Running Head: INTERNET PRIVACY SCALES

Development of measures of online privacy concern and protection for use on the Internet.

Tom Buchanan

University of Westminster

Carina Paine and Adam N. Joinson

Institute of Educational Technology, The Open University

Ulf-Dietrich Reips

University of Zurich

Address: Department of Psychology, University of Westminster, London, W1B 2UW

Telephone: +44 207 911 5000 ext 2017   Email: buchant@wmin.ac.uk

Correspondence concerning this article should be sent to Tom Buchanan, Department of

Psychology, University of Westminster, 309 Regent Street, London W1B 2UW, United

Kingdom. Email buchant@wmin.ac.uk

Abstract

As the Internet grows in importance, concerns about online privacy have arisen. We describe the development and validation of three short Internet-administered scales measuring privacy related attitudes (*'Privacy Concern'*) and behaviors (*'General Caution'* and *'Technical Protection'*). In Study 1, 515 people completed an 82-item questionnaire from which the three scales were derived. In Study 2, scale validity was examined by comparing scores of individuals drawn from groups considered likely to differ in privacy-protective behaviors. In Study 3, correlations between the scores on the current scales and two established measures of privacy concern were examined. We conclude that these scales are reliable and valid instruments suitable for administration via the Internet, and present them for use in online privacy research.

Over the past decade, the Internet has become an important and ubiquitous feature 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. For 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).

This article is not about privacy *per se* (for a recent review of psychological issues relating to privacy and the Internet, see for example Joinson & Paine, 2006). However, it is motivated by the recognition that there are important privacy issues related to online activities as mundane as buying your weekly groceries over the Web (e.g. does the retailer store information on your purchases? Is it sold to third parties so they can send you targeted junk mail?), or as specialized as online psychological research (e.g. is identifying information gathered about participants? Can confidentiality be guaranteed?) or teaching (e.g. if Virtual Learning Environments allow student behavior to be tracked, what are the ethical implications? Would awareness of this affect students' willingness to

use the technology?). Awareness of these issues may affect people's behavior in a wide range of contexts. It is therefore important to have methods of identifying and quantifying people's privacy concerns, as a tool for research on how people behave both on and off the Internet.

Privacy

There have been several attempts to define privacy. In a legal context, privacy has been considered to be largely synonymous with a 'right to be let alone' (Warren & Brandeis, 1890). However, others have since argued that privacy is only the right to prevent the disclosure of personal information to others (e.g. Westin, 1967). Despite the many attempts to create a synthesis of existing literature, a unified and simple account of privacy has yet to emerge. The highly complex nature of privacy has resulted in an alternative way of defining it - through its various dimensions. 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; access to information about home addresses that people might wish to keep private and so on). 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). In this way, expressive privacy restricts external social control over choices about lifestyle, and improves internal control over self-expression and the ability to build interpersonal relationships. This dimension overlaps with Burgoon's *Social / Communicational* dimension of privacy, which is an individual's ability and effort to control social contacts (Altman, 1975).

Central to these dimensions is the desire to keep personal information out of the hands of others, or in other words *privacy concern* (Westin, 1967), and the ability to connect with others without interference. In a systematic discussion of the different notions of privacy, Introna and Pouloudi (1999) developed a framework of principles that explored the interrelations of interests and values for various stakeholders where privacy concerns have risen. In this context, concern for privacy is a subjective measure—one that varies from individual to individual based on that person's own perceptions and values. In other words, different people have different levels of concern about their own privacy.

Privacy Measurement

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. While some have been critical of the methodology and interpretation of such polls (Harper & Singleton, 2001), they do appear to provide evidence that people recognize the existence of threats to their privacy while online.

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.

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 individual's 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. Other dimensions such as expressive privacy are not addressed.

As described above, studies tend to focus on informational privacy and privacy scales are usually approached with a view of privacy as a one-dimensional construct. Harper and Singleton (2001) suggest that one of the main defects of most privacy surveys and studies is that they do not separate out all of the different factors that could be considered privacy

issues. It is clear from the definitions of privacy provided above that it is a multifaceted concept, and therefore that scales attempting to measure concern should tap these different facets about which people may be concerned. For instance, Paine, Reips, Stieger, Joinson & Buchanan (2006) used an automated interview agent to collect Internet users' privacy concerns, and report a wide variety of non-information types – including viruses and spam.

