A Contextual Method for Evaluating Privacy Preferences

Caroline Sheedy[1] and Ponnurangam Kumaraguru[2]

[1] Dublin City University
Glasnevin, Dublin 9, Dublin, Ireland
csheedy@computing.dcu.ie
[2] School of Computer Science
Carnegie Mellon University,
Pittsburgh, PA 15213
ponguru@cs.cmu.edu

Abstract. Identity management is a relevant issue at a national and international level. Any approach to identity management is incomplete unless privacy is also a consideration. Existing research on evaluating an individual's privacy preferences has shown discrepancies in the stated standards required by users, and the corresponding observed behaviour. We take a contextual approach to surveying privacy, using the framework proposed by contextual integrity, with the aim of further understanding users self reported views on privacy at a national level.

1 Introduction

Privacy is an inherent concern for users in electronic transactions. Such concerns are based in the lack of anonymity afforded by electronic transactions. Some personal data is usually required for completion of these transactions, such as name, address, preferences. Identity management systems aim to manage various partial identities, whereas a privacy-enhanced identity management system should prevent linkability between the partial identities of a user [10]. Brands argues that "Schemes in which users do not have control over their own personal data offer zero privacy " [2] . Privacy should be a design consideration of identity management schemes, not an add-on to a finished product. Privacy is pertinent to a wide range of arenas: social, legal, technical etc.

Forming an understanding about privacy perceptions and concerns of individuals is generally achieved by conducting privacy surveys [6, 8]. The most prolific privacy surveyor is Dr. Alan Westin [7].

1.1 Motivation

By analysing a privacy survey using a contextual method, we hope to garner further insight into users privacy attitudes. As privacy is increasingly considered at a national and international level, it is necessary to have a consistent and effective means of comparing surveys.

Please use the following format when citing this chapter:

Sheedy, C. and Kumaraguru, P., 2008, in IFIP International Federation for Information Processing, Volume 261; *Policies and Research in Identity Management*; Eds. E. de Leeuw, Fischer-Hübner, S., Tseng, J., Borking, J.; (Boston: Springer), pp. 139–146.

1.2 Privacy Surveys

Privacy surveys are conducted to with a view to identifying people's conception of privacy, and their attitudes on how their personal information is collected and used [4]. They suffer from the 'talk is cheap' problem [3]. That is, users may state any preferences they like, without due consideration of the consequences.

Existing privacy survey methods have some associated issues [3, 11]. One such consideration is the survey design as it can skew results and manipulate responses. Another is finding a correlation between what respondents say and what they actually do [3]. The lack of comparability of independent studies is yet another issue. The factors affecting this include context, wording and sample size.

A valid question is "What should privacy surveys results be used for?". Surveys may be used as a means to evaluate public opinion, rather than dictate policy. As the need for international concurrence in this area increases, so too does the requirement for a means to effectively evaluate findings of such surveys.

2 Background

2.1 Contextual Integrity

Contextual integrity (CI) [9] was developed as an alternate benchmark for evaluating privacy breaches, in response to greater challenges from emerging technologies. Two of the fundamental concepts underlying CI are *contexts* and *norms*.

Contexts model societal structure, reflecting the core concept that society has distinctive settings. For example, society distinguishes between the social contexts of a hospital and a university. CI allows individuals to describe their privacy expectations, by associating norms of behaviour with contexts. The notion of a context and its norms mirror societal structure. In contrast to other privacy theories, CI associates the context with the subject's attribute being passed. Whether or not the data in question is confidential is often not the issue - information may only be deemed sensitive with respect to certain contexts.

CI uses norms of transmission to describe the accepted ways in which information may be passed. This reflects the accepted data flows within a context. These norms are used to describe a context, and they facilitate the sharing of data in a prescribed manner. Data gathering and dissemination must be appropriate to the stated norms of a context. For example, a student may accept their examination results being known within their academic department, as this is a norm within a university. They may not accept their results being passed outside of the university, as this would change the context in which the data flows.

