Measuring Internet Privacy Attitudes and Behavior: A multi-dimensional approach

Adam Joinson and Carina Paine
Institute of Educational Technology
The Open University
Milton Keynes, UK, MK7 6AA
+44 1908 654811
A.N.Joinson@open.ac.uk

Tom Buchanan
University of Westminster

Ulf-Dietrich Reips
University of Zurich

ABSTRACT
In this position paper we present a series of tools developed as part of our ongoing research into privacy and self-disclosure (www.prisl.net). These tools are designed for use in Internet-based studies of privacy and user behavior, and include measures of privacy attitudes, behaviors and willingness to disclose personal information to websites. Importantly, they are based on a multi-dimensional approach to understanding privacy, and we would argue provide a better basis for the study of Internet privacy concern and behavior than measures based on broad-brush definitions of privacy.

Keywords
Privacy, measures, self-disclosure, internet

INTRODUCTION
Over the past decade, the Internet has become an important and ubiquitous component of daily life in the developed world. As is often the case, the technology is somewhat of a double-edged sword. While it may enhance our lives in many ways, as our world becomes an ‘information society’ it also raises new concerns about privacy. Much of that information relates to not just things but to people. Information about us is accessed, stored, manipulated, data mined, shared, bought and sold, analyzed and potentially lost, stolen or misused by countless government, corporate, public and private agencies, often without our knowledge or consent. When we communicate, interact or even just go shopping, both online and offline, we leave data trails and digital footprints behind us, generating information about our lives and activities as we go. As recognition of this phenomenon grows, the issue of privacy has increased in salience. Research and articles about online privacy are now appearing regularly in the academic and popular press (e.g. Vise, 2005).

Measuring Privacy
Given the increased concern about privacy, the issue has not gone unexamined by researchers. Many public opinion surveys and polls about privacy have been conducted and have been one of biggest factors in privacy debate. For example, Jupiter Research (2002) reported 70% of American consumers worry about online privacy. However, some have been critical of the methodology and interpretation of such polls (Harper & Singleton, 2001), and people’s understanding of what privacy actually means is rarely examined in any detail.

For instance, the Harris Poll is a privacy survey that has been conducted regularly since 1995 by telephone across the United States among approximately 1,000 people. The survey includes the Westin privacy segmentation (Harris and Associates Inc. & Westin, 1998), which is a scheme for categorizing individuals’ different levels of privacy concerns. It divides respondents into one of three categories depending on their answers to three statements: Privacy Fundamentalists, who view privacy as an especially high value which they feel very strongly about; Privacy Pragmatists, who too have strong feelings about privacy but can also see the benefits from surrendering some privacy in situations where they believe care is taken to prevent the misuse of this information; and Privacy Unconcerned who have no real concerns about privacy or about how other people and organizations are using information about them. However, this process, based as it is on only three questions, is weak methodologically, and is not nuanced enough for a detailed examination of privacy attitudes.

Several studies have also attempted to measure privacy concerns in more detail and to identify different types of privacy concern. The Concern for Information Privacy (CFIP) scale was developed by Smith, Milburg and Burke (1996). It was the first measure of its kind and measured individuals’ concern regarding organizational practices. It identified four factors – collection, errors, secondary use and unauthorized access to information as the dimensions of an individuals concern for privacy. Later research (e.g. Stewart & Segars, 2002) argued that the CFIP needed to be re-evaluated and developed following advances in technology, research and practice.

More recently, Malhotra, Kim and Agarwal (2004) operationalized a multidimensional notion of Internet Users Information Privacy Concerns (IUIPC). Their model (and measuring instrument) recognizes that there are multiple aspects of informational privacy: they identify attitudes towards the collection of personal information, control over personal information; and awareness of privacy practices of companies gathering personal information as being components of a second order construct they label IUIPC. While this model does consider multiple aspects of privacy, all of these aspects still lie within the domain of informational privacy.

However, most definitions of privacy outside of the CHI community now adopt a multi-dimensional approach to understanding the subject. Burgoon, Parrott, LePoire,
Kelley, Walther and Perry (1989) and DeCew (1997) have both developed multidimensional definitions of privacy.

The dimension Informational Privacy appears in both Burgoon et al.’s and DeCew’s definitions. Burgoon et al state that informational privacy relates to an individual’s right to determine how, when, and to what extent information about the self will be released to another person (Westin, 1967) or to an organization. The dimension Accessibility Privacy, as defined by DeCew, overlaps with informational privacy in cases where “acquisition or attempted acquisition of information involves gaining access to an individual” (DeCew, 1997, p. 76). However, it also extends to cases where physical access is at stake (for example, ‘intrusions’ by spam mail or computer viruses). This dimension overlaps with Burgoon’s Physical dimension of privacy, which is the degree to which a person is physically accessible to others.

Finally, DeCew identified Expressive Privacy, which “protects a realm for expressing ones self-identity or personhood through speech or activity. It protects the ability to decide to continue or to modify ones behavior when the activity in question helps define oneself as a person, shielded from interference, pressure and coercion from government or from other individuals” (DeCew, 1997, p77). This dimension overlaps with Burgoon’s social / communicational dimension of privacy, which is an individual’s ability and effort to control social contacts.