Another issue not addressed by privacy scales published thus far is that there may sometimes be *benefits* to the decrease in privacy online: collection and storage of information can permit personalized services, convenience and efficiency. In some situations, expressive privacy may be obtained through the loss of informational privacy to a third party. For example, one may disclose personal details and credit card information in order to have the convenience of completing an online transaction. Therefore, it is appropriate to consider people's views of such benefits when measuring privacy attitudes.

As well as attitudes and concerns about privacy, it is important to consider behaviors people may adopt to safeguard their privacy. For example, have you ever provided false or incomplete personal information when registering on some website, rather than giving your real name and address? We suspect most people would answer yes. There is likely to be a complex relationship between attitude and behavior in this context. For example, one's computer being infected by a virus can be seen as an invasion of privacy. We may be concerned about the possibility, and accordingly take steps to prevent it (use anti-virus

software, or an operating system less vulnerable to viruses). Concern prompts us to take preventative measures, but knowing that measures have been taken could reduce our level of concern (Paine et al, 2006, found that some people reported that they were not concerned about privacy, and when asked why stated that they had taken action to protect their privacy). It is likely that only asking people about their concerns will produce an incomplete picture: we also need to ask about privacy related behaviors.

The Present Study

As indicated above, existing privacy scales could benefit from expansion in a number of ways. In particular, the range of constructs tapped by the measures could be increased, and behaviors as well as attitudes addressed. Additionally, as far as we are aware none of the measures describe above have been validated for use on the Internet. The web is a convenient and increasingly accepted medium for psychological research, and seems ideally suited for investigation of peoples' concerns about online privacy. However, it remains important to ensure that psychological measures used online really are valid tests of the constructs they purport to address (e.g. Buchanan, Ali, Heffernan, Ling, Parrott, Rodgers & Scholey, 2005; Buchanan, Johnson & Goldberg, 2005; Buchanan & Smith, 1999; Reips, 2000; Reips, 2002).

The aim of the present study is to develop a robust, reliable measure of privacy concerns and behavior suitable for administration via the Internet. As outlined above, it is important not to look only at threats to informational privacy, but also to address other aspects of privacy and the privacy-related behaviors people may adopt.

Study 1

Method

 Materials.

Following an examination of existing published privacy literature, definitions and

surveys (including Burgoon et al, 1989; DeCew, 1997; Fox, Rainie, Horrigan, Lenhart,

Spooner & Carter, 2000; Georgia Tech Research Corporation, 1999; Stark, 2004), 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?"; "How acceptable is it that law enforcement agencies

track users of websites to track criminals?") 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* (e.g. "Are you concerned who might access your medical records

electronically?"). For these questions responses had to be made on a five -point scale labeled at each end, from 'not at all' to 'very much'. Both the attitudinal and behavioral sets of items were designed to address all the various aspects of privacy outlined above.

The final web-based questionnaire consisted of seven pages, five of which included the privacy attitude and privacy behavior items. The remaining pages incorporated a number of items relating to participants' previous Internet experience and a measure of their willingness to disclose personal information. These additional data are not relevant to the current analyses, which focus purely on the privacy related items, and are reported elsewhere (Joinson, Paine, Buchanan & Reips, 2006).

Participants.

Participants were members of a volunteer research panel of Open University (OU) students (the OU is an adult distance learning institution based in the UK, with nearly all students studying part time from home or work). Panel members study a range of subjects at the OU, and had volunteered to be contacted about up to six research-based surveys each year. In total 685 members of the research panel were invited by e-mail to complete the web-based questionnaire and 515 did so (response rate: 75%). Of the 515 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).

Procedure.

All members of the research panel were sent an invitation via e-mail to complete a web-based questionnaire. Members were told that the questionnaire consisted of a series of questions about their use of the Internet and in particular, any privacy concerns they may have, and any actions they take to address these concerns. They were informed that responses would be used to develop a measure of online privacy attitudes and behaviors. They were also informed that in order to develop a highly accurate measure, a large number of questions were included, some of which were very similar. They were asked to answer *all* of the questions as then the most suitable items could be selected for the measure. Participants were informed that all information they provided would remain confidential.

Participants were prompted to use the full scale when responding and not only the labeled response options. At the end of each page, participants' responses were submitted. The website was left open for two and a half weeks. Participants took, on average, 17 minutes to complete the questionnaires.

