International Journal of Human-Computer Studies 71(12), 1144-1162 (Special Issue on Privacy Methodologies in HCI)

Dimensionality of information disclosure behavior

Bart P. Knijnenburg [a,b] – bart.k@uci.edu (corresponding author)
Alfred Kobsa [a] – kobsa@uci.edu
Hongxia Jin [b] – hongxia.jin@sisa.samsung.com

[a] Donald Bren School of Information and Computer Sciences, University of California, Irvine
6210 Donald Bren Hall, Irvine, CA 92697, USA

[b] Samsung Information Systems of America
75 West Plumiera Drive, San Jose, CA 95134, USA

Abstract

In studies of people's privacy behavior, the extent of disclosure of personal information is typically measured as a summed total or a ratio of disclosure. In this paper, we evaluate three information disclosure datasets using a six-step statistical analysis, and show that people's disclosure behaviors are rather *multidimensional*: participants' disclosure of personal information breaks down into a number of distinct factors. Moreover, people can be classified along these dimensions into groups with different "disclosure styles". This difference is not merely in degree, but rather also in kind: one group may for instance disclose location-related but not interest-related items, whereas another group may behave exactly the other way around. We also found other significant differences between these groups, in terms of privacy attitudes, behaviors, and demographic characteristics. These might for instance allow an online system to classify its users into their respective privacy group, and to adapt its privacy practices to the disclosure style of this group. We discuss how our results provide relevant insights for a more user-centric approach to privacy and, more generally, advance our understanding of online privacy behavior.

Keywords

Privacy behavior, privacy attitude, information disclosure, measurement, factor analysis, latent class analysis, structural equation modeling.

1 Introduction

Privacy is a very active research topic: every year, over 1200 new books and journal articles have been published with this word in the title (Patil and Kobsa, 2009). Even in the sub-realm of online privacy, these publications cover a wide range of disciplines, such as human-computer interaction (Iachello and Hong, 2007), information systems (Bélanger and Crossler, 2011), personalization (Kobsa, 2007), behavioral economics (Acquisti and Grossklags, 2008), marketing (Caudill and Murphy, 2000), and social psychology (Joinson and Paine, 2007; Joinson et al., 2010). Although attempts to integrate the contributions of these different fields exist (Knijnenburg and Kobsa, 2013a; Smith et al., 2011), this has proven to be a difficult task, as each discipline has its own conceptualization of the notion of "privacy" (Smith et al., 2011).

Two notions that do recur across these disciplines are the concepts of *privacy attitudes* and *privacy behaviors*. Moreover, researchers seem to agree that attitudes have a significant impact on behavior. In fact, the fundamental theory behind integrative models of privacy research (Knijnenburg and Kobsa, 2013b; Li, 2011; Smith et al., 2011; Xu et al., 2008) is Ajzen and Fishbein's Theory of Reasoned Action, which describes the link between attitudes and behaviors (Ajzen and Fishbein, 1977). However, Ajzen and Fishbein noted an "attitude-behavior gap" in their seminal work: the link between attitudes and behavior is not very strong. There is strong evidence in privacy research that such a gap exists for privacy as well (Acquisti and Grossklags, 2005; Acquisti, 2004; Metzger, 2006; Norberg et al., 2007; Spiekermann et al., 2001; van de Garde-Perik et al., 2008). Norberg (2007) et al. use the term "privacy paradox" to refer to this discrepancy between stated privacy attitudes and actual privacy behavior. Ajzen and Fishbein remedy this problem by introducing the mediating concept of *behavioral intentions*, but they note that even these intentions are not always perfectly correlated with actual behavior (Ajzen, 1991). Privacy researchers suggest that the paradox can be overcome by studying people's behavior in realistic situations instead of lab experiments (Bennett, 1995; Smith et al., 2011).

For the measurement of attitudes, several scales have been developed, often through a reasonably rigorous scale-development process (Dinev and Hart, 2004; Malhotra et al., 2004; Smith et al., 1996; Stewart and Segars, 2002). In each of these efforts, privacy concerns turned out to be multidimensional: privacy attitude is not a single construct, but an interplay of correlated but conceptually distinct aspects, such as "control", "collection", "identification", "improper access", "unauthorized use" and "awareness".

In contrast to attitudes, comparatively few studies have been conducted on privacy-related behavior, and specifically on information disclosure behavior[1]. The research in this area can be broadly divided into two approaches (see section 2 for a detailed literature overview). The first approach regards the disclosure of each item of personal information (e.g. location, gender, income) as a separate decision, and makes no assumptions about correlations between these decisions. In the absence of a theory of how different disclosure behaviors are related, this work does not define an overall measure of a person's rate of disclosure (or *disclosure tendency*). This work typically also does not try to explain how disclosure behaviors come about, or how they can be influenced.

The other approach treats the aggregate of individual disclosure behaviors as a single scale (see section 2). By summing individual disclosures into an overall measurement of disclosure tendency, these researchers make an implicit assumption of unidimensionality of the information disclosures (i.e. they assume that all items belong to the same scale), and even exchangeability of the disclosed items (i.e. they assume that each item contributes the same amount of "evidence" to the scale). The construction of a disclosure tendency scale allows these researchers to, e.g., find antecedents in terms of covariates and manipulations of disclosure behavior. In doing so, they might however oversimplify the actual structure of the disclosure behavior (i.e. some behaviors may be more strongly related than others), thereby violating one of the preconditions of unidimensional measurement.

[1] In this paper we consider both behavioral intentions and actual behaviors, and we provide results in section 7.5 that suggests the two are sufficiently related. We therefore refer to both of them as "behavior", unless our argument calls for a distinction.

In this paper, we argue that information disclosure behaviors are in fact multidimensional, i.e. that different people have different tendencies to disclose different types of information. Our statistical results suggest that regardless of people's overall tendency to disclose information, different categories of information can be distinguished and people be classified into distinct groups that behave differently with regard to the disclosure of information belonging to these categories.

Classifying people according to their privacy concerns is not a new idea; in fact, one of the most cited results in privacy research is that people can be divided into three broad categories: privacy fundamentalists, pragmatists, and unconcerned (Harris et al., 2003a; Harris, 2000; Westin and Maurici, 1998; Westin et al., 1981). Our classification is different though in two ways: First, we classify on behavior rather than attitudes, which, given the mentioned attitude-behavior gap, is an arguably more accurate classification. Second, we argue that privacy categorization should not just consider a difference in degree, but also a difference in kind: for example, one group may be less likely to disclose their location, while another group may be less likely to disclose their opinions.

Although the notion of multidimensional information disclosure behaviors and the idea of classifying people along these dimensions seems fairly straightforward, this approach has to date hardly ever been considered in the privacy literature. The multidimensional analyses by, e.g., Phelps, Nowak and Ferrell (2000), Spiekermann et al. (2001), Olson et al. (2005), Koshimizu et al. (2006) and Lusoli et al. (2012) are notable but limited exceptions.

The next section describes existing research that measures information disclosure behaviors, and shows that (with the mentioned exceptions) these measurements either do not make any assumptions about dimensionality, or are unidimensional in nature. Section 3 explains in more detail what it means for disclosure behaviors to be multidimensional, and discusses why it is important for researchers to conceptualize disclosure behaviors in this way. Section 4 details the 6-step analysis that we performed to describe the dimensionality of disclosure behavior, and clarifies in what way this approach is a step beyond the aforementioned multidimensional analyses. Sections 5 to 7 describe the results of our analyses of the dimensionality of disclosure behavior in three datasets and also discuss how different groups of users behave differently along the discovered dimensions. Two of the discussed datasets were previously collected (one of them by other researchers), and we try to uncover the dimensionality in these datasets *ex post*. The third dataset was specifically collected for this paper, and we formulate *ex ante* hypotheses about its underlying dimensional structure. Finally, section 8 draws conclusions and makes suggestions for future work.

2 Related work

Studies that focus on the disclosure of small amounts of personal information typically treat users' disclosure of each requested item as independent:

- Acquisti, John and Loewenstein (2011, study 1) investigate the effect of social information on participants' tendency to admit having engaged in six sensitive behaviors. They treat these behaviors independently, and show an effect of social information on all of them.
- Joinson et al. (2008, study 2) test the effect of priming participants with a privacy policy on their subsequent disclosure behavior. They allow participants to opt out of disclosure by either choosing "prefer not to say", or by "blurring" their answer (providing a less

concrete value). They use three items (income, religion and ethnicity), and show that a privacy policy has an effect when these three behaviors are summed together. However, when treated separately, the items on which their manipulation has an effect are different for men (income and religion) and women (ethnicity). It thus remains unclear whether the separated or rather the summated results should be considered as the best representation of participants' behavior.

Treating the disclosure of different items as independent behaviors is certainly not wrong, but it typically comes at the cost of reduced statistical power. One also has to consider the family-wise error of performing a large number of statistical tests: when comparing users' disclosure of 20 different items of personal information, one is likely to find one item that tests significant at the $p < .05$ level by pure chance. Studies that consider the disclosure of a larger number of items therefore typically use a summated composite score to represent disclosure behavior:

- In a series of studies, Metzger (2007, 2006, 2004) examines the effects of privacy policies, trust and previous experience on information disclosure to an online retailer. She creates two composite scores of disclosure: a simple sum, and a sum weighed by the relative sensitivity of the items.
- Similarly, Joinson et al. (2010) measure the effect of privacy and trust on a summated score of information disclosure. In study 1, they show that the effect of perceived privacy on a composite score of disclosure is mediated by trust. In study 2, they manipulate trust and privacy through interface cues, and measure the effect on another composite score of disclosure (comprising the four most sensitive items from study 1). They find that disclosure is substantially lower only when the system employs both a weak privacy policy and cues designed to reduce trust.
- Similarly, John, Acquisti and Loewenstein (2011) investigate the effect of contextual cues (a frivolous survey versus a serious survey) on admittance of sensitive behaviors. They sum up participants' answers into a single score (the "affirmative admission rate"), and show that these rates are higher for frivolous surveys than for serious surveys.

It would be interesting to revisit the data from each of these experiments and analyze the dimensionality of the disclosure behaviors. For example, a closer inspection of John et al.'s (2011) results reveals that the effect of contextual cues differs per behavior. For instance, in study 1B, the effect seems to be strongest for financial behaviors, whereas in study 2A the effect is most pronounced for legal sexual acts. A factor analysis of the specific behaviors could categorize the behaviors in interesting ways, possibly leading to new insights. The same is true for the studies of Metzger and Joinson et al., because the disclosure items requested in these studies span a wide range of domains.

There are other studies that group items into a number of distinct scales, but most often these groups are based on the sensitivity of the item instead of a tested underlying dimensionality:

- In study 1 of Joinson et al. (2008), they distinguish between sensitive and non-sensitive items. Their results show that priming had an effect on both types of information.
- In study 2 of Acquisti, John and Loewenstein (2011), they study the effect of request order on "tame", "moderate" and "intrusive" items. Request order only seems to influence the disclosure of intrusive items.
- Knapp and Kirk (2003) test the effect of different survey administration methods (touch-tone phone, Internet and paper) on the disclosure of 60 items ranging from innocuous ("Do you own pet?") to very sensitive ("Have you ever been in jail?"). They test the effect of administration method on each of these items separately, and do not find any

effect. Subsequently, they sum items for three different sensitivity levels, but still find no effect.