Originating in a social and legal setting, CI postulates that *there are no arenas of life not governed by norms of information* [9]. It aims to provide users with an intuitive way to state their preferences. For instance, students naturally describe the limited sets of people that should access their examination results. They reference the context (university), the agents involved (employees, other students), their associated roles (finance, teaching), and the actual data (fees, results) to be passed. This facilitates the prescription of acceptable data flows as well as describing existing ones.

There are two types of norms: *norms of appropriateness* and *norms of distribution. Norms of appropriateness* address what information it is pertinent to disclose about a subject in a given context. For example, it is relevant for academic affairs to know details of a student's fee status to facilitate registration. Equally relevant, however, is to know what is not appropriate, such as details of a student's fee status being passed on to another student. This highlights the importance of relationships in norms. *Norms of distribution*, or flow, address the movement of information from one party to another. A finance office employee typically may only pass the students fee details to an employee of academic affairs. Thus, in order to ensure norms are respected, both the appropriateness and the distribution are considered.

2.2 Relevance of Context

Hine and Eve [4] observe that no single type of information is considered personal in all situations. Described as 'situated privacy', users associate a context with their privacy choices. Notions of privacy are found to be socially constructed, and CI was introduced as a means to mirror socially accepted norms of privacy. Combining these concepts, we analyse a survey on privacy preferences using a CI structure. This allows us to examine the complex nuances of privacy with a novel approach.

One suggestion [1] is to allow users to view their information in different circumstances. For example, the information they would be prepared to give to a retailer may differ from that which they would give to a marketer. CI offers a means to do this, by showing the associated contexts and norms of flow for a specific piece of information.

The need for context as a consideration of privacy surveys has been identified [6]. People will reveal data which they consider to be the 'norm', or typical for the given group or context. By discovering such norms using CI, survey results could aid in the design of an identity management system. Respondents were also found to only reveal atypical data which paints them in a positive light - people who has a slightly below average weight were happy to publicise this. Context is once again emphasised as an important factor of privacy.

2.3 Sample Survey

A study of international attitude differences with respect to privacy between the USA and India is carried out in [8]. They use the method of "mental models", with one-on-one interviews with 57 subjects. They drew interesting results on the differences in the national privacy perceptions of the subjects from the United State and those from India. We re-evaluate the responses from the Indian participants using a contextual approach. We aim to derive accepted norms of flow from the responses.

2.4 Culture

Hofstede [5] identifies two types of cultures, *collectivism* and *individualism*, and discusses the divide between them. A collectivist society uses 'we' as a major source of identity, whereas individualist societies expects individuals to identify as 'I'. Hofstede

[5] develops a number of cultural indices which measure the differences between societies. Of particular relevance is the Individualism Index (IDV), which measures how collectivist or individualist a society is. As India has a low IDV score, it is considered a collectivist society. Collectivist societies consider harmony with one's social environment a key virtue. Group interests prevail over those of the individual. In contrast to this, the USA is the highest ranking individualist society in Hofstede's study. He details the difficulties associated with culturally biased approaches to studies, and evaluating them. The survey used [8] follows Hofstede's recommendation of involving researchers from different cultures when developing questions.

CI [9] was developed by western minds, used to a culture of individualism. We examine its applicability to a non-individualistic culture by applying it to the surveys of the Indian respondents [8].

3 Analysis

3.1 Survey Details

The 14 questions posed in the survey protocol [8] were examined. It was decided to re-evaluate questions 1-10 only, as they covered the main areas of interest. They include general understanding and concerns about privacy, awareness of and concerns about privacy and technology, concerns about identity theft and knowledge of and concerns about data collection in organisations and government. The responses were considered in terms of dichotomies such as sensitive or non-sensitive, public or private and safe or unsafe, with the aim of uncovering accepted norms of information flow.

The survey was conducted via one-on-one open-ended interviews with respondents from India and the USA. Unfortunately, the original transcripts are unavailable from the participants in the USA. We had access to the 29 transcripts from the Indian subjects. The population sampled is not to be considered as statistically representative of any particular community or of technology users [8], consisting largely of college educated people who are familiar with the Internet.