In this study, and the two that follow, questionnaires were administered using an online surveying system called ELSA developed at the OU. This is a powerful web-based surveying system that enables personalized URLs, sophisticated authentication, and automated reporting functions. Panel-based participants are sent URLs directing them to a web-based survey with MD5 encrypted personal identifiers (PINs) embedded within the URL. This PIN is then unencrypted and linked to the volunteering students' demographic

information on submission. A single PIN is only allowed a single response. Session based

cookies and IP number tracking are used in 'open' surveys (i.e. those that are not using

encrypted PINs) to enable the identification of any multiple submissions.

Results

As a preliminary check score distributions on each item were examined to ensure that

none suffered from restricted range (i.e. the full range of response options was being

used). This was the case for all items bar one that had a 2-5 rather than 1-5 range. The

item was retained for the time being. There were some missing data: on average, each

item was unanswered by around 12 participants ($M = 11.72$, $SD=7.02$, Range= 1-56[1]).

Accordingly, sample sizes vary in the analyses depending on how many participants

answered the relevant items. In the analyses that follow, attitudinal and behavioral items

were treated separately.

    Behavioral items.

A principal components analysis was performed, identifying 9 factors with eigenvalues

greater than 1. The Scree plot suggested a three-factor solution was tenable, with factors

after the third accounting for smaller and smaller proportions of variance.

The three-factor solution was rotated to simple structure using Varimax. Items were

identified as markers of each factor on the basis of the commonly used benchmark of a

---

[1] The 56 missing values were on the item "Are you concerned that an email you send may be read by someone else besides the person you sent it to?". There is no obvious reason why there should be a higher level of non-response to this item, and it was retained in the analyses. Apart from this item, the highest number of missing values was 21.

loading greater than .3. To maximize 'factor purity' (i.e. create a set of factor-univocal scales) a further procedure, suggested by Saucier (1994), was adopted. Saucier's criterion for a factor-pure item was that the item's loading on the marked factor should be at least twice the value of the next highest loading. Application of these two criteria led to identification of 6 marker items each for Factors 1 and 2, and 4 for Factor 3. For the Factor 1 items, Cronbach's alpha was .75 (N=495). For Factor 2, alpha was .74 (N=484). For Factor 3, alpha was only .44 (N=498). This indicates Factors 1 and 2 may form acceptable scales if they are interpretable, while Factor 3 is probably too unreliable to pursue further (unsurprisingly, given it only comprises 4 items). Loadings and item content for Factors 1 and 2 are shown in Table 1.

Table 1

Examination of item content suggests that Factor 1 reflects general caution and concern with protection of privacy in a number of ways and might be labeled something like '*general caution*'. Factor 2 seems to reflect the use of technology to protect privacy and prevent intrusion, and is likely contingent on awareness of these options. It might be labeled something like '*technical protection of privacy*'. Ability to answer the Factor 2 items positively would seem to be related to level of technical competence - for example, one would need to know what spyware and cookies were, and know how to operate the relevant software to control them. This is not the case for Factor 1, for which technical know-how is much less important.

Despite the fact that these factors arise from an orthogonal rotation and are clearly separable in terms of item loadings, they are correlated (r=.246, n= 481, p<.0005). This is likely to be because scores on Factor 1 influence the behavior tapped by Factor 2 in people who have the technical know-how to do these things: those who are protective of their privacy in general, are likely to use technology to this end if they are capable of doing so.

Neither factor seems to directly embody any single one of the dimensions of privacy previously outlined. For example, General Caution's Item 1 would seem to reflect informational privacy while Item 2 reflects physical privacy. Technical Protection's Item 4 reflects informational privacy, while Item 6 relates to accessibility. It seems rather that the clusters of behaviors identified may be motivated by concerns about multiple aspects of privacy.

Attitude items.

A principal components analysis indicated that there were 11 components in the dataset with eigenvalues greater than 1. However, examination of the Scree plot indicated a solution with considerably fewer factors would be appropriate. The gradient of the Scree slope suggested that a solution with between two and five factors would be tenable. However, the proportion of variance explained by each of these factors was small (4.8% to 6.8%) in comparison to the first factor, which accounted for 27% of the variance.