Although these studies make an effort to test the effect of their manipulation on different groups of items, we contend that a factor analysis of the reported behavior could categorize the disclosure of the behaviors differently, potentially leading to interesting new results. For instance, the Acquisti et al. items seem to fall along the dimensions of sexual, financial, larcenous, work-related, and impression management behaviors. It would be interesting to see how their manipulations have different effects on each of these different types of data. Although it is not certain that a domain-related grouping is more insightful than a sensitivity-related grouping, a data-driven dimensionality approach would arguably have resulted in more robust dimensions. For example, Knapp and Kirk acknowledge that the Cronbach's alphas of their composite scores are low (between 0.37 and 0.61) because items within each group are from non-related domains. Domain-specific composite scores would likely have resulted in higher Cronbach's alphas.

Studying a music recommender, Van de Garde-Perik et al. (2008) make a distinction between music preference items and personality trait items. Their use of two different information types is in line with our approach, but the authors unfortunately do not request users to make a separate decision per individual item (only per information type), and they do not report the correlation between participants' disclosures of the two information types.

Finally, Norberg, Horne and Horne (2007) test the effect of trust on the intentions to disclose, and on the actual disclosure of 17 items of personal information. They show that participants' intentions to disclose information are lower than their actual levels of disclosure, regardless of whether the receiving party is a trustworthy bank or a less trustworthy pharmaceutical company. Moreover, they show that perceptions of risk, but not of trust, are related to participants' intention to disclose. Interestingly, the authors use different items for the pharmaceutical company (more health-related items) than for the bank (more finance-related items). This invalidates their additional finding that both intention to disclose and actual disclosure are higher for the bank than for the pharmaceutical company, because the metrics taken in both cases are incomparable. The authors should have taken a multi-dimensional approach instead, and should have considered how separate measures of general, financial and health-related disclosures differ between the two scenarios.

3 Multidimensionality: why bother?

The previous section demonstrates that most existing research on disclosure behavior either does not make any assumptions about its dimensionality, or regards behavior as unidimensional. We argue instead that disclosure behaviors are in fact multidimensional. What does that mean? Let's take a hypothetical website that asks users to (optionally) disclose ten items of personal information, $I_{1...10}$. Most researchers would agree that people's tendencies to disclose these individual items are correlated (i.e. people have an overall *disclosure tendency* that holds for any type of information). However, it may be the case that the correlations among the disclosure tendencies for items $I_{1..5}$ are stronger than the correlations for these items with the other items, and the same may hold true for $I_{6...10}$. In that case, there are potentially two (correlated) *factors* of disclosure behavior underlying the disclosure of these ten items. In other words: there are two disclosure tendencies: the tendency to disclose $I_{1...5}$ and the tendency to disclose $I_{6...10}$. This essentially means that although there may be some people who have no problems disclosing all ten items and some people who do not disclose any of them, there also exists a sizable group of

people who tend to disclose $I_{1...5}$ but rather not $I_{6...10}$ and/or[2] a group of people for whom the opposite holds true.

If this is indeed the case, then a multidimensional conceptualization of disclosure behavior will provide a more accurate description of individual behaviors than a "compound" measure that just sums up all individual disclosures (Gardner, 1996). To see the problem of compound measurement, imagine adding up people's weight (in pounds) and height (in inches) and then comparing them on this compound measure. Three persons who each score 230 on this scale may have rather different appearances: an obese person measuring 5 feet, a 6 feet person of normal weight, and an underweight person measuring 6'5". Although certain versions of such compound measures have valid uses (e.g. the Body-Mass Index), describing people in terms of both their weight and height separately arguably portrays their appearance much better.

Better descriptions of people's disclosure behaviors will increase the robustness of behavioral measurements, which will in turn improve the statistical quality of studies that use such measures. For example, if a privacy seal influences people's disclosure of financial information but not their disclosure of medical information, this effect can only be uncovered by treating information disclosure as a multidimensional concept. The fictitious example in Table X demonstrates that when the two types of information are summed, the measurement error (*sd*) increases and the effect of the seal on financial information will be "muddled" by the absence of an effect on medical information. Because of this, a simple t-test finds no effect, even though there actually is an effect for financial information.

Participant	Seal?	Financial items	Medical items	Total items (sum)
1	Yes	7	9	16
2	Yes	11	4	15
3	Yes	14	11	25
4	Yes	10	6	16
5	No	6	9	15
6	No	9	4	13
7	No	2	6	8
8	No	3	11	14
		$M_{seal} = 10.5$ $sd_{seal} = 2.89$ $M_{noseal} = 5.0$ $sd_{noseal} = 3.16$ $t(7) = 2.47, p = .043$	$M_{seal} = 7.5$ $sd_{seal} = 3.11$ $M_{noseal} = 7.5$ $sd_{noseal} = 3.11$ $t(7) = 0.00, p = 1.00$	$M_{seal} = 18$ $sd_{seal} = 4.69$ $M_{noseal} = 12.5$ $sd_{noseal} = 4.36$ $t(7) = 1.68, p = .137$

Table 1: Fictitious example of an experiment that shows how summing two types of items can obscure the effect that an experimental manipulation has on the disclosure of only one of the two types of information.

We do not claim that information disclosure behavior is always multidimensional (e.g. if the example in Table X only considered financial items, these could have safely been considered unidimensional), but we argue that it is prudent to test all behavioral data in privacy research for multidimensionality, as part of the "best practice" to increase the likelihood of making valuable discoveries.

[2] Only one of these two groups is necessary to get two factors, although the factors are stronger when both groups exist.

Moreover, in the realm of online privacy, researchers are beginning to advocate a *personalized* approach (Kobsa, 2001; Wang and Kobsa, 2007), in which a system tailors the amount of requested personal information or the justifications for those requests to the user's overall disclosure tendency (Knijnenburg and Kobsa, 2013a). Systems that follow this personalized privacy approach typically represent the systems' predictions of the user's privacy concerns in a *user model*, which the system then employs to decide what privacy practices to follow. For example, if the user model predicts a user to be very sensitive about privacy, a recommender system could request less sensitive personal information from that user and a social network could disclose less information publicly, thereby respecting the user's privacy needs. If the user model instead predicts the user to be unconcerned about privacy, the system can request or disclose more personal information, which may improve the user's experience.

A multidimensional conceptualization of disclosure tendency will improve the predictive accuracy of such personalized systems, and classifying users as proposed in this paper will make these systems considerably more powerful. For example, a social network like Facebook can make use of a finding, e.g., that most of its users fall into one of five groups with fundamentally different information disclosure behaviors. If the system determines that user X belongs to group A, it can deduce, e.g., that the user does not want to disclose location information but is okay with disclosing her opinions and activities. If user Y belongs to group B, the system knows that this user is, e.g., okay with the disclosure of location information and activities, but not of opinions. Based on this knowledge, the system can give user X the "group A treatment": refrain from "geo-tagging" her status updates, but publicly display her political preference on her profile page. User Y instead would get the "group B treatment": her posts are geo-tagged, but her political opinions are hidden from the public eye. Only a multidimensional measurement of information disclosure behaviors and a classification of users can capture such preferences in the system's user model and enable the correct personalized privacy practices.

Aside from its value for personalized privacy, a classification of users' behaviors can be very informative for the design of privacy-related interfaces in general. Interface designers can use the classification to simplify privacy settings interfaces, e.g. by creating shortcuts to settings for specific user-classes, or by grouping certain settings according to the classes that are most likely to use them. The classification can even inform personas for user-centered design (e.g. designers can ask themselves what each type of user would think about a new feature) and user testing (e.g. testers can select at least one of each user from each class for think-aloud usability tests).

In summary, a multidimensional approach to behavioral measurement can have important implications for Human-Computer Interaction. It is therefore no surprise that this approach has already gained wide acceptance in other fields, such as consumer information seeking behavior (Kiel and Layton, 1981) and the study of counterproductive work behavior (Gruys and Sackett, 2003). The aim of our paper is to contribute to the advancement of behavioral measurement in privacy research as well.

4 Method of analysis

We first present the methods used in our 6-step analysis, and contrast them with previous studies on the dimensionality of information disclosure behaviors. The sections thereafter describe the results of applying our dimensionality analysis to three different information disclosure datasets.

4.1 Our analyses: Factor Analysis and Latent Class Analysis

The three datasets used in our study consist of coded disclosure behaviors (1 for disclosure, 0 for no disclosure) or disclosure intentions (7-point scales) for various items and participants. Although we may have specific hypotheses about the relations between items and thus the number of dimensions, we first submit each dataset to a series of Exploratory Factor Analyses (EFAs) to discover the inherent dimensionality of the data (step 1 in Figure 1). In our specific EFAs, we use a Weighted Least Squares extraction method (a weighted version of Minres) and an oblique Geomin rotation method. The WLS estimator treats disclosure intentions as ordinal variables, and disclosures as binomial. For several possible numbers of factors, the Bayesian Information Criterion (BIC; a measure of the parsimony of a model) and Loglikelihood (LL) are measured. The optimal solution is at a minimum of BIC, its successor does not fit significantly better (p-value[3] > .05), and the loglikelihood levels off for successive models.

Next, based on the EFA results and hypothesized relations between items, a Confirmatory Factor Analysis (CFA) is set up to create a "clean" factor model (i.e. setting off-factor loadings to zero). Our specific CFAs use a Weighted Least Squares estimator to estimate the parameters of the model. The CFA is iteratively improved by assessing the communality, cross-loadings, and residual correlations of the individual items. Items that do not fit the model are removed from the analysis. The final model describes the dimensionality of the disclosure behavior (step 2). For the final CFA model, we report the overall model fit[4], convergent validity[5], and discriminant validity[6].

Subsequently, we perform a series Mixture Factor Analyses (MFAs; Muthén, 2007) to classify participants on these factors (step 3). We estimate our MFAs using a maximum likelihood estimator that calculates non-normality robust standard errors. This MLR estimator treats disclosure intentions as (non-normal) linear variables[7], and disclosures as binomial. Like in the EFA case, the model with the most appropriate number of classes is at a minimum of BIC, its successor does not fit significantly better (p-value[8] > .05), and the loglikelihood levels off for

---

[3] For the EFAs, the p-values come from a chi-square test of the difference between the model and its predecessor in terms of -2LL and number of parameters.

[4] A good model has a χ^2 that is not statistically different from a saturated model ($p > .05$). However, this statistic is regarded as too sensitive, and researchers have proposed other fit indices (Bentler and Bonett, 1980). Hu and Bentler (1999) propose cut-off values for these indices to be: $CFI > .96$, $TLI > .95$, and $RMSEA < .05$, with the upper bound of its 90% CI falling below 0.10.