3.2 Template

Question 1 "When you hear the word privacy, what comes to mind?"
Respondents considered privacy as one of two things: physical privacy and informational privacy. Over half of the respondents focused solely on privacy as a physical issue. They cited the contexts of home and work. The remainder considered privacy in terms of personal information. The respondents cited the contexts of social life, political, economic and online. Two respondents considered privacy as both information and physical.

A general trend in the nature of privacy concerns in respondents who considered privacy to be physical space throughout the survey was noted. Their concerns were focused on more traditional contexts, for example people looking at their monitor in an office or looking at paper based bank statements and credit cards in a bank. Respondents who considered privacy in terms of personal information focused on privacy concerns caused by existing and emerging technologies, such as data being disseminated electronically without their permission.

Question 2 "Do you have any (other) concerns about privacy?"
Out of the 29 respondents, 14 stated they had no privacy concerns. The remainder stated they had privacy concerns, and gave examples containing a context and sensitive data. This highlights the data that is considered private or sensitive such as financial, email, religion, background. The associated contexts identified were personal sphere, professional sphere and family.

With almost half of the respondents stating they had no privacy concerns, some cultural insight is given by one respondent: *".. the Indian culture has a system which says we should not hide anything...everything is common"*. This is reflective of a collectivist society.

Question 3 "Keeping computerised information secure, and out of the hands of people and groups that should not have it, is a problem that overlaps the issue of privacy. Tell me any concerns you may have."
Responses here were consistent. All users, except one, felt that some control mechanism was required to house the data securely. Four respondents specifically mentioned control over information dissemination. The respondent who was unconcerned felt that increasing accessibility to all data was a positive thing.

Users are happy to store their data electronically as long as access control mechanisms are place. Context was not a feature of most responses. Many respondents felt that either their personal data was not of interest to others, again a feature of the group 'we' mentality of a low IDV culture.

Question 4 "Data security and privacy are not really a problem because I have nothing to hide."
The majority of respondents, 22, disagreed with this. The remaining 7 deemed it acceptable to have all data public.

The responses to this question included free choice, confidentiality and necessity as factors in information flow. These factors correlate to the prominent norms of flow of CI.

Question 5 "Do you feel that information about you is exclusively yours, and the people who you ask to take the information?"
24 respondents agreed with this statement, with 18 stating that an individual maintains control of the information is disseminated about them, and 6 stating that this should be true, but was unrealistic in many contexts, such as online. The remaining 5 did not agree, as they felt data should be public.

The majority of respondents who agreed with this statement focused the situation where the information is passed by the individual. There was no mention of their data being passed on by others. No respondents brought up contentious issues, such as the entitlement of a sexual partner to know of their partner's HIV status.

Question 6 "Are you concerned about authorities misusing personal data of yours or members of your family?"
The 12 respondents who expressed concern here focused on potential misuse of their

data by individuals within an organisation, rather than a corporate misused of their data. The remaining 17 had no concerns.

This question highlights the difference between low and high IDV cultures. CI emphasises the right to control information about oneself. This is a facet of a a high IDV culture. The trust placed in authorities by the majority of respondents is reflective of the belief system of low IDV cultures. Should majority opinion be enforced by policy or law in this case, a free for all with regard to data mining, dissemination and collection would occur. This is an example of CI highlighting areas of concern.

Question 7 "Are you concerned about the government misusing personal data of yours, or members of your family?"
Just over half, 16, stated they are unconcerned. Reasons such as the government is restricted by laws and they can only garner statistical knowledge of the data were given. The remaining 13 stated they are concerned.

This supports the observations of question 6 above. Applying the government-private dichotomy of political and legal inquiry of CI to this question results in a stale mate, with almost equal number on either side. A core principle of privacy theory is to protect the privacy of individuals against intrusive governments [9]. This opens the question as to what to do at a international level in the case of cultures where such concerns are not the norm. CI has been designed from a high IDV point of view, and considers the desire for privacy protection from invasive governments an objective. Thus, it is valid to question if CI can be applied to low IDV cultures.