A series of exploratory factor analyses were performed, with extraction and Varimax rotation of solutions with between two and eleven factors, and application of the same criteria for item selection as the behavioral items. These analyses yielded several short, internally consistent sets of items. However, interpretation of these sets of items was problematic: for instance, one group seemed to centre on concern about viruses; another reflected concern that people might misrepresent themselves online (e.g. pretending to be someone else). It is likely that these reflect real dimensions of peoples' attitudes toward the Internet. However, none of the solutions seemed to embody the constructs we wished to measure (for example, there was no group of items related to expressive privacy). Accordingly, we decided to examine a one-factor solution, based on the first unrotated component (which explained by far the largest proportion of variance in the dataset), with the goal of identifying a general index of concern about privacy in online interactions.

We acknowledge that interpretation of unrotated solutions is a somewhat unconventional approach. For example, the first unrotated factor often reflects a response bias (respondents answering positively or negatively to all items) and may thus be an artifact rather than reflecting a real construct. However, in this instance 14 out of the 45 items had trivial loadings (less than .3) on the first factor, suggesting it did not just reflect a response bias. Also, there are precedents in the literature for instances where unrotated factors can be meaningfully interpreted – consider for example the Self-Monitoring Scale (Revised) of Gangestad and Snyder, who argue that while their instrument can be decomposed into three interpretable rotated factors, most of the meaning of their

construct (and variance in the correlation matrix) is captured by the first unrotated factor (Gangestad & Snyder, 1985).

Therefore, marker items for the first extracted component were identified on the same basis as in the previous analyses (loading on first factor of .3 or greater, and at least double the loading on any other factor). The unrotated eleven-factor solution was used for this purpose. Application of these criteria resulted in a set of 16 factor-univocal marker items, shown in Table 2.

Table 2

This set of items does seem to encompass concerns about a variety of aspects of privacy on the Internet. Alpha for this group of items is .93 (N=443). Other latent variables can probably be found among the discarded items, but this factor does appear to capture the essence of general concern about privacy online.

Discussion

On the basis of this study, a set of three short, internally consistent and interpretable scales has been developed. Two address different aspects of things people do to protect their privacy: exercising General Caution, and Technical Protection. The third scale, Privacy Concern, is attitudinal rather than behavioral, and reflects general concerns about privacy on the Internet. It is likely that there are a number of more specific latent variables addressed by the attitudinal items that were discarded, accounting for small

amounts of variance in the dataset. It might well be profitable to explore these, but that is beyond the scope of the current project where the emphasis is on privacy concern.

The next step is to examine the validity of the scales: do they really measure the constructs we suggest? Some initial evidence may be drawn from the intercorrelations of the attitude and behavior scales. The Privacy Concern scale correlates significantly with General Caution ($r=.333, n=435, p <.0005$) but not strongly with the Technical Protection factor ($r=.094, n=425, p = .053$). The former finding is consistent with the notion that both scales have a degree of validity: one would expect people with higher levels of privacy concern to be more cautious about protecting it. The latter finding also makes sense: it could be that people who score high on technical protection are actually less concerned about their privacy being violated because they are taking steps to prevent it, so the relationship between the two variables is unlikely to be simple, e.g. linear.

However, further evidence of validity is required, especially given the way the Privacy Concern scale was developed. This is addressed in Study 2, which examines whether the scales are capable of discriminating between groups who should differ in their level of privacy concern. The university at which the project was based hosts a number of in-house bulletin boards for students on various courses. We hypothesized that students on technology-based courses (e.g. computing) might be more aware of privacy threats on the Internet, and thus more likely to take measures to protect their privacy than would students on less technically-oriented courses (e.g. social sciences, humanities). Accordingly, higher scores would be expected on the privacy behavior measures.

Study 2

Method

    Materials.

The refined set of 16 privacy attitude items and 12 privacy behavior items (including both General Caution and Technical Protection items) from Study 1 were used to create a web-based questionnaire for use in this validation study. The final web-based questionnaire consisted of these items plus demographic questions, and was administered through the ELSA system.

    Participants.