[5] Convergent validity refers to the degree to which the behaviors in each factor are consistent enough to constitute a single dimension. Convergent validity is adequate when AVE > .50. Cronbach's Alpha is acceptable at $\alpha > 0.70$, and good at $\alpha > 0.80$.

[6] Discriminant validity refers to whether the behaviors in two different factors are different enough to constitute separate dimensions. Discriminant validity is adequate when the square root of the AVE of each factor is larger than the highest correlation with other factors.

[7] MFAs are computationally expensive, and treating the disclosure intentions as ordinal variables would exponentially increase the complexity of the analyses.

[8] For the MFAs, the p-values come from a Lo-Mendell-Rubin adjusted LRT test between the model and its predecessor.

successive models. Additionally, we want classes to show good separation on the different factors (i.e. we may prefer certain solutions on substantive grounds). The final MFA solution is compared to a simple Latent Class Analysis (LCA) with the same number of classes, where not the factors but the items themselves are used for the classification (step 4). The LCA uses the same estimator as the MFA. This step is taken to validate that the same grouping occurs when classification is performed without the restrictions imposed by the factor analysis.

Finally, we test for an attitude−>behavior link by measuring the effect of attitudinal factors on the behavioral factors using Structural Equation Modeling (SEM) with a WLS estimator (step 5). Additionally, we test whether there are significant differences between classes[9] in terms of these attitudes, as well as participants' demographics and related behaviors (step 6). Specifically, the attitudes are regressed on the classes using a Multiple Indicators and Multiple Causes (MIMIC) model with a WLS estimator. Demographics and behaviors are regressed on the classes in a simple linear regression model.

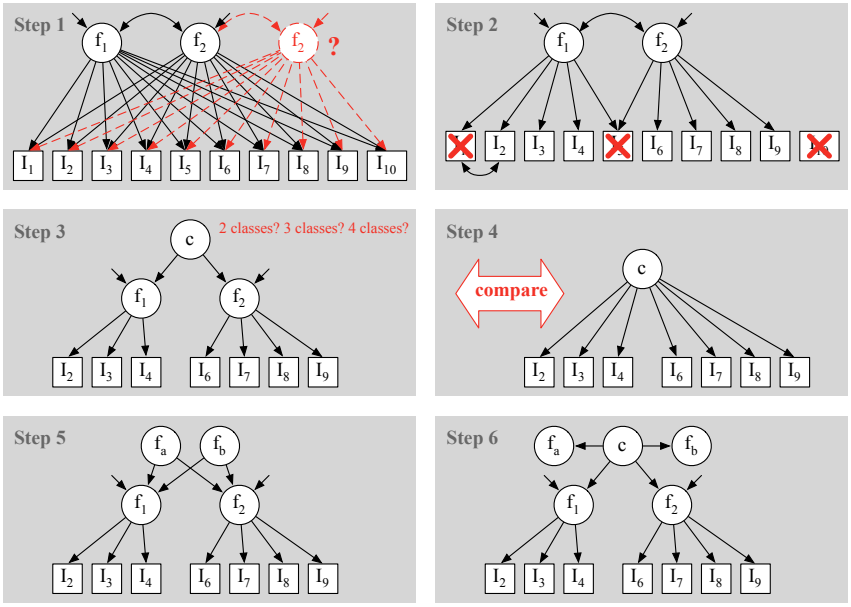

Figure 1: The steps involved in our analysis of the dimensionality of information disclosure behaviors

4.2 Comparison to related work on dimensionality of disclosure behavior

Several previous studies have looked at the dimensionality of information disclosure behaviors and/or at clustering people in terms of their behavior. We now describe these studies, and subsequently explain how our method of analysis takes a step beyond these prior works.

In an experiment with a customer loyalty club at a grocery store, White (2004) investigates whether depth of relationship with the vendor and customized benefits have an effect on participants' willingness to disclose identity-related contact data (address and phone number) and embarrassing information (purchase histories of Playboy/Playgirl and condoms). They show that participants are less willing to disclose embarrassing information than contact data, except when the benefits are customized and the relationship with the vendor is shallow. It is worth noting

[9] To reduce computational complexity, predicted class membership is saved as a categorical variable before running these models.

though that the two distinguished dimensions of information are predetermined *ex ante*, and that no measures of convergent and discriminant validity are reported.

Phelps, Nowak and Ferrell (2000) ask participants about their willingness to disclose items in five predefined categories (demographic, lifestyle-related, purchase-related, personal identifiers, and financial). They show that participants are more likely to disclose the former three categories than the latter two. However, no attempt is made to let the categories/factors emerge from the data. In (Phelps et al., 2001), the authors define three categories on the same data (lifestyle and shopping, personal financial, and demographic), and show that these groups have a high Cronbach's alpha (0.92, 0.80 and 0.87, respectively), indicating high convergent validity. However, they do not discuss the correlations between the different categories; if these are high, this could mean that the categories show low discriminant validity (and thus in fact represent only one factor).

Khalil and Connelly (2006) perform an experience sampling study for the development of a context-aware telephony service. They ask participants at random intervals what they would tell someone (who could be a friend, boss, colleague, family or unknown other) if that person would call them. Options were "where you are", "what you are doing", "whether you are in a conversation" and "whether or not you have company". They find that location and activity disclosures are highly correlated, as are conversation and company disclosures. They also find this latter group of items to be less private than the former group. Although the correlations among four items constitute insufficient evidence to establish two robust classes of information, these results are an interesting precursor for our work on multidimensionality.

Buchanan et al. (2007) use exploratory factor analysis to uncover two dimensions of privacy behavior among 12 items: a general caution dimension and a technical protection dimension[10]. These factors were significantly correlated, despite the fact that they resulted from an orthogonal rotation. Interestingly, they find that their attitudinal measure of "Privacy Concern" is correlated with the general caution dimension, but not the technical protection dimension. The opposite was true for the Westin Privacy Score (Westin and Maurici, 1998). The IUIPC scale (Malhotra et al., 2004) correlated with both general caution and technical protection. Buchanan et al.'s findings provide a very interesting insight into the attitude−>behavior link, but the authors do not try to classify participants in terms of their disclosure behavior.

Koshimizu et al. (2006) apply exploratory factor analysis and clustering to data on participants' feelings about a community-based video surveillance system. They find seven factors and three main clusters of participants differing in their attitudes towards social and authoritative surveillance. Note that the feelings surveyed in this study are more closely related to attitudes than to behaviors or behavioral intentions. Moreover, with only 32 participants (and only 8-15 participants in each cluster), the results may be an artifact of the sample. The authors do not provide the statistical fit of their factor model nor a statistical justification for selecting three clusters.

Lusoli et al. (2012) report a wealth of data collected in a large-scale pan-European survey of privacy practices. Most closely related to our approach is their exploratory factor analysis of personal data disclosed on eCommerce sites. The authors find four factors: social information, biographical information, sensitive information, and security information. They show that

[10] As such these behaviors are not information disclosure behaviors, but we believe that the conducted analysis is interesting and related enough to be reported here nevertheless.

disclosure of these four types of information is quite uniform across European countries, with some differences between northern and southern Europe. The authors make no attempts to cluster participants on these dimensions. Lusoli et al. (2012) also find six dimensions of protection behaviors: reactive practices (e.g. spam- and spyware filters), proactive practices (e.g. contacting websites about their privacy practices), withholding information, minimizing disclosure, avoiding the use of technology, and lying.

Olson et al. (2005) perform a small-scale study on disclosure behavior in an interpersonal privacy context. Specifically, they ask 30 participants how likely they are to disclose 40 different items to 19 different types of people (e.g. sibling, friend, boss, spouse). They find six different dimensions among their 40 types of information:

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

For most of these dimensions, we see no clear theme among the items that would justify their relatedness. Moreover, with only 30 participants the results may again be an artifact of the sample. Olson et al. do not report the statistical quality of their results, nor do they provide statistical evidence for the particular number of dimensions.

Ackerman et al. (1999) study people's intentions to disclose personal information in a generic e-commerce setting. They first cluster participants based on their behavior rather than attitudes. They then contend that participants in different clusters vary in their levels of comfort to disclose several information items, but that the relative sensitivity to these items is consistent across clusters (no statistical evidence is provided for these claims). This suggests that the measured behavioral intentions are unidimensional, but they do not directly test this suggestion.

Similarly, De Souza and Dick (2009) measure disclosure behaviors as a single score, classify participants into two clusters based on attitudes, and then show a difference in behavior between the two clusters. No statistical tests are reported to justify the unidimensional measure or the selection of merely two clusters.

Spiekermann et al. (2001) study disclosure behavior in an ecommerce system with an anthropomorphic online shopping bot. They perform similar clustering on the same attitudinal questions as Ackerman et al. (1999), but their resulting four clusters fall onto two attitudinal dimensions: identity disclosure and profile disclosure. For each of these clusters they then measure participants' tendency to disclose their address (identity-related behavior) and their tendency to answer shopping bot questions (profile-related behavior). They find that behavior differs per cluster, and is in line with the expectations for each cluster. These results suggest that there may be two dimensions of information disclosure behavior in this study, but this dimensionality is derived from the dimensionality of attitudes, and not directly tested on behavior.

Our work improves upon these existing practices in the following ways:

- It derives the dimensionality of the behavior directly from the behavioral data itself;

- It provides statistical justifications for the chosen number of dimensions;
- It classifies[11] participants on their behavior (and not on their attitudes);
- It provides statistical justifications for the selected number of classes.

Each of the six steps proposed in section 4.1 uses a state-of-the-art statistical evaluation technique. Although these techniques have all been used earlier, the current paper is to our best knowledge the first to combine them into an integral procedure to analyze the dimensionality of behavior and clustering of participants on these dimensions. The next three sections will apply the six steps proposed in section 4.1 to three different information disclosure datasets.

5 Dataset 1: Disclosure behavior towards a mobile app recommender

5.1 Study description

This dataset comes from a study with 493 participants (266 female, 223 male; median age group: 25-30, ranging from 18 to older than 60) who were asked to interact with a mobile application that recommends new apps to its users based on their phone usage ("context data") and personal information ("demographics data"). Although the study used a web-based mockup of the mobile application, care was taken to make the study realistic and to ensure that participants had "skin in the game": by hosting the disclosure part of the study on the developer's website, participants were led to believe that their data would be disclosed to the developer of the application.

The system first gave participants a short introduction to the app recommender, including two examples of how their data would be used to generate recommendations. The mockup of the system then requested 31 items, each on a separate screen: 12 context data items and 19 demographic items. The demographic items were further divided into four categories: 5 interests items, 4 person-related items, 5 household-related items, and 3 life-related items. Users could grant or deny permission to the system to collect requested context data with a simple 'yes' or 'no', while for the demographics requests they had to select the actual information from a pull-down menu.