The tools presented in the present position paper aim to tap into this multi-dimensional approach to understanding privacy, as well as proposing a method for the study of people’s willing disclosure to web-based services. We argue that by focusing on either opinion-poll type questions, or informational privacy alone, researchers will miss many of the important interactions between types of privacy.

Identifying Privacy Concerns

We began our study of multi-dimensional Internet privacy concerns using a Dynamic Interview Program (DIP) which interviews ICQ-users automatically (Stieger and Reips, 2005). DIP can ask two sorts of questions - with or without pre-defined answers required. A question without a set of required pre-defined answers (open format) shows only a minimal set of ‘steering answers’ in the square brackets: [DECLINE / EXIT / INFO]. Questions that may only be answered with one of the options listed (closed or multiple choice format) are recognizable via the additional possible answers in the square brackets, e.g. [ YES / NO / DECLINE / EXIT / INFO ].

DIP is able to adaptively choose questions depending on the respondent’s answers (dynamic branching), similar to branched testing in diagnostic assessments. This type of sensitive automated interviewing spares the respondent from having to answer superfluous questions (with demotivating effects) and saves time for a deeper investigation of the specific respondent’s opinion.

For information about the specific question asked, DIP offers an INFO response option. By typing “INFO” the respondent is presented with a question specific information text that explains the question in more detail.

We (Paine, Reips, Steiger, Joinson and Buchanan, 2006) used DIP to interview Internet users about their privacy concerns and precautionary behaviors. An interview request was sent to 79,707 randomly selected ICQ users. From 1,507 subjects came some kind of response. The response rate was 1.9%, with response defined as responding to the minimum of one question. Following data cleaning 530 participants’ responses were fully analysed. These results illustrated the breadth of topics that people consider to be privacy concerns – for instance, the most common concern was ‘viruses’ (16.1% of people), followed by ‘Spam’ (10.5%), and then ‘Spyware’ (9.9%). A multi-dimensional approach to understanding privacy would accommodate this by arguing that viruses and spam, while not a threat to informational privacy, do threaten accessibility privacy.

PRIVACY CONCERN AND BEHAVIOR SCALES

Following the DIP study, we (Buchanan, Paine, Joinson & Reips, 2006) developed, refined and validated a measure of Internet related privacy concerns and behavior. This measure was developed to reflect the various dimensions of privacy concern, rather than simply threats to informational privacy.

Item development

Following an examination of existing published privacy literature, definitions and surveys (including Burgoon et al, 1989 and DeCew, 1997), a set of 82 privacy items was collated, including both novel items and some drawn from these sources. In addition to informational privacy (e.g. “Are you concerned that you are asked for too much personal information when you register or make online purchases?”), questions relating to all of the theoretically distinct aspects of privacy outlined above were included. Thus, items intended to address accessibility (e.g. “Are you concerned that information about you could be found on an old computer?”), physical privacy (e.g. “Are you concerned about people viewing your screen over your shoulder when you are online?”), expressive privacy (e.g. “Are you concerned that an email you send someone may be inappropriately forwarded to others?”), and possible benefits of surrendering privacy (e.g. “How acceptable is it that personal information provided online can be used to speed up log in / purchases?”) were included. Thirty-four of the items addressed privacy related behavior (e.g. “Do you clear your Internet browser history regularly?”). Each of these questions required responses to be made on a five-point scale that was labeled at each end, from ‘never’ to ‘always’. Forty-eight of the items related to privacy attitudes. For these questions responses had to be made on a five-point scale labeled at each end, from ‘not at all’ to ‘very much’.

Scale development

The 82 items were administered to 515 members of an online volunteer research panel at the Open University. Of the respondents, 43% (220) were male and 57% (286) were female (demographic data was unavailable for nine participants). The mean age of the sample was 43.9 years (range: 22 – 77 years, SD= 10.4).
Factor analysis
Factor analysis of the 82 items yielded two privacy behavior scales, and a single privacy attitude scale. The items comprising these scales are shown below.

Table 1: Privacy Behavior factor loadings.

<table>
<thead>
<tr>
<th>Content</th>
<th>Factor 1 Loading</th>
<th>Factor 2 Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Caution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you shred / burn your personal documents when you are disposing of them?</td>
<td>.365</td>
<td>.162</td>
</tr>
<tr>
<td>Do you hide your bank card PIN number when using cash machines / making purchases?</td>
<td>.329</td>
<td>.077</td>
</tr>
<tr>
<td>Do you only register for websites that have a privacy policy?</td>
<td>.701</td>
<td>.066</td>
</tr>
<tr>
<td>Do you read a website’s privacy policy before you register your information?</td>
<td>.777</td>
<td>-.041</td>
</tr>
<tr>
<td>Do you look for a privacy certification on a website before you register your information?</td>
<td>.790</td>
<td>-.030</td>
</tr>
<tr>
<td>Do you read license agreements fully before you agree to them?</td>
<td>.676</td>
<td>-.009</td>
</tr>
<tr>
<td><strong>Technical Protection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you watch for ways to control what people send you online (such as check boxes that allow you to opt-in or opt-out of certain offers)?</td>
<td>.188</td>
<td>.407</td>
</tr>
<tr>
<td>Do you remove cookies?</td>
<td>.215</td>
<td>.600</td>
</tr>
<tr>
<td>Do you use a pop up window blocker?</td>
<td>.030</td>
<td>.745</td>
</tr>
<tr>
<td>Do you check your computer for spyware?</td>
<td>.047</td>
<td>.750</td>
</tr>
<tr>
<td>Do you clear your browser history regularly?</td>
<td>.150</td>
<td>.616</td>
</tr>
<tr>
<td>Do you block messages / emails from someone you do not want to hear from?</td>
<td>.212</td>
<td>.451</td>
</tr>
</tbody>
</table>