Participants were recruited through two sets of online bulletin boards at the OU. The bulletin boards for technology-oriented students were one associated with the course "Vandalism in Cyberspace: Understanding and Combating Malicious Software", and also "The Technology Café" – a more general bulletin board for technology based students. The bulletin board used to recruit less technically-oriented students was associated with the course "Child Development". In total, 69 students responded. Thirty-eight were from the technology-based course, of whom 18 (47.4%) were male, and all were aged between 26 and 62 with a mean age of 44.2 and *SD* of 10.6 years. Thirty-one were from the social science based course, of whom 4 (12.9%) were male, and all were aged between 20 and 67 with a mean age of 35.7 and *SD* of 10.2 years.

Procedure.

A message was posted on each online bulletin board asking for participants for an Internet privacy questionnaire. The message included details about the importance of developing new scales of privacy concern and behavior. On accessing the survey, participants were informed that all information provided would remain confidential. The survey site was left open for three weeks. Participants took, on average, five minutes to complete the questionnaire.

Results

The technical and non-technical students did not differ significantly in their level of online Privacy Concern ($t_{(62)} = .29$, $p = .83$). Mean scores were 57.11 ($SD = 11.43$) and 56.54 ($SD = 8.96$) respectively.

Technical students did have significantly higher scores on the General Caution behavioral scale ($t_{(65)} = 1.91$, $p = .03$, one tailed), with the mean score for technical students being 22.08 ($SD = 4.56$) and that for non-technical students 19.74 ($SD = 5.50$).

The technical and non-technical students also differed significantly ($t_{(51.58)} = 2.55$, $p = .005$, one tailed) on Technical Protection of privacy. Mean scores were 26.06 ($SD = 3.17$) and 23.52 ($SD = 4.70$) respectively. There was significantly higher variance in the 'non-technical' condition (Levene's $F = 5.46$, $p = .02$). This could be attributable to the fact that there is likely to be more variance in technical competence among these social science

students (e.g. all the technical students are likely to know about things like removing cookies, not all the social science students may do so).

Correlations between the scales were also examined, across both conditions. Online Privacy Concern correlated positively with General Caution (r=.26, n=63, p=.04). It did not correlate significantly with the second behavioral factor, Technical Protection of privacy (r=-.13, n=62, p=.33).

Discussion

The two groups differed in the predicted manner in their scores on the two behavioral scales: students on technically-oriented courses reported more General Caution and a higher use of Technical Protection, despite the fact that they did not differ in their levels of Privacy Concern. This is consistent with the notion that they might be more aware of threats to privacy online, and do more to counteract them.

The pattern of correlations between the attitudinal and behavioral scales is the same as in Study 1: people with higher levels of Privacy Concern reported higher levels of General Caution but not Technical Protection. This is logical: we would only expect a positive correlation with Technical Protection for those individuals who had the relevant technical awareness and skills – and in fact they might do these things as a matter of course, irrespective of their privacy concerns.

Study 2 provides some evidence for the construct validity of the behavioral scales.

Further evidence of validity would accrue if links between the scales under development

and other measures of privacy can be demonstrated. This is addressed in Study 3.

Study 3

Method

Materials.

The Privacy Concern (attitude) and the General Caution and Technical Protection

(behavioral) scales developed in Study 1 and used in Study 2 were again used here as part

of a larger web-based questionnaire. The main questionnaire was concerned with

people's attitudes towards identity cards within the United Kingdom. Data pertaining to

that aspect of the study are presented elsewhere (Joinson, Paine, Buchanan & Reips, in

press). For current purposes, only those elements related to the scales described in this

paper are described.

In addition to our own privacy scales, two other privacy measures were also included.

The first of these measures was the Westin Privacy segmentation (Harris et al, 1998).

This measure requires participants to respond to three statements on a four-point scale.

On the basis of their scores, they can be divided into one of three categories of privacy

concern: *privacy fundamentalists*; *privacy pragmatists; or, privacy unconcerned.* For the

purposes of this study, participants were not assigned to categories: instead a total privacy

concern score was derived by summing scores across the three items. The second of these

existing measures was the Internet Users Information Privacy Concerns scale (IUIPC;

Malhotra et al, 2004), which requires responses to 10 items on seven-point scales and gives an index of respondents concerns about several aspects of informational privacy. We are not aware of either measure having previously been administered online.

Participants.