Question 8 "Do you feel that more information collected on you and others will increase domestic security? Does it make you feel safer?"
The majority of respondents, 22 felt safer with the data being collected. Of these, 8 required that the data collected should be used in the context of national security only. The remaining 7 felt it was invasive, and open to abuse should the data flow out of context.

The context of the data played a big part of the response. A norm could be derived stating that data collected for national security cannot be used for any other reason, and that it must be securely stored. This would appease the concerns of those in favour of it, as well as addressing those against it.

Question 9 "Are you concerned about identity theft?"
17 respondents had no concerns regarding identity theft, feeling that no one would want to do this to them. A cultural norm of reflecting this trust could be drawn. The 12 who claimed to be concerned said they counteracted it using protection such as control over dissemination of information and passwords.

The data associated with identity theft by the respondents was tangible - passwords, passports, credit cards. It was felt self protection was possible, with one respondent stating that they didn't believe such things to happen in India. Thus the norm stating no concerns, so long as protection measures such as passwords and firewalls are used is possible here.

Question 10 "Consider technologies that exist today, or that soon might be developed. Are there some that you think pose a threat to privacy or data security?"
17 respondents expressed concern regarding technology, citing phones, access control cards etc as examples. The rest were aware of issues, but overall felt that technology is improving privacy and advances were generally viewed in a positive light.

Existing technological concerns focused on mobile phones and cameras.

4 Discussion

The re-evaluation of the data using a contextual method further highlights the differences between the expectations of a high IDV culture and the choices of a low IDV one. CI is proposed as a framework for stating privacy concerns. However, our findings suggest that CI needs to be extended to incorporate a low IDV culture. CI expects users to be concerned about privacy invasiveness. A significant number of survey respondents do not consider privacy concerns in terms of their personal data. They either trust the government and authorities to do the correct thing, or they consider privacy in terms of personal space. So there is a need to think about the underlying model of CI to incorporate the low IDV culture expectations.

Further work is required to design a privacy survey which captures the attitudes of international respondents. We believe that context should be a factor, as well as how to pose questions which garner the views of high and low IDV societies alike.

Acknowledgements

The authors would like to thank Dr. Stephen Blott from Dublin City University and Dr. Lorrie Cranor, Dr. Raj Reddy, Dr. Granger Morgan, and Elaine Newton from Carnegie Mellon University.

References

1. M.S. Ackerman, L.F. Cranor, and J. Reagle. Privacy in e-commerce: examining user scenarios and privacy preferences. *Proceedings of the 1st ACM conference on Electronic commerce*, pages 1–8, 1999.
2. S.A. Brands. *Rethinking Public Key Infrastructures and Digital Certificates: Building in Privacy*. MIT Press, 2000.
3. J. Harper and S. Singleton. With a Grain of Salt: What Consumer Privacy Surveys Dont Tell Us. *Competitive Enterprise Institute, June*, 2001.
4. C. Hine. Privacy in the Marketplace. *The Information Society*, 14(4):253–262, 1998.
5. G.J. Hofstede. *Cultures and Organizations: software of the mind*. McGraw-Hill, 2005.
6. BA Huberman, E. Adar, and LR Fine. Valuating Privacy. *Security & Privacy Magazine, IEEE*, 3(5):22–25, 2005.
7. P. Kumaraguru and L.F. Cranor. Privacy Indexes: A Survey of Westins Studies. *Institute for Software Research International*, 2005.
8. P. Kumaraguru, L.F. Cranor, and E. Newton. Privacy Perceptions in India and the United States: An Interview Study. *The 33rd Research Conference on Communication, Information and Internet Policy (TPRC), Sep*, 30:26–30, 2005.

9. H. Nissenbaum. Privacy as Contextual Integrity. *Washington Law Review*, 79(1):119–157, 2004.
10. A. Pfitzmann and M. Hansen. Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management-A Consolidated Proposal for Terminology. *Version v0*, 27:20, 2006.
11. M. Teltzrow and A. Kobsa. Impacts of User Privacy Preferences on Personalized Systems: a Comparative Study. *Designing Personalized User Experiences for eCommerce*, 2004.