The study, reported in (Knijnenburg and Kobsa, 2013a), tested several *justifications* to encourage participants to disclose each piece of information. Surprisingly, these justifications did not have the expected positive effect, because they increased privacy concerns. The study also manipulated the *order* in which items were requested: the main manipulation was context first versus demographics first; within the demographics questions the order of the four categories was rotated (interest−>person−>household−>life, person−>household−>life−>interest, etc.). Only the main manipulation had an effect: if context was requested first, context data disclosure went up and demographics disclosure went down (compared to requesting demographics first). In the present analyses, we ignore these manipulations and focus on the dimensionality of the behaviors and on the classification of participants on these dimensions.

5.2 Dimensions of behavior

Table 2 shows all items requested in the app recommender study, as well as the percentage of participants who disclosed that piece of information. Several items are disclosed by a very large

[11] In MFA and LCA terminology clusters are called 'classes'.

majority. Participants' behavior on these items thus has a very low variability, which could make the inclusion of these items in a factor analysis problematic.

Type of data	ID	Items	Level of disclosure	Factor loading
Context	1	Recommendation browsing	87.0%	
	2	Location	84.8%	0.767
Alpha: 0.79	3	App usage	82.2%	0.749
AVE: 0.652	4	App usage location	67.1%	
	5	App usage time	73.2%	
Factor correlation: 0.432	6	Web browsing	48.3%	0.874
	7	Calendar data	62.9%	0.835
	8	E-mail messages	36.7%	0.940
	9	Phone model	84.6%	0.659
	10	Accelerometer data	65.3%	
	11	Microphone	50.9%	
	12	Credit card purchases	20.1%	0.796
Demographics	13	Favorite sports (fan)	86.8%	0.718
	14	News interests	92.7%	
Alpha: 0.86	15	Amount of TV watching	92.3%	
AVE: 0.784	16	Amount of reading	93.5%	
	17	Phone data plan	87.6%	0.905
Factor correlation: 0.432	18	Gender	94.9%	
	19	Age	93.3%	
	20	Education	92.7%	
	21	Field of work	83.6%	0.915
	22	Housing situation	87.4%	
	23	Population density of area	90.7%	
	24	Relationship status	88.6%	0.911
	25	Children	89.3%	
	26	Household income	74.2%	0.964
	27	Household savings	66.3%	0.957
	28	Household debt	64.5%	
	29	Race	89.1%	
	30	Political preferences	86.4%	0.802
	31	Workout routine	90.1%	

Table 2: The items used in the app recommender study, along with their average rate of disclosure and the factor loadings of the CFA. Within each group, the ID is numbered in the order of the requests. The dashed lines delineate the four categories of demographics items, the order of which was randomized (these categories did not produce different factors).

5.2.1 *Step 1: Exploratory Factor Analysis (EFA)*

Table 3 and Figure 2 compare the different factor solutions. The two-factor solution has the lowest BIC, the three-factor solution makes the last significant improvement in model fit, and the loglikelihood levels off at the two-factor solution. The two-factor solution nicely splits context items and demographics items, whereas the three-factor solution has an additional "nested" factor with financial items (26, 27, and 28). We thus adopt the two-factor solution, due to its better parsimony.

	BIC	LL	# of par.	*p*-value
1 factor	10316	-4965.574	62	
2 factors	**9174**	-4301.779	92	< .001
3 factors	9213	-4231.195	121	< .001
4 factors	9351	-4213.488	149	**.158**
5 factors	9426	-4167.549	176	< .001

Table 3: A comparison of the fit of different factor solutions.

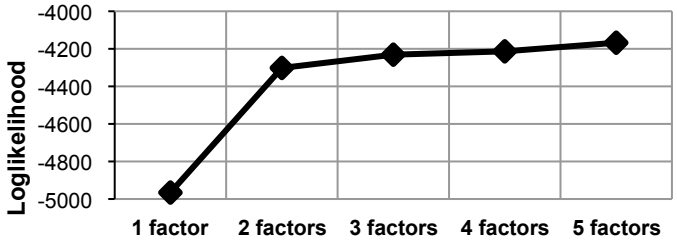

Figure 2: Change in loglikelihood between subsequent factor solutions.

5.2.2 *Step 2: Confirmatory Factor Analysis (CFA)*

The factor loadings of the final CFA solution are presented in Table 2 (removed items have no factor loading). This final solution has a good fit ($\chi^2(76) = 152.15$, $p < .001$; *CFI* = .989, *TLI* = .987; *RMSEA* = .045, 90% CI: [.035, .055]), and the factors show a good convergent and discriminant validity. The two factors are correlated with $r = 0.432$ (significant at $p < .001$).

5.3 Clustering participants

5.3.1 *Step 3: Mixture Factor Analysis (MFA)*

Table 4 and Figure 3 compare the different MFA solutions. The BIC has no maximum, the entropy is highest for four classes, and the five-class solution also does not fit significantly better. The loglikelihood levels off at three classes (see Figure 3), and the likelihood ratio test for the four-class solution is only just significant ($p = .040$). We therefore adopt both the three-class solution and the four-class solution.

	BIC	Entropy	LL	# of par.	*p*-value
1 class	7084		-3455.005	28	
2 classes	6069	0.821	-2944.375	29	.0016
3 classes	5602	0.861	-2701.611	32	< .001
4 classes	5482	**0.865**	-2632.467	35	.0400
5 classes	5414	0.837	-2589.267	38	**.2572**

Table 4: A comparison of the fit of MFA models with different numbers of classes.

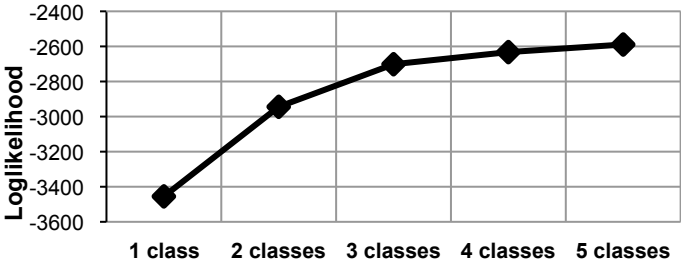

Figure 3: Change in loglikelihood between subsequent MFA models.

The three-class solution (Figure 4, left) shows 97 participants with rather low disclosure tendencies on both context and demographics items (from hereon called "LowD" for low disclosure), 196 participants who are very likely to disclose either context items or demographics items ("HiD" for high disclosure), and 200 participants who are okay with disclosing demographics items, but would rather not disclose context items ("DemoD" for demographics disclosure, but no context disclosure).

The four-class solution (Figure 5, left) shows the same three classes of 67 LowD participants, 179 HiD participants, and 176 DemoD participants. The new class contains 71 participants who are moderate in their disclosure, but their moderation is equally spread among context items and demographics items ("MedD" for medium disclosure).

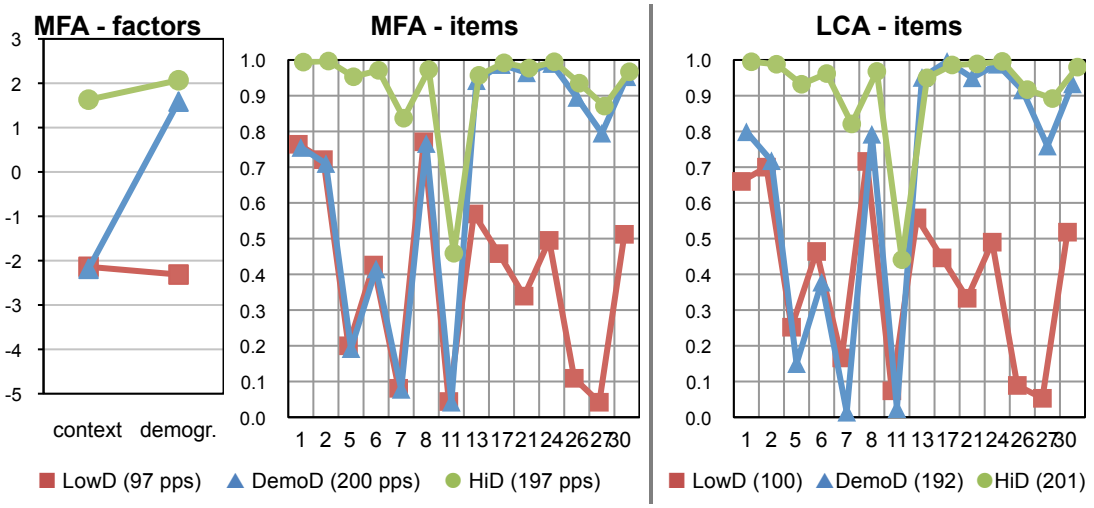

Figure 4: The standardized factor values and item disclosure probabilities for the three-class MFA solution (left), and the item disclosure probabilities for the three-class LCA solution (right).

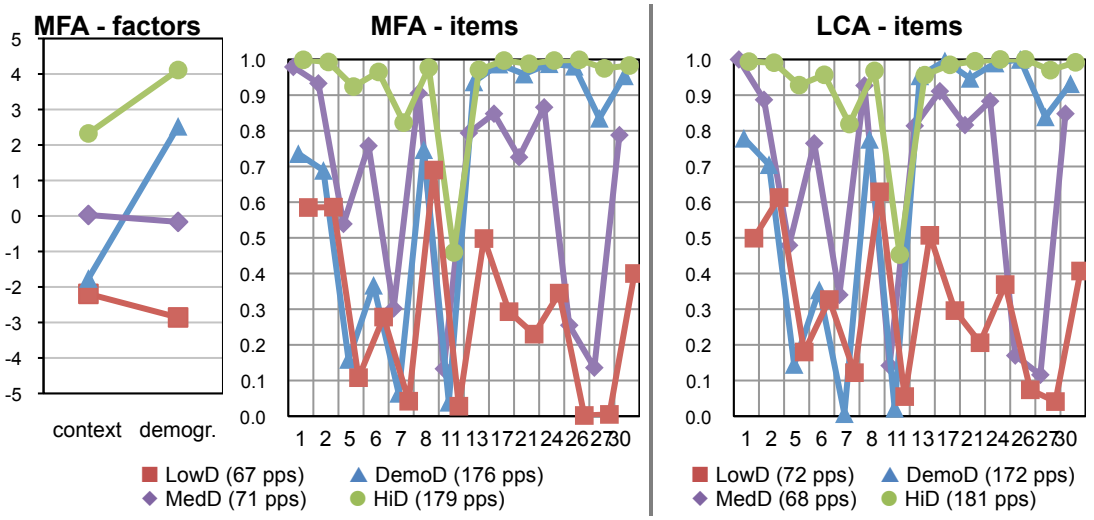

Figure 5: The standardized factor values and item disclosure probabilities for the four-class MFA solution (left), and the item disclosure probabilities for the four-class LCA solution (right).

5.3.2 *Step 4: Latent Class Analysis (LCA)*

The right sides of Figure 4 and Figure 5 show the LCA results. The number of participants in each class differs slightly between the MFAs and the LCAs, but the overall results are similar, indicating that the factors are an adequately simplified representation of participants' behavior.