Participants were 1122 members of a research panel of OU students called 'PRESTO'. PRESTO is a different panel of students from the panel used for Study 1, so none of the current participants would have completed Study 1. In total 1935 members of the research panel were invited by e-mail to complete the web-based questionnaire. Panel members study a range of subjects at the OU. Of the 1122 respondents (response rate: 58%), 40% (449) were male and 60% (672) were female (demographic data were not available for one participant). The mean age of the sample was 42.3 years, (range: 17 – 84 years, $SD = 11.1$).

Procedure.

All members of the research panel were sent an invitation via e-mail to complete a web-based questionnaire. Members were told that the questionnaire consisted of a series of questions about any privacy concerns they may have when they use the Internet, and their privacy related behavior. Participants were informed that all information provided would remain confidential. The data collection website was left open for two weeks. Participants took, on average, 13 minutes to complete the measures.

Results

Correlations of Privacy Concern, General Caution, and Technical Protection with the Westin and IUIPC scale scores are shown in Table 3. With one exception (Westin privacy scale and General Caution, where the correlation was not significant), all were positive and significant. Privacy Concern was correlated as before with General Caution (r=.311 n=752 p<.0005) and also Technical Protection (r=.145 n=753 p<.0005).

Table 3

Discussion

While the Westin and IUIPC scales have not previously been validated for use on the Internet, the fact that they correlate as expected with the attitudinal measure (Privacy Concern) developed in this study provides evidence for the construct validity of both sets of scales.

General Discussion

Prior to Study 1, we had anticipated that privacy concern would be multifactorial in nature, and had included items related to the various dimensions of privacy that DeCew and others have outlined. However, our analysis only identified one interpretable attitudinal factor, which appears to map well onto the general concept of *Privacy Concern* (Westin, 1967). This still represents an advance over the brief (e.g. single item)

measures of privacy concern adopted by much previous research, as it taps multiple aspects of privacy (e.g. Item 6 reflects accessibility; Item 7 reflects informational privacy; Item 12 reflects expressive privacy; Item 13 reflects expressive privacy). It also has the advantage that, unlike some of the other existing scales, its validity for use in an online research environment has now been demonstrated. We therefore consider that the scale is useful, even though it provides a composite measure rather than separate indices of all theoretically delineated aspects of privacy.

Why did separate attitudinal dimensions not emerge from the factor analysis? This may be for a number of reasons. One possibility is that we failed to include sufficient numbers of questions related to each different aspect of privacy in our original question pool for item clusters to emerge from the factor analysis. Another is that these theoretical constructs are so closely related and interdependent (e.g. loss of informational privacy may lead to loss of expressive privacy) that they cannot be meaningfully separated in a questionnaire. A third is that the initial pool of items chosen did not adequately capture the relevant dimensions. Given the substantive conceptual overlap between many of the dimensions (e.g. DeCew's accessibility dimension and Burgoon et al's physical dimension), it may be somewhat naïve to expect to be able to develop 'pure' measures of each. Related to this is the possibility that the theoretical constructs put forward do not accurately reflect the way people actually think about privacy in their everyday lives, or at least the way in which they answer privacy questionnaires. Accordingly, further research aimed at more fine-grained understanding of Internet users' real-life privacy concerns is desirable (Paine, Reips, Stieger, Joinson & Buchanan, 2006).

Another interesting observation related to the attitudinal items is that all reflect concern about privacy. There are instances where surrendering privacy may have benefits for an individual (e.g. when one visits an e-commerce site, and receives recommendations for books or music one might like based on past purchases). Items pertaining to such benefits were included in our original pool. However, none of these items loaded on the general concern measure (unsurprisingly) and furthermore did not cluster together to form a 'positive' factor at any stage of the analysis.

One possible reason for the non-emergence of positive attitudes might be that the framing of the questionnaire inadvertently introduced a biased response set. The questionnaire instructions referred to *concerns* people might have, and most of the items also refer to concerns about or the acceptability of various things. It might be more appropriate for possible benefits of online information sharing to be assessed using a separate questionnaire, with instructions and phrasing less likely to induce a 'distrustful' response set.

In the case of privacy-related behaviors, we were able to identify two separate factors underpinning the actions people may take to protect their privacy online. One appeared to represent *General Caution*, taking common sense steps to protect personal information (e.g. shredding documents). The second appears to relate to the sophisticated use of hardware and software as tools for *Technical Protection* of privacy (e.g. checking for spyware, deleting cookies). There are indications that while everyone can engage in the

behavior reflected by General Caution, a higher level of technical training or awareness is required for Technical Protection. Different people may thus be able to protect their privacy in different ways.

