to this account and vice versa. Additionally, 5 out of the 9 project teams and their customers were asked to install PRISM. The other 4 teams and their customers were assigned to the control group. The teams that installed PRISM were asked not to divulge the installation of the plugin to those in the control groups. We used a short verification questionnaire to ensure that all subjects had installed GAIM and/or PRISM successfully, and that they were aware of its functionality. However, in order to avoid biasing the participants we took care not to make references to privacy in any of the questionnaires.

Subjects were then asked to use the class IM account throughout eight weeks of the rest of the quarter for the purposes of collaborating with their project partners, customers, instructors and other classmates. During this period, a server collected logs of user interactions with PRISM. To account for the time required for learning how to use GAIM and PRISM, we discarded the first two weeks of logs. At the end of the quarter, subjects filled out a post-study questionnaire, which asked them in detail about their experiences with PRISM.

3. LESSONS LEARNED

To our surprise and disappointment, we were unable to achieve sufficient usage to be able to evaluate the privacy mechanisms provided by PRISM. Upon reflecting on the reasons for the lack of usage, we believe that the following important lessons can be learned.

3.1 Class project collaboration falls short of simulating collaboration in a knowledge-work organization.

A major motivation behind our IM studies was to explore the utility of IM as a means for collaboration and communication in knowledge-work. As a result, many of our design ideas were targeted at users engaged in collaborative knowledge work across multiple work spheres [7]. We expected that a course with a collaborative project, which required interactions with one's team members, other classmates, instructors, and customers, would be sufficient as an approximation of a collaborative knowledge-work environment. However, we discovered that the amount of shared context and simultaneous online time among students taking the same course is far lower than among knowledge workers collaborating on a project. As a result, we discovered that most collaborative activities of the students took place either during scheduled face-to-face meetings or completely asynchronously via email. Knowledge workers, in contrast, spend a large portion of their work time online in front of a computer with significant overlaps in their work hours. This fact, coupled with the shared context of the organizational affiliation, creates much greater incentives and opportunities for IM usage.

There are at least three major domains in which systems for interpersonal interaction are used: professional, social and educational. Our experience suggests that careful attention must be paid to the similarities and differences between these domains as well as the rigidity (or fluidity) of the boundaries placed by an individual when moving between them.

3.2 Undergraduates are not representative users.

Ideally, it is desirable to conduct a user study on a sample of the target population. Often times though, access to the target population is prohibitively difficult. The ease of access to undergraduate students makes them an attractive population for conducting user studies.. However, the use of undergraduate populations in a study could jeopardize its external validity. For privacy studies, this sampling bias has an even greater impact because undergraduates are known to have different privacy attitudes and behaviors than older adults [5, 16, 19]. Moreover, a person's age is known to have an effect on privacy concerns [4]. In order to mitigate the impact of these factors we utilized an upper-division course with older undergraduates, and provided the context of collaborative team projects. Unfortunately, we found that a course that meets only three hours each week is not enough to transcend the impacts of age and of the "undergraduate lifestyle" (for instance, undergraduates take several classes, work part-time jobs, and are often mobile across campus locations). A possible compromise is to utilize graduate students, faculty, and staff as subjects.

3.3 A longitudinal study does not guarantee the usage of privacy management mechanisms.

As discussed above, a longitudinal study is necessary for an effective evaluation of privacy management designs. However, we found that running a study over a long period of time may not be *sufficient* for ensuring that the privacy management mechanisms are used by the subjects. This situation arises because privacy management is a secondary function in the overall system usage. A user's desire and attention are focused on the primary function of interpersonal inter-

action; privacy management comes into play only when required. As a result, privacy management functionality forms a very small portion of the overall system usage to begin with. Infrequent use also leads to a vicious cycle where users do not utilize the privacy management functionality, even when desired, because they forgot about its existence and/or because they are less familiar with its operation, owing to the lack of sufficient use. Further, some users may never engage in additional privacy management if the default system behavior and preferences satisfy their privacy needs adequately.

These observations suggest that the length of such studies needs to be longer than that for a typical longitudinal user study. The study could also introduce external stimuli that require the user to use one or more of the privacy management mechanisms. Study confederates who deliberately engage in privacy-insensitive behavior is an example of such a stimulus.

3.4 Prototypes cannot overcome switching costs.

PRISM worked only with GAIM[1]. To ensure that students would use GAIM instead of the IM client they normally used, we required the creation of a separate ID, and mandated that only this ID be used for all matters related to the class. Although responses to the post-study questionnaire reported an occasional lapse, our subjects did comply with this policy overall. We did not prohibit the use of GAIM and PRISM for non-class IM activities. Yet, for all other (i.e., non-class) IMing purposes (which represent the vast majority of their IM activities), the subjects switched to their regular IM program. The enhancements of the plugin, which targeted the secondary function of privacy management, did not provide sufficient incentive to switch from other programs that provided a more familiar, convenient and polished user experience for the primary IM functions. Additionally, unlike the other IM programs, GAIM and PRISM were not available on the lab computers which are used frequently by undergraduate students.