5.4 Attitude−>Behavior link

In the app recommender study, we used 24 five-point scale items to measure three attitudes (based on (Malhotra et al., 2004; Smith et al., 1996)) and two personal characteristics. Table 5

displays the results of the five-factor CFA. Note that general privacy concerns and collection concerns are very highly correlated (r = .800).

Considered aspects	Items	Factor loading
General privacy concerns Alpha: 0.76 AVE: 0.600	All things considered, the Internet causes serious privacy problems	
	Compared to others, I am more sensitive about the way online companies handle my personal information	0.766
	To me, it is the most important thing to keep my privacy intact from online companies	0.727
	I believe other people are too concerned with online privacy issues	
	I am concerned about threats to my personal privacy today	0.828
Collection concerns Alpha: 0.86 AVE: 0.812	It usually bothers me when online companies ask me for personal information	0.826
	When online companies ask me for personal information, I sometimes think twice before providing it	
	It bothers me to give personal information to so many online companies	0.873
	Online companies may collect any information about me because I have nothing to hide	-0.723
	I'm concerned that online companies are collecting too much personal information about me	0.832
	I'm not bothered by data collection, because my personal information is publicly available anyway	-0.760
Control concerns Alpha: 0.58 AVE: 0.749	Online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used, and shared	0.605
	Control of personal information lies at the heart of online privacy	0.869
	I do not want to think about who controls my personal information	
	I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction	
	I do not feel the need to control my personal information	
Mobile Internet usage Alpha: 0.88 AVE: 0.865	I regularly use my phone to browse the Internet	0.965
	I regularly use my phone to check my e-mail	0.894
	I would feel lost without my cellphone	
	I only have a few apps installed on my phone (or none at all)	
Tech-savvyness Alpha: 0.85 AVE: 0.739	I am not very good with technology	
	People ask me to fix their computer	0.960
	I have detailed knowledge of how most technological products work	0.780
	I usually ask others to fix my computer	-0.829

Table 5: The items used in the app recommender study to measure attitudes and personal characteristics, along with the factor loadings of the CFA (items that do not have a factor loading in the rightmost column were excluded from the analysis).

5.4.1 *Step 5: Predicting behavioral dimensions with attitudes/characteristics*

We regressed the two behavioral factors on these attitudes and characteristics. For context data disclosure, we find significant causal effects of collection concerns ($\beta = -.455$, $SE = .065$, $p < .001$) and mobile Internet usage ($\beta = .251$, $SE = .066$, $p < .001$). For demographics disclosure, we also find effects of collection concerns ($\beta = -.221$, $SE = .068$, $p = .001$) and mobile Internet usage ($\beta = .148$, $SE = .068$, $p = .029$), albeit slightly smaller.

5.4.2 *Step 6: Differences in attitudes/characteristics between classes*

Figure 6 reveals significant differences between classes, for the four-class solution (the results for the three-class solution are similar). On average, participants in the DemoD and LowD classes have higher general privacy concerns than participants in the HiD and MedD classes. In terms of

collection concerns, participants in the HiD class score significantly lower than those in the other classes. Finally, participants in the DemoD and LowD classes make significantly less use of the mobile Internet than in the HiD and MedD classes.

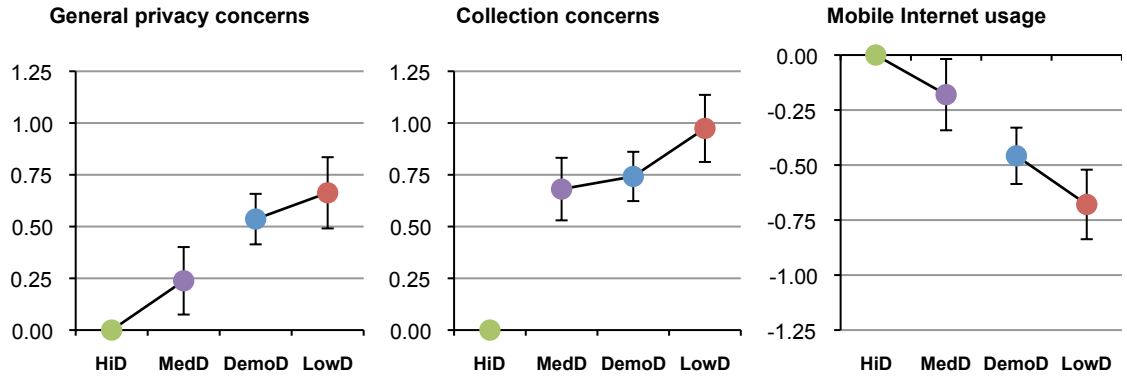

Figure 6: Differences between classes in terms of general privacy concerns, collection concerns, and mobile Internet usage. Points that are not connected are significantly different from one another. Since factor scores have no inherent scale, HiD is fixed to zero, and the vertical axes are scaled in sample standard deviations of the measured factor (i.e. 95% of the participants fall within a range of 4 units). Error bars are ±1 standard error of the difference with HiD.

5.5 Discussion of the results from dataset 1

Our results show that the concept of "disclosure behavior" in the app recommender study is not unidimensional, but involves two correlated dimensions: context data disclosure and demographics disclosure. Moreover, our clustering solution(s) reveal the existence of a class of users who readily disclose their demographics, but who are conservative towards disclosing context items. In a unidimensional model, knowing the context-related behavior of these participants would lead to false assumptions about their demographics-related behavior, and vice versa. The multidimensional model is thus indispensible in describing or predicting people's disclosure behavior in this app recommender.

In terms of an attitude−>behavior link, we observe that collection concerns are a significant predictor of both demographics and context data disclosure. Also, participants who accessed the Internet on their mobile phone were more likely to disclose demographics and context items. This means that systems can gauge the amount of information that users are willing do disclose by observing or asking them about their mobile Internet usage and/or collection concerns. Unfortunately, though, we found no antecedents that are able to distinguish between the two types of disclosure (i.e., the found differences in the antecedents are not significant between the DemoD group and the LowD group).

Some reservations need to be made regarding our results. First of all, some items were disclosed at a very high rate, resulting in very little variance, thereby making them unsuitable for factor analysis. The fact that our results show two dimensions only, despite the wide variety of demographics questions, may also be a result of the low information density in the measured behaviors.

Another methodological issue is that the dimensions are in line with the order and style of requests in this study: both the demographics items and the context items were always requested consecutively and used slightly different user interface elements (pull-down menu for demographics items, "yes" and "no" buttons for context items). Users may have therefore seen

them as two distinct groups, and may have (unconsciously) tried to behave consistently within each group, thereby artificially inflating the robustness of our results. Moreover, the order was actually manipulated between demographics items and context items. The request order influenced the amount of disclosure: demographics-first participants disclosed more demographics, while context-first participants disclosed more context items. This manipulation may thus have further separated the two groups of items. However, we also conducted our analysis separately for each order, and observed similar results. To alleviate any remaining concerns that the thematic grouping of items had a major effect on the discovered dimensions we asked questions in a random order in the experiment that generated dataset 3 (see section 7).

6 Dataset 2: Intentions to make Facebook data publicly accessible

6.1 Study description

This data originated from a cross-cultural comparison of Facebook privacy concerns by Wang et al. (2011). We used the subset of the data that came from the United States participants, with a total of 359 responses (222 female, 137 male; median age: 28, ranging from 18 to 75). After answering a number of questions about their demographics and their Facebook usage, participants in this study indicated on a seven-point scale their level of comfort with disclosing 16 different Facebook profile items to "everyone on the Internet". The order of these questions was fixed, and the answers to them constitute the behavioral intentions we will consider in this section. An additional 54 seven-point scale items and 7 open questions measured various related attitudinal concepts.

6.2 Dimensions of behavior

Table 6 shows all items requested in the Facebook study. The items were phrased as: "How comfortable are you with everyone on the Internet seeing your [item]", each with a seven-point scale anchored at "Not at all comfortable", "Neutral", and "Very comfortable".

Type of data	ID	Items	Level of comfort							Factor loading
			1	2	3	4	5	6	7	
Facebook activity	1	Wall	83	44	32	47	52	53	48	0.820
	2	Status updates	80	56	43	39	55	44	42	0.953
Alpha: 0.93	3	Shared links	64	45	36	68	54	45	47	0.885
AVE: 0.790	4	Notes	102	55	44	55	37	32	34	0.907
	5	Photos	132	55	38	34	37	31	32	0.874
	16	Friend list	60	34	50	73	51	43	48	
Location	6	Hometown	62	50	32	63	61	44	47	0.924
	7	Location (your current city)	72	62	41	56	46	37	45	0.960
Alpha: 0.95	8	Location (your current state/province)	60	54	42	58	53	40	52	0.958
AVE: 0.919										
Contact info	9	Residence (your street address)	240	33	23	23	19	10	11	0.884
	11	Phone number	262	29	15	22	19	4	8	0.933
Alpha: 0.85	12	Email address	159	58	35	41	28	22	16	0.849
AVE: 0.792										
Life and interests	13	Religious views	45	30	27	109	50	33	65	0.740
	14	Interests (favorite movies, books, etc.)	32	26	36	79	72	51	63	0.913
Alpha: 0.88	15	Facebook groups that you are a member of	35	33	38	79	65	51	58	0.942
AVE: 0.756										
	10	Employer	110	53	36	73	43	24	20	

Table 6: The items used in the Facebook study, along with the frequencies at each level of comfort, and the factor loadings of the CFA. The ID parallels the request order.

6.2.1 *Step 1: Exploratory Factor Analysis (EFA)*

Table 7 and Figure 7 compare the different solutions. The four-factor solution has the lowest BIC, and the five-factor solution does not fit significantly better. Moreover, the loglikelihood clearly levels off at four factors. We therefore adopt the four-factor solution.

	BIC	LL	# of par.	*p*-value
1 factor	20611	-10164.489	48	
2 factors	20207	-9918.105	63	< .001
3 factors	19574	-9560.411	77	< .001
4 factors	**19320**	-9395.040	90	< .001
5 factors	19360	-9379.961	102	**0.237**
6 factors	19402	-9368.779	113	0.428

Table 7: A comparison of the fit of different factor solutions.

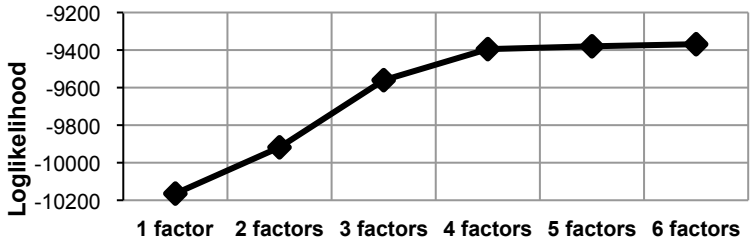

Figure 7: Change in loglikelihood between subsequent factor solutions.

6.2.2 *Step 2: Confirmatory Factor Analysis (CFA)*

The factor loadings of the final CFA solution are presented in Table 6. This model shows some misfit ($\chi^2(71) = 370.19$, $p < .001$; *CFI* = .985, *TLI* = .980; *RMSEA* = .108, 90% CI: [.098, .119]), but the factors have a good convergent and discriminant validity. Table 8 shows the factor correlations.