A final issue worth considering is the nature of the sample used to develop and validate the scale. Are the findings generalizable to populations beyond panels of OU students? The panels in question are selected using stratified sampling based on age, gender, faculty of study, and region of residence, and weighted based on knowledge of volunteering patterns to ensure representativeness of the OU student population. However, that population could be characterized as better educated, slightly more likely to be female, and of higher socio-economic status than the United Kingdom population average. There is no immediately obvious reason why the structure of attitudes or behavioural tendencies towards privacy should differ between populations (though the strength with which those attitudes are held is likely to, in the same way as the groups in Study 2 differed in behavioral tendencies). However, we are currently gathering data from other populations that will be informative about whether the current findings are more widely applicable, and whether the scales 'work' with different groups of participants.

We interpret the findings of Studies 2 and 3 as indicating that the scales have a degree of construct validity as online measures of privacy related attitudes and behaviors. We will continue to use them in our own work, and present them here for the convenience of other researchers. The items and instructions to participants are shown in Tables 1 and 2.

As previously indicated, though, more fine-grained examination of privacy concern is desirable both to permit the development of appropriate questionnaires and to further elucidate models of how people perceive and think about their privacy in Internet contexts.

References

Altman, I. (1975). *The environment and social behavior.* Monterey, CA: Brooks/Cole.

Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World-Wide Web. *British Journal of Psychology, 90,* 125-144.

Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J., & Scholey, A. B. (2005). Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods, 37,* 148-154.

Buchanan, T., Johnson, J. A., & Goldberg, L. (2005). Implementing a Five-Factor personality inventory for use on the Internet. *European Journal of Psychological Assessment, 21,* 115-127

Burgoon, J. K., Parrott, R., LePoire, B. A., Kelley, D. L., Walther, J. B. & Perry, D. (1989). Maintaining and restoring privacy through communication in different types of relationship. *Journal of Social and Personal Relationships, 6,* 131-158

DeCew, J., (1997). *In Pursuit of Privacy: Law, Ethics, and the Rise of Technology.* Ithaca: Cornell University Press

Fox, S., Rainie, L., Horrigan, J., Amanda Lenhart, A., Spooner, T., Carter, C. (2000). *Trust and privacy online: Why Americans want to rewrite the rules.* Retrieved December 8, 2005, from http://www.pewinternet.org/pdfs/PIP_Trust_Privacy_Report.pdf

Gangestad, S. W., & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review, 92,* 317-340.

Georgia Tech Research Corporation (1999). GVU's 10th WWW User Survey. Retrieved December 8, 2005, from http://www.gvu.gatech.edu/user_surveys/survey-1998-10/tenthreport.html

Harper, J., & Singleton, S. (2001). *With a grain of salt: What consumer privacy surveys don't tell us.* Retrieved November 29, 2005, from http://www. cei. org/PDFs/with_a_grain_of_salt. pdf

Harris and Associates Inc., & Westin, A. (1998). E-commerce and privacy: What net users want. *Privacy and American Business and Pricewaterhouse Coopers LLP.* Retrieved June 20, 2005, from http://www. pandan. org/ecommercesurvey. html

Introna, L. D., & Pouloudi, A. (1999). Privacy in the Information Age: Stakeholders, interests and values. *Journal of Business Ethics, 22,* 27-38.

Joinson, A. N., & Paine, C. (2006). *Self-disclosure, Privacy and the Internet.* Manuscript submitted for publication.

Joinson, A. N., Paine, C., Buchanan, T., & Reips, U. -D. (2006). *Development of a behavioral measure of self-disclosure for use on the Internet.* Manuscript in Preparation.

Joinson, A. N., Paine, C., Buchanan, T., & Reips, U. -D. (in press). Watching me, watching you: Privacy attitudes and reactions to identity card implementation scenarios in the United Kingdom. *Journal of Information Science.*

Jupiter Research (2002). *Security and privacy data.* Presentation to the Federal Trade Commission Consumer Information Security Workshop. Retrieved June 20, 2005.

from http://www. ftc. gov/bcp/workshops/security/0205201leathern. pdf

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale and a causal model. *Information Systems Research, 15*, 336-355.