Are you concerned about people online not being .741 who they say they are?
Are you concerned that information about you .586 could be found on an old computer?
Are you concerned who might access your .630 medical records electronically?
Are you concerned about people you do not know .683 obtaining personal information about you from your online activities?
Are you concerned that if you use your credit card .738 to buy something on the internet your card number will obtained / intercepted by someone else?
Are you concerned that if you use your credit card .717 to buy something on the internet your card will be mischarged?
Are you concerned that an email you send may be .682 read by someone else besides the person you sent it to?
Are you concerned that an email you send .683 someone may be inappropriately forwarded to others?
Are you concerned that an email you send .629 someone may be printed out in a place where others could see it?
Are you concerned that a computer virus could .611 send out emails in your name?
Are you concerned about emails you receive not .629 being from whom they say they are?
Are you concerned that an email containing a .674 seemingly legitimate internet address may be fraudulent?

Table 2 Privacy Attitude factor loadings.

<table>
<thead>
<tr>
<th>Content</th>
<th>Factor 1 Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>In general, how concerned are you about your .688 privacy while you are using the internet?</td>
<td></td>
</tr>
<tr>
<td>Are you concerned about online organisations not .726 being who they claim they are?</td>
<td></td>
</tr>
<tr>
<td>Are you concerned that you are asked for too .577 much personal information when you register or make online purchases?</td>
<td></td>
</tr>
<tr>
<td>Are you concerned about online identity theft? .753</td>
<td></td>
</tr>
</tbody>
</table>

SELF-DISCLOSURE MEASURE
Self-disclosure is increasingly required for us to engage effectively with online systems – for instance, most personalization, registration sites and social software require some disclosure of personal information to work. It is closely related to online privacy since we would expect that many privacy concerns are related to the request and use of personal information. While such disclosure can be studied through the use of mock-up websites, this is often problematic since participants either must suspend disbelief and engage in a ‘what if’ scenario exercise, or the outcome measure lacks power since it only tells us whether or not a person chose to submit the information requested. As part of the PRISD project, we (Joinson, Paine, Buchanan & Reips, 2006; Reips, Joinson, Buchanan and Paine, 2006) have developed a tool to enable to measurement of self-disclosure using the provision of an ‘I prefer not to say’ option to sensitive questions. Choice of this option is taken as a measure of non-disclosure that does not require the participant to exit the process. It also enables the researcher to distinguish between a wish to protect privacy (called ‘active’ non-disclosure) and people who simply do
not submit any response or run through on default (called ‘passive’ non-disclosure).

Method
Participants were the same 515 members of the volunteer research panel of that completed the privacy scale development.

Self-disclosure measure
Participants completed an 18-item measure of self-disclosure titled ‘About you and your life experiences’. Six of the items were non-sensitive (e.g. “What season were you born in”) and twelve sensitive (e.g. “How many sexual partners have you had?”). In total, 24 sensitive items were developed. To reduce participant effort, the sensitive items were grouped for face validity and split into two 12-item versions. The six non-sensitive items were presented in both versions.

Participants responded to the items using a both drop down menus and open text boxes (for instance, when height and weight was asked). For drop down responses, the default option was “Please select” with “I prefer not to say” the second choice. For text boxes, a radio button labelled “I prefer not to say” was provided next to the text box.

Results
The mean number of times the ‘I prefer not to say’ option was used was 1.08 (SD = 2.59, range 0-18). The question, “How many different sexual partners have you had” was non-disclosed 27.3% of the time (the highest amount), followed by a question on the number of serious relationships since age 18 (14%), whether these partners were the same or opposite sex, or a mixture (13.26%), weight (12.8%) and length of longest relationship (10.2% non-disclosure). Items about height, season of birth, handedness and academic plagiarism showed the lowest use of the ‘I prefer not to say option’.

SUMMARY
We argue that tools to measure privacy and its related outcome measures need to reflect the breadth of privacy concerns that people report, and the multi-dimensional nature of privacy. We suggest that a combination of measures – both open-ended and psychometrically developed scales – should be used to fully appreciate the different dimensions of privacy on the Internet. Full copies of the measures outlined here, and reports of their development, can be accessed via the URL: http://www.prisd.net

ACKNOWLEDGMENTS
Funding from the UK Economic and Social Research Council E-Society Programme (RES-341-25-0011) supported the work reported in this paper.

REFERENCES