This lesson regarding the costs of switching from the user's primary system is not limited to IM. For example, if privacy management functionality for a Web browser is packaged as a plugin available only for the Internet Explorer browser, users of other Web browsers, as well as other operating systems besides Microsoft Windows, will most likely choose to forgo the secondary enhancements than incur heavy switching costs for the primary activity of browsing the Web. An ideal solution to this problem is to develop the privacy plugin for all possible browsers on all possible platforms – a task that is daunting, if not infeasible. A more modest alternative is to develop privacy mechanisms in various domains of interpersonal interaction systems (e.g., IM, social networking, etc.) as cross-platform open standards similar to P3P (http://www.w3.org/P3P). Having such standards enables implementation by a wider community of institutional or individual software developers.

3.5 Meaningful evaluation requires involvement of the entire set of contacts.

As mentioned earlier, systems for interpersonal interaction involve entire sets of people who are interconnected in the form of a social network. By studying one such network, viz. the students, instructors and customers of the project course, we believed that we would overcome the limitation of single-user usability studies that do not take into account parties besides the user himself or herself. However, in the case of privacy, it turned out that investigating a small sub-network did not suffice; the entire network of IM users needed to be included. As mentioned earlier, privacy management practices typically require adoption over time and/or co-evolution, either of which is unlikely to occur within a sub-network of users, especially one much smaller than the larger network that has no access to the privacy management enhancements.

Consider, for example, a situation in which privacy management functionality in the context of mobile phones was available only to the users of a specific handset. Splitting one's communication network between those who posses the handset and those who do not would undermine the collective evolution of the shared experience surrounding privacy management mechanisms, hampering the attempts at evaluating their effectiveness. Access to entire networks of users is possible only with cooperation from the owners and administrators (organizations or individuals) of the specific systems. Achieving such cooperation requires actively pursuing collaborative ventures with industry partners.

3.6 Defining what constitutes success is complicated.

We asked subjects to rate the utility of the different pieces of privacy management functionality added by our plugin, as well as their likelihood of adopting these enhancements. We expected that higher averages would be a measure of success for a given functionality. However, subject responses indicate that users have opposing opinions regarding some of the enhancements. For example, some subjects found a given feature to be great value, while another one was indicated to be not so useful. Another set of subjects, however, expressed nearly opposite opinions regarding the same two features. In retrospect, it seems natural that opinions regarding a nuanced, personal, and context-dependent concept such as privacy could evoke opposite opinions. Secondly, it is quite likely that different users would find differential value in different privacy management features.

This situation makes it difficult to decide which metrics should be used to measure the "success" of privacy management designs. Should success be gauged for each feature separately or as an entire "privacy management user experience"? How does one isolate the impact of each piece of functionality on the overall experience of privacy management? An additional factor to consider is the mismatch between the stated opinions of the users and their captured actions, as reflected in the usage logs and observations of interaction [3]. How should these mismatches be reconciled for measuring success? Moreover, to account for learning and the evolution of practices, it is also important to measure the relevant metrics at multiple points in time during the study.

4. DISCUSSION AND CONCLUSION

Improving the usability of privacy management mechanisms in information systems is an ongoing endeavor. Meeting this challenge requires effective user studies that measure the extent of usability improvement, and that allow designers to rate and rank alternative design choices. Our experience shows that designing and conducting such studies is hard. Our lessons about usability evaluations of privacy-enhancing designs are not specific to IM, but apply to all

[1]GAIM was chosen because – unlike the other commercially developed, IM-system-specific clients – it is open source, cross-platform, plugin-based.

information systems that involve interpersonal interactions. Further, they need to be considered *collectively* and *simultaneously*; addressing only some of the lessons will not be adequate.

We believe that the lessons we learned are useful for those who wish to design and conduct such studies in the future. Further discussion and research is needed for developing creative methodological solutions to address these challenges. We have suggested a few possible avenues for addressing these challenges. Alternatively, it may be worthwhile to explore whether multiple studies could be run, each of which could deal with a subset of the discussed issues. We also hope that the discussions initiated by these lessons will spark new research on evaluation methodologies for privacy-enhancing mechanisms as well as to the development of metrics and guidelines to help designers rate and rank privacy management mechanisms.

5. ACKNOWLEDGEMENTS

We acknowledge all the subjects of this study. We also wish to acknowledge the help of Steve Abrams, Joel Ossher, and Xinru Page for commenting on the drafts of this paper. This research has been supported by NSF Grant Nos. 0205724 and 0808783.

6. REFERENCES

[1] V. Bellotti. What You Don't Know Can Hurt You: Privacy in Collaborative Computing. In *HCI '96: Proceedings of HCI on People and Computers XI*, pages 241–261, London, UK, 1996. Springer-Verlag.