Paine, C., Reips, U. -D., Stieger, S, Joinson, A. N., Buchanan, T. (2006). *Internet users' perceptions of 'privacy concerns' and 'privacy actions'*. Manuscript in preparation.

Reips, U.-D. (2000). The Web Experiment Method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-118). San Diego, CA: Academic Press.

Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology, 49*, 243-256.

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment, 63,* 506–516.

Smith, J. H., Milberg, S. J. & Burke, S. J. (1996). Information privacy: Measuring individuals concerns about organizational practices. *MIS Quarterly, June*, 167-196.

Stark, D. (2004). *TNS-TRUSTe Consumer Privacy Index Q4 2004: Consumer behaviors and attitudes about privacy.* Retrieved December 8, 2005, from http://www.truste.org/pdf/Q4_2004_Consumer_Privacy_Study.pdf

Stewart, K. A. & Segars, A. H. (2002). An empirical examination of the concern for information privacy instrument. *Information Systems Research, 13,* 36-49.

Vise, D. (2005, November 20). Fears over Big Brother stalked Gmail. *The Sunday Times*. Retrieved December 8, 2005, from http://business.timesonline.co.uk/article/0,,9075-1879562,00.html

Warren, S., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review. 4,* 193-

220.

Westin, A., (1967). *Privacy and Freedom*. New York: Atheneum.

Table 1
Privacy Behavior factor loadings.

Item	Content	Factor 1 Loading	Factor 2 Loading
General Caution			
1	Do you shred / burn your personal documents when you are disposing of them?	.365	.162
2	Do you hide your bank card PIN number when using cash machines / making purchases?	.329	.077
3	Do you only register for websites that have a privacy policy?	.701	.066
4	Do you read a website's privacy policy before you register your information?	.777	-.041
5	Do you look for a privacy certification on a website before you register your information?	.790	-.030
6	Do you read license agreements fully before you agree to them?	.676	-.009
Technical Protection			
1	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)?	.188	.407
2	Do you remove cookies?	.215	.600
3	Do you use a pop up window blocker?	.030	.745
4	Do you check your computer for spy ware?	.047	.750
5	Do you clear your browser history regularly?	.150	.616
6	Do you block messages / emails from someone you do not want to hear from?	.212	.451

Note. The instructions accompanying the scales were "*For this part of the survey, we are interested in your privacy related behavior in general and when online. Please answer every question using the full scale provided*". Participants responded using a 5-point scale for each item (never – always).

Table 2
Privacy Attitude factor loadings.

Item	Content	Factor 1 Loading
Privacy Concern		
1	In general, how concerned are you about your privacy while you are using the internet?	.688
2	Are you concerned about online organisations not being who they claim they are?	.726
3	Are you concerned that you are asked for too much personal information when you register or make online purchases?	.577
4	Are you concerned about online identity theft?	.753
5	Are you concerned about people online not being who they say they are?	.741
6	Are you concerned that information about you could be found on an old computer?	.586
7	Are you concerned who might access your medical records electronically?	.630
8	Are you concerned about people you do not know obtaining personal information about you from your online activities?	.683
9	Are you concerned that if you use your credit card to buy something on the internet your credit card number will obtained / intercepted by someone else?	.738
10	Are you concerned that if you use your credit card to buy something on the internet your card will be mischarged?	.717
11	Are you concerned that an email you send may be read by someone else besides the person you sent it to?	.682
12	Are you concerned that an email you send someone may be inappropriately forwarded to others?	.683
13	Are you concerned that an email you send someone may be printed out in a place where others could see it?	.629
14	Are you concerned that a computer virus could send out emails in your name?	.611
15	Are you concerned about emails you receive not being from whom they say they are?	.629
16	Are you concerned that an email containing a seemingly legitimate internet address may be fraudulent?	.674

Note. The instructions accompanying the scale were "*For this part of the survey, we are interested in any privacy concerns you might have when online. Please answer every question using the full scale provided*". Participants responded using a 5-point scale for each item (not at all – very much).

Table 3

Correlations (Pearson's r) between privacy measures.

	Privacy Concern	Privacy Behavior: General Caution	Privacy Behavior: Technical Protection
Westin privacy score	.308*** (N=749)	.051 (N=750)	.120*** (N=751)
Total IUIPC score	.246*** (N=753)	.172*** (N=753)	.089* (N=754)

*p<.05, ***p<.005.