Location	.732		
Contact	.711	.642	
Interests	.775	.696	.490
	Activity	**Location**	**Contact**

Table 8: Correlations between factors (all are significant at $p < .001$).

6.3 Clustering participants

6.3.1 *Step 3: Mixture Factor Analysis (MFA)*

Table 9 and Figure 8 compare the different MFA outcomes. For three classes, the BIC is at a minimum, and four classes do not fit the data significantly better. The five-class solution shows a nice distribution of classes over factors, and we adopt this solution for this reason: a classification that shows how groups of people exhibit substantially different behaviors on the four factors is arguably more useful (e.g., for user modeling) than a low−medium−high classification.

	BIC	Entropy	LL	# of par.	*p*-value
1 class	16837		-8277.147	48	
2 classes	16578	0.973	-8133.179	53	.0069
3 classes	**16442**	**0.998**	-8050.552	58	.0002
4 classes	16468	0.998	-8048.736	63	**0.407**
5 classes	16482	0.878	-8041.459	68	0.999
6 classes	**16351**	0.897	-7960.902	73	0.812
7 classes	16359	0.852	-7950.412	78	0.893

Table 9: A comparison of the fit of MFA models with different numbers of classes.

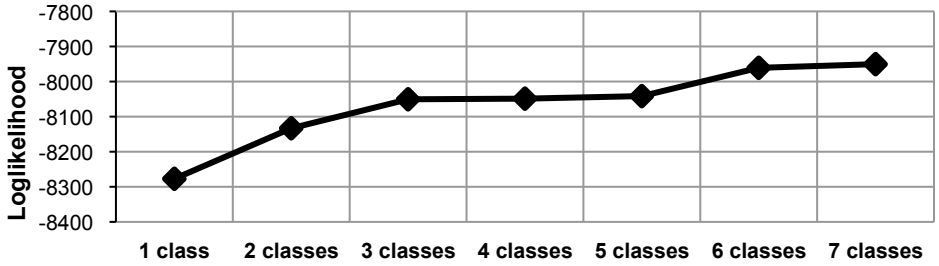

Figure 8: Change in loglikelihood between subsequent MFA models.

The three-class solution (Figure 9, left) shows 291 participants with rather low disclosure tendencies on all dimensions (LowD), 56 participants who are very likely to disclose any type of information (HiD), and 12 participants who are more or less in between the two other classes (MedD).

The five-class solution (Figure 10, left) shows 159 LowD participants; 59 HiD participants; a class of 65 participants with a low intention to disclose contact information ("Hi−ConD"); a class of 50 participants who have a low intention to disclose contact information and Facebook activity, but a high intention to disclose location and interests ("Loc+IntD"); and a class of 26 participants with a low intention to disclose contact information and location, but a high intention to disclose Facebook activity and interests ("Act+IntD").

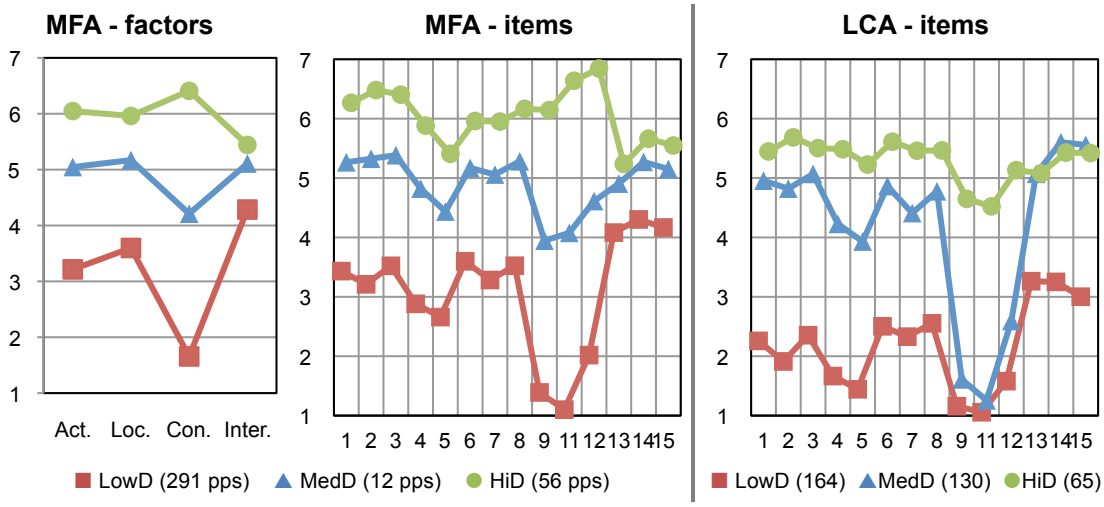

Figure 9: The factor values and item disclosure tendencies for the three-class MFA solution (left), and the item disclosure tendencies for the three-class LCA solution (right).

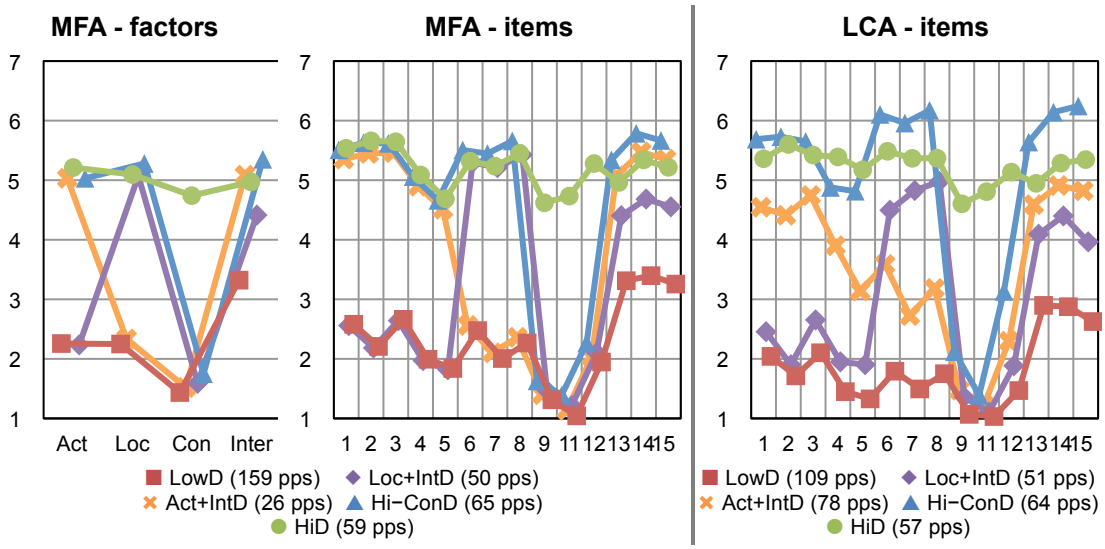

Figure 10: The factor values and item disclosure tendencies for the five-class MFA solution (left), and the item disclosure tendencies for the five-class LCA solution (right).

6.3.2 *Step 4: Latent Class Analysis (LCA)*

The right sides of Figure 9 and 10 show the LCA results. For the three-class solution, MedD in the LCA (130 participants) is very different from the MFA (only 12 participants). This means that the three-class solution is not very robust. The five-class LCA resembles the MFA much better, which indicates that the five-factor solution is an adequately simplified representation of participants' behavior. The only difference is the Act+IntD class, which is less pronounced on the low location disclosure intentions in the LCA than in the MFA.

6.4 Attitude−>Behavior link

We extracted three attitudinal factors from the Facebook survey based on 9 survey items[12]. Table 10 displays the results of the three-factor CFA.

Considered aspects	Items	Factor loading
Knowledge about privacy policy Alpha: 0.82 AVE: 0.679	I have read Facebook's privacy policy thoroughly	0.949
	I did not read Facebook's privacy policy in detail	-0.862
	I stay up to date on Facebook's privacy policy changes	0.629
Trust in Facebook Alpha: 0.74 AVE: 0.531	I trust Facebook with my personal information	0.818
	I feel that Facebook employs trustworthy staff members	0.688
	I feel that data on Facebook's servers is secure against intruders	0.672
Need for consent Alpha: 0.72 AVE: 0.577	Facebook should not be able to share my information unless I specifically give them permission	0.710
	Facebook should announce any planned changes in advance	0.856
	Facebook should ask for user input before making changes	0.702

Table 10: The attitudinal items used in the Facebook study, along with the factor loadings of the CFA.

[12] Although the survey contained 54 attitudinal items, only these 9 converged to a stable factor solution.

6.4.1 *Step 5: Predicting behavioral dimensions with attitudes*

Table 11 shows the results of the regression of the four behavioral factors on the three attitudes. Interestingly, people with a high desire for consent are much less likely to give out their contact information, while no correlation exists for interests. People's knowledge of Facebook's privacy policy makes them less likely to disclose their location and interests. Trust has a positive effect on all disclosure behaviors, and especially on interest-related items.

	Knowledge about privacy policy	**Trust in Facebook**	**Need for consent**
Activity	*ns*	$\beta = .303$ (.066), $p < .001$	$\beta = -.254$ (.066), $p < .001$
Location	$\beta = -.100$ (.047), $p = .035$	$\beta = .333$ (.069), $p < .001$	$\beta = -.144$ (.066), $p = .030$
Contact	*ns*	$\beta = .283$ (.079), $p < .001$	$\beta = -.580$ (.072), $p < .001$
Interests	$\beta = -.161$ (.050), $p = .001$	$\beta = .489$ (.066), $p < .001$	*ns*

Table 11: Regression coefficients (standard errors) regressing the behavioral factors on the attitudinal factors (attitude −> behavior).

6.4.2 *Step 6: differences in attitudes/characteristics between classes*

Figure 11 reveals significant differences in trust and need for consent for the five-class solution. On average, participants in the LowD class have less trust in Facebook, while participants in the HiD and Hi−ConD classes have more trust as well as a lower need for consent. Knowledge about Facebook's privacy policy was not significantly related to class membership. We did however find differences in age and gender between classes (Figure 12). Specifically, participants in the HiD class are on average significantly younger than participants in the LowD and Hi−ConD classes, indicating that younger people tend to disclose more. There is also a significant gender difference between the HiD class and all other classes, indicating that males tend to disclose more on average.