[2] V. Bellotti and A. Sellen. Design for Privacy in Ubiquitous Computing Environments. In *ECSCW '93: Proceedings of the Third European Conference on Computer-Supported Cooperative Work*, pages 77–92, Norwell, MA, USA, 1993. Kluwer Academic Publishers.

[3] B. Berendt, O. Günther, and S. Spiekermann. Privacy in E-commerce: Stated Preferences vs. Actual Behavior. *Communications of the ACM*, 48(4):101–106, 2005.

[4] A. J. Campbell. Relationship Marketing in Consumer Markets: A Comparison of Managerial and Consumer Attitudes about Information Privacy. *Journal of Direct Marketing*, 11(3):44–57, 1997.

[5] C. J. Dommeyer and B. L. Gross. What Consumers Know and What They Do: An Investigation of Consumer Knowledge, Awareness, and Use of Privacy Protection Strategies. *Journal of Interactive Marketing*, 17(2):34–51, 2003. DOI 10.1002/dir.10053.

[6] P. Dourish. Culture And Control In A Media Space. In *ECSCW '93: Proceedings of the Third European Conference on Computer-Supported Cooperative Work*, pages 125–137, Norwell, MA, USA, 1993. Kluwer Academic Publishers.

[7] V. M. González and G. Mark. "Constant, Constant, Multi-tasking Craziness": Managing Multiple Working Spheres. In *CHI '04: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 113–120, New York, NY, USA, 2004. ACM.

[8] R. E. Grinter and L. Palen. Instant Messaging In Teen Life. In *CSCW '02: Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, pages 21–30, New York, NY, USA, 2002. ACM.

[9] J. D. Herbsleb, D. L. Atkins, D. G. Boyer, M. Handel, and T. A. Finholt. Introducing Instant Messaging And Chat In The Workplace. In *CHI '02: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 171–178, New York, NY, USA, 2002. ACM.

[10] J. I. Hong, J. D. Ng, S. Lederer, and J. A. Landay. Privacy Risk Models for Designing Privacy-Sensitive Ubiquitous Computing Systems. In *DIS '04: Proceedings of the 2004 Conference on Designing Interactive Systems*, pages 91–100, New York, NY, USA, 2004. ACM Press.

[11] S. E. Hudson and I. Smith. Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems. In *CSCW '96: Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work*, pages 248–257, New York, NY, USA, 1996. ACM.

[12] A. Kobsa, S. Patil, and B. Meyer. Privacy in Instant Messaging; An Impression Management Model. *(Under Review)*, 2009.

[13] M. Langheinrich. Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. In *UbiComp '01: Proceedings of the 3rd International Conference on Ubiquitous Computing*, pages 273–291, London, UK, 2001. Springer-Verlag.

[14] S. Lederer, J. Hong, A. K. Dey, and J. Landay. Personal Privacy through Understanding and Action: Five Pitfalls for Designers. *Personal Ubiquitous Computing*, 8(6):440–454, 2004.

[15] A. Lee, A. Girgensohn, and K. Schlueter. NYNEX Portholes: Initial User Reactions and Redesign Implications. In *GROUP '97: Proceedings of the International ACM SIGGROUP Conference On Supporting Group Work*, pages 385–394, New York, NY, USA, 1997. ACM.

[16] G. R. Milne and M.-E. Boza. Trust and Concern in Consumers' Perceptions of Marketing Information Management Practices. *Journal of Interactive Marketing*, 13(1):5–24, 1999. DOI 10.1002/(SICI)1520-6653(199924)13:1<5::AID-DIR2>3.0.CO;2-9.

[17] L. Palen. Social, Individual and Technological Issues for Groupware Calendar Systems. In *CHI '99: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 17–24, New York, NY, USA, 1999. ACM.

[18] L. Palen and P. Dourish. Unpacking "Privacy" for a Networked World. In *CHI '03: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 129–136, New York, NY, USA, 2003. ACM.

[19] S. Patil and A. Kobsa. Instant Messaging and Privacy. In *Proceedings of HCI 2004*, pages 85–88, 2004. http://www.ics.uci.edu/ kobsa/papers/2004-HCI-kobsa.pdf.

[20] S. Patil and A. Kobsa. Privacy in Collaboration: Managing Impression. In *The First International Conference on Online Communities and Social Computing*, 2005.

[21] S. Patil and A. Kobsa. Uncovering Privacy Attitudes and Practices in Instant Messaging. In *GROUP '05: Proceedings of the 2005 International ACM SIGGROUP Conference On Supporting Group Work*, pages 109–112, New York, NY, USA, 2005. ACM.

[22] S. Patil and A. Kobsa. Privacy Considerations in Awareness Systems: Designing with Privacy in Mind. In S. Verlag, editor, *Awareness Systems: Advances in Theory, Methodology and Design*. Panos Markopoulos and Boris de Ruyter and Wendy Mackay, 2008.