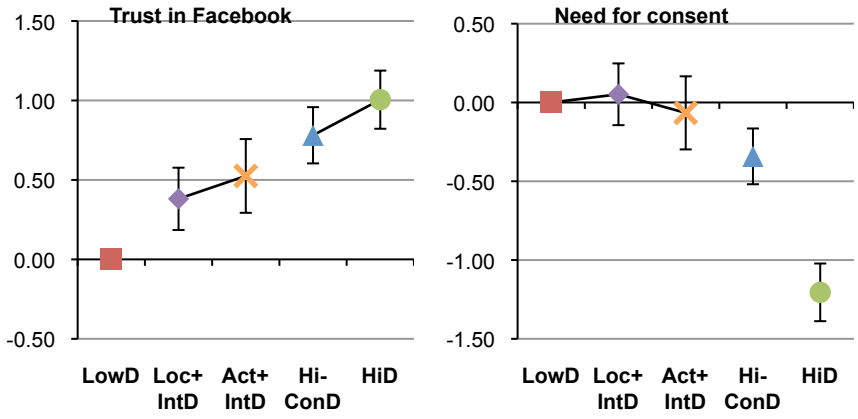

Figure 11: Differences between classes in terms of trust in Facebook and need for consent. Points that are not connected are significantly different from one another (except for Act+IntD and Hi−ConD). LowD is fixed to zero, and the vertical axes are scaled in sample standard deviations of the measured factor. Error bars are ±1 standard error of the difference with LowD.

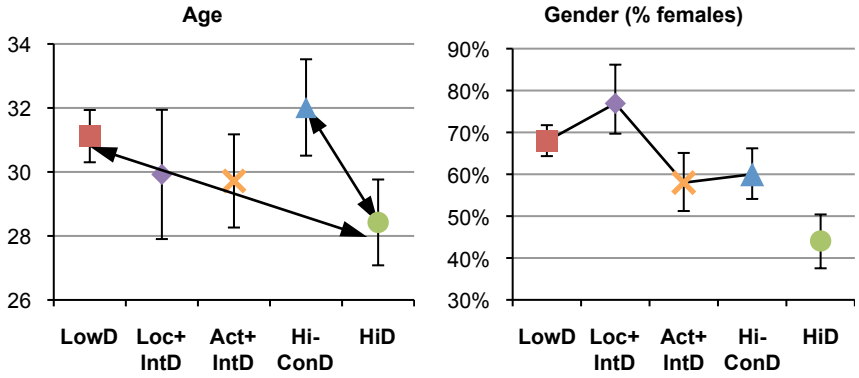

Figure 12: Differences between classes in terms of age and gender. For age, arrows indicate a significant difference between HiD compared to LowD or Hi−ConD. For gender, points that are not connected are significantly different from one another. Error bars are ±1 standard error of each measurement.

6.5 Discussion of the results from dataset 2

The data from the Facebook privacy study has four dimensions, and five different groups of users seem to exhibit inherently different privacy behaviors along these dimensions. Any system that wants to make optimal use of users' personal information should thus distinguish the different types of information and users, and try to identify which user belongs to which group. This identification can consider demographic characteristics such as age and gender (young, male participants are more likely to fall into the HiD class), but could also measure attitudes such as trust in Facebook or need for consent. All this information can be used to predict the user's group, and to request appropriate information accordingly (e.g., don't solicit contact data from people who are unlikely to provide it).

In this dataset, the most interesting finding is the difference between the Loc+IntD and the Act+IntD groups: whereas the former tends to disclose their location but not their Facebook activities, the latter does the opposite. To achieve the highest disclosure rate of users' location and Facebook activity, systems such as Facebook apps will therefore need to determine which users fall in the activity-concerned group and which ones in the location-concerned group (and of course, which users fall in neither or both). This is particularly important if the number of items that the system may request is limited. Unfortunately though, information about attitudes and age/gender do not provide much help in distinguishing these groups.

7 Dataset 3: Intentions to disclose to an online retailer

7.1 Study description

The "online retailer dataset" was gathered specifically for this paper, in order to broaden its empirical basis, to test the claim of multidimensionality as an *ex ante* hypothesis, and to alleviate concerns that the thematic grouping of items in the two previous studies had a major effect on the discovered dimensions. We constructed our 24 items in such a way that we could formulate hypotheses about their dimensionality: 6 of them were related to health, 6 to interests, 6 to work, and 6 to more general issues including contact information. We asked these questions in a random order, so that our hypothesis about dimensionality would not be confounded with the grouping of requests.

154 people participated in our study (69 females, 84 males; median age: 29, ranging from 18 to 65). For each of the 24 items, we first asked them to enter the answer into a text field, with the option to rather not disclose it. We then asked for each item how likely they were to provide the answer to an online retailer. The 24 answers to this question constitute the behavioral intentions used for the analysis in this section. Subsequently, we measured participants' privacy attitudes with 19 additional questionnaire items.

To control for order effects, we defined two random orderings of the 24 requests, one being exactly the reverse of the other. Like in the app recommender dataset, order had a very strong effect on disclosure.

7.2 Dimensions of behavior

Table 12 shows all items from the online retailer study. The items were phrased as: "How likely are you to submit your [item] to an online retailer?", each with the seven possible answers "very unlikely", "unlikely", "somewhat unlikely", "neutral", "somewhat likely", "likely", and "very likely".

Type of data	ID	#	Items	Level of comfort							Factor loading
				1	2	3	4	5	6	7	
Health	A1	8	Physical health	21	20	13	19	31	31	19	
Alpha: 0.92	A2	23	Number of doctor visits in the past month	36	21	18	23	19	17	20	0.919
AVE: 0.782	A3	20	Weight (lbs)	25	26	19	24	24	22	14	0.839
	A4	22	Dietary restrictions	19	17	15	26	27	27	23	0.837
	A5	12	Whether you use birth control	43	26	15	26	12	17	15	0.897
	A6	14	Whether you have diabetes	22	27	13	20	25	20	27	0.924
Interests	B1	7	Favorite pastime	13	13	10	22	31	35	30	
	B2	2	Favorite musical band/artist	6	12	11	23	27	44	31	0.894
Alpha: 0.91	B3	4	Favorite food	5	11	5	24	26	49	34	0.929
AVE: 0.850	B4	21	Favorite movie	7	10	9	20	33	40	35	
	B5	24	Last holiday location	23	16	11	26	25	24	29	
	B6	6	Relationship status	16	23	9	19	31	33	23	
	B7	3	Computer software you are familiar with	8	9	11	28	31	41	26	0.908
Work	C1	5	Highest completed degree	15	16	14	24	25	33	27	
	C2	9	Work experience (years)	18	24	7	26	27	29	23	0.943
Alpha: 0.93	C3	18	Current/previous occupation	27	27	18	19	27	19	17	0.916
AVE: 0.823	C4	13	Current/previous field of work	15	25	16	25	31	24	18	0.920
	C5	19	Current/previous income level	40	31	20	23	17	14	9	0.884
Contact info	D1	1	Name	16	21	15	17	25	36	24	
	D2	16	Gender	5	9	4	22	29	43	42	
Alpha: 0.87	D3	15	Age	8	11	7	27	37	37	27	
AVE: 0.761	D4	17	Address	43	32	4	18	24	17	16	0.912
	D5	10	E-mail address	15	30	13	19	35	23	19	0.860
	D6	11	Phone number	50	37	14	21	19	7	6	0.844

Table 12: The items from the online retailer study, with frequencies at each comfort level and factor loadings of the CFA. '#' indicates the request order (randomized, reversed in the second condition)

7.2.1 *Step 1: Exploratory Factor Analysis (EFA)*

Table 13 and Figure 13 compare the different factor solutions. The five-factor solution has the lowest BIC, but the loglikelihood seems to somewhat level off at four factors. Comparing the optimal five-factor CFA solution with the four-factor solution (step 2), we find an additional factor with only the items Gender and Age, which are admittedly different from the other contact

info items. We feel however that two items are not enough to constitute a separate "type" of information, and therefore choose to adopt the four-factor solution.

	BIC	LL	# of par.	*p*-value
1 factor	13000	-6318.468	72	
2 factors	12749	-6135.381	95	< .001
3 factors	12550	-5980.160	117	< .001
4 factors	12418	-5861.396	138	< .001
5 factors	**12399**	-5801.487	158	< .001
6 factors	12409	-5758.677	177	< .001

Table 13: A comparison of the fit of different factor solutions.

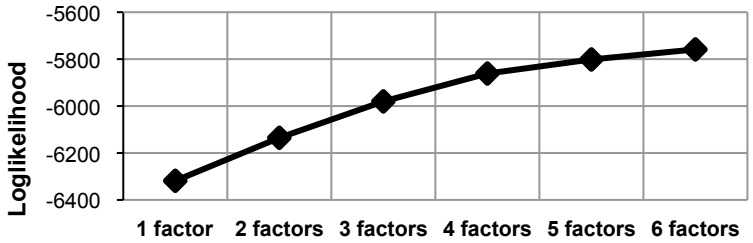

Figure 13: Change in loglikelihood between subsequent factor solutions.

7.2.2 *Step 2: Confirmatory Factor Analysis (CFA)*

The factor loadings of the final CFA solution are presented in Table 12. Note that all items fall on the hypothesized factors, except for the "computer software" item. We predicted that this item would load on the work factor, but instead it loaded on the interest factor. In hindsight, this makes perfect sense.

The model shows some misfit ($\chi^2(84) = 221.77$, $p < .001$; *CFI* = .985, *TLI* = .981; *RMSEA* = .103, 90% CI: [.087, .120]), but the factors exhibit good convergent and discriminant validity. Table 14 shows the factor correlations. Note that the contact info factor is *negatively* related to the other factors, which means that participants who disclose their contact information are *less* likely to disclose the other items, and vice versa. The contact info factor thus differs particularly strongly from the other factors.

Interests	.769		
Work	.880	.758	
Contact	-.488	-.308	-.327
	Health	**Interests**	**Work**

Table 14: Correlations between factors.

7.3 Clustering participants

7.3.1 *Step 3: Mixture Factor Analysis (MFA)*

Table 15 and figure 14 compare the different MFA solutions. Interestingly, the two-class solution is not significantly better than the one-class solution, indicating that classifying participants may not be necessary to adequately represent the data. In other words: many participants in this study behaved very similarly (this may be due to a comparatively low number of participants). On the other hand, the BIC of the four-class solution is at a minimum, and this is also the point where the loglikelihood seems to level off. We therefore adopt this solution.

	BIC	Entropy	LL	# of par.	*p*-value
1 class	7854		-3798.693	51	
2 classes	7849	.856	-3783.708	56	0.170
3 classes	7843	.892	-3767.770	61	0.364
4 classes	**7842**	.858	-3754.985	66	0.572
5 classes	7854	.893	-3748.100	71	0.515
6 classes	7868	.850	-3743.010	76	0.627

Table 15: A comparison of the fit of MFA models with different numbers of classes.

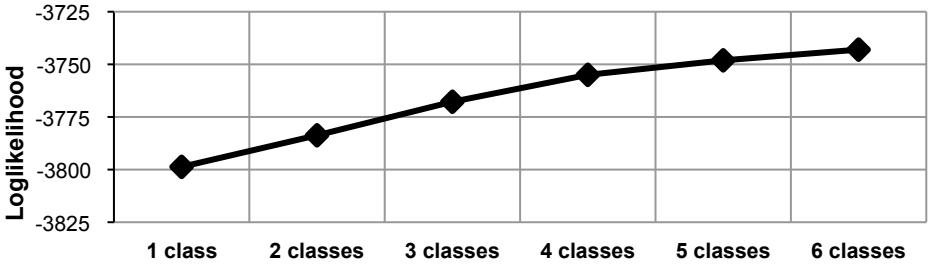

Figure 14: Change in loglikelihood between subsequent MFA models.

The four-class solution (Figure 15, left) shows 26 participants with rather low disclosure tendencies on all dimensions except contact info (ConD), 40 participants who are very likely to disclose any type of information except contact information (Hi−ConD), 65 participants who have a low intention to disclose contact info and medium tendencies on all other factors (Med−ConD), and 23 participants who have low tendencies to disclose health and work, and medium tendencies to disclose interests and contact info (Int+ConD).

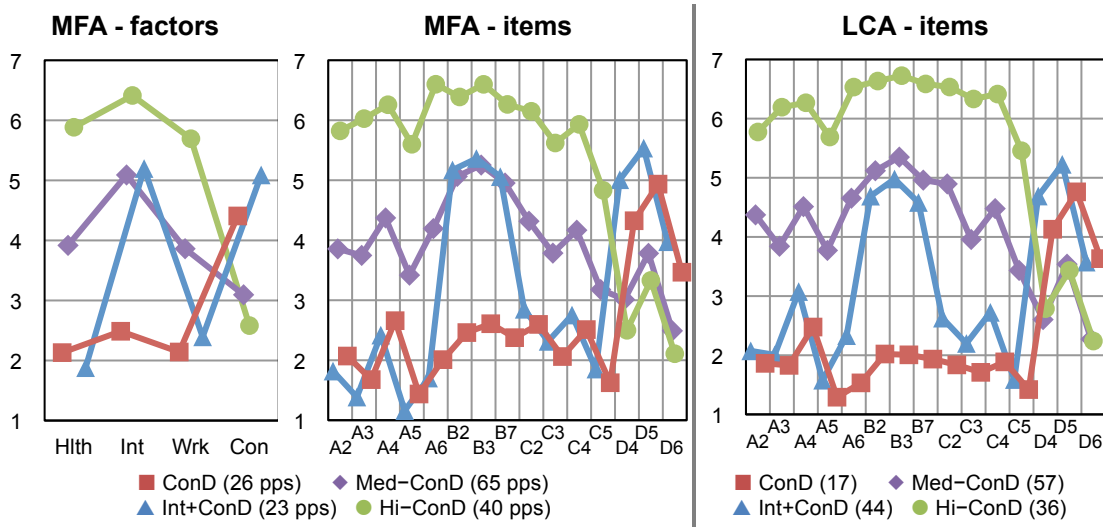

Figure 15: The factor values and item disclosure tendencies for the five-class MFA solution (left), and the item disclosure tendencies for the five-class LCA solution (right).

7.3.2 *Step 4: Latent Class Analysis (LCA)*

The right side of Figure 15 shows the LCA results, which clearly match the MFA results, the only difference being slightly different class sizes. This indicates that the factors are an adequately simplified representation of participants' behavior.

7.4 Attitude−>Behavior link

In the online retailer study, we used 19 seven-point scale items to measure three privacy attitudes (based on Malhotra et al. (2004) and Smith et al. (1996)). Table 16 displays the results of the CFA of these three factors. Note that general privacy concerns and collection concerns are very highly correlated (r = .760).

Considered aspects	Items	Factor loading
General privacy concerns Alpha: 0.80 AVE: 0.560	I only deal with online companies if I am certain that they will respect my privacy	
	I do not let concerns about privacy get in the way of interactions with online companies	
	Compared to others, I am more sensitive about the way online companies handle my personal information	0.708
	To me, it is the most important thing to keep my privacy intact from online companies	0.626
	In dealing with companies online, I try to keep things private as much as possible	0.787
	Online privacy is an overblown problem	
	I am concerned about threats to my personal privacy today	0.851
Collection concerns Alpha: 0.86 AVE: 0.670	It usually bothers me when online companies ask me for personal information	0.852
	It bothers me to give personal information to so many online companies	0.939
	Online companies may collect any information about me because I have nothing to hide	-0.727
	I am concerned that online companies are collecting too much personal information about me	0.820
	I am not bothered by data collection, because my personal information is publicly available anyway	-0.747
Control concerns Alpha: 0.65 AVE: 0.526	Online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used, and shared	0.682
	Control of personal information lies at the heart of online privacy	0.805
	Having the option to keep certain information for myself is a sufficient way to protect my privacy	
	The more options I have to share or not share my information, the better it is for my privacy	0.682
	Consumers should have the right to control what online companies do with their personal information	
	I have a strong need to control what companies do with my personal information	
	I am not worried about my privacy as long as I can control what happens with my personal information	

Table 16: The items used in the app recommender study to measure attitudes and personal characteristics, along with the factor loadings of the CFA (items that do not have a factor loading in the rightmost column were excluded from the analysis).

7.4.1 *Step 5: Predicting behavioral dimensions with attitudes/characteristics*

Table 17 shows the results of the regression of the four behavioral factors on the control concerns and collection concerns; general privacy concerns did not show any significant effect, likely due to the high correlation with collection concerns. Collection concerns decrease the disclosure intentions for all types of information, especially contact information. Interestingly, control concerns *increase* disclosure of interests and work related items. This is in line with Nowak and Phelps (1995) and Taylor et al. (2009), who argue that when people perceive to be in control of their information disclosure, this actually reduces the significance of privacy threats and may thus increase their disclosure. Our study gave participants "double control": for each piece of solicited

information, they could disclose or not disclose it to the experimenters, and separately indicate their willingness or refusal to disclose it to the online retailer.

	Collection concerns	Control concerns
Health	$\beta = -.161$ (.049), $p = .031$	*ns*
Interests	$\beta = -.205$ (.100), $p < .040$	$\beta = .229$ (.106), $p = .030$
Work	$\beta = -.274$ (.077), $p < .001$	$\beta = .226$ (.114), $p = .048$
Contact	$\beta = -.445$ (.106), $p < .001$	*ns*

Table 17: Regression coefficients (standard errors) regressing the behavioral factors on the attitudinal factors (attitude −> behavior).

7.4.2 *Step 6: differences in attitudes/characteristics between classes*

Although attitudes were able to predict behavioral dimensions (Step 5), we did not find any differences in attitudes between classes. We also found no differences in gender, but we did find some age differences, as shown in Figure 16. The groups that have low contact info disclosure tendencies (Hi−ConD and Med−ConD) are generally younger than the other two groups.

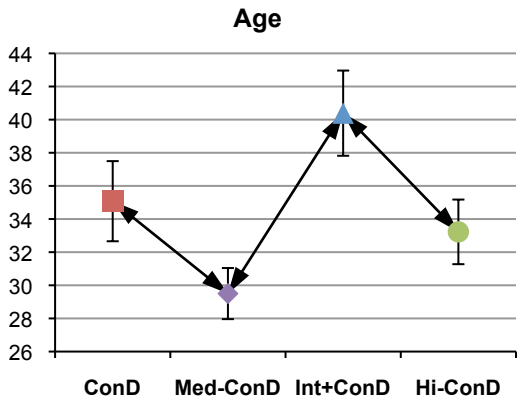

Figure 16: Differences between classes in terms of age. Arrows indicate significant differences. Error bars are ±1 standard error of each measurement.

7.5 Behaviors versus behavioral intentions

In the online retailer experiment, we did not only measure behavioral intentions but also actual disclosure behaviors. Although the measured behavioral intentions have a different target than the actual behaviors (an online retailer versus the experimenters), the two measurements can be used to explore the link between intentions and actual behaviors, and to test whether the dimensionality and classification results are the same for intentions and behaviors.

In this dataset, intention is a strong predictor of behavior ($\beta = 0.330$, $SE = .022$, $p < .001$). Participants answering "very likely" to the behavioral intention question of a certain item were on average 2.69 times more likely to disclose the item than participants answering "neutral". Similarly, participants answering "very unlikely" were 2.69 times less likely to disclose the item than those answering "neutral".

We conducted our dimensionality and classification procedure on the behavioral data as well, and compared the results with those for the intentions data (presented in this section). For the behavioral data, step 1 and step 2 resulted in the same four dimensions as the optimal solution for the intentions data. The correlation of contact information with the other types of information is

not negative like in the intentions dataset, but is instead non-significant (with health and interests) or low (with work).

The classification of participants was rather different: in the behavioral data we found a small class of participants with very high disclosure (HiD; 14 participants), a small class of participants with low disclosure (LowD; 17 participants), and a very large class of participants that discloses everything except contact information (Hi−ConD; 123 participants). A likely reason why this classification deviates from our results on intentions is that these disclosures are directed towards the university conducting the survey, and not the online retailer.

7.6 Discussion of the results from dataset 3

The online retailer data has four dimensions, and four different groups of users seem to behave inherently differently along these dimensions. The most interesting finding in this dataset are the distinctions that participants apparently make between contact info and all other data: some participants are okay with providing all kinds of information to retailers, as long as they cannot be contacted. Others see the merit of contact information (which may for instance be necessary for making purchases), but then minimize other types of disclosure. Reflecting on this finding, we start to realize how consumers can get "tricked" into disclosing their information to an online retailer: first they submit a lot of information under the assumption that they cannot be contacted about it, but when they finally want to purchase something, they cannot avoid submitting their contact information (Kobsa & Teltzrow (2005) present findings how the purchase ratio in such situations can be influenced by trust-enhancing explanations at the interface).

Another interesting aspect is that privacy attitudes do not predict whether someone belongs to the group that discloses contact info only, or to the group that discloses everything but contact info. Age is a much better predictor, but for practical applications additional predictors need to be found. Finally, intentions and actual behaviors have the same dimensionality in this dataset. The classification results are different, but one of the most interesting results—the strong distinction between contact information and other types of information—is upheld in this dataset.

8 General discussion

Using three datasets of online information disclosure intentions and behaviors, we demonstrate that information disclosure behaviors are not unidimensional but instead consist of multiple related dimensions. Furthermore, we show that people can be classified into distinct groups that show very different behaviors along these dimensions. Importantly, these distinct groups do not necessarily differ in their overall degree of disclosure (as in the now almost classical trichotomy of privacy fundamentalists, pragmatists and unconcerned (Harris et al., 2003b)), but rather in their disclosure tendencies per *kind* of information.

It is important to make such distinctions, since they may reveal that groups of people with the same amount of overall disclosure can show very different "disclosure profiles" if one looks at more than one dimension. Our datasets contain a number of examples for this. In dataset 1, two groups of participants exhibit the same medium level of disclosure, but one group discloses both context items and demographics items at a medium rate while the other group discloses almost all demographics items but almost no context items. In dataset 2, one group has high intentions to disclose location-related items but low intentions to disclose activity-related items, and another group has the opposite intentions (a third group has high, and a fourth group low, intentions on both). Similarly, in dataset 3 all participants seem to have different disclosure tendencies for

contact information compared with all other types of information, but in different extent and direction.

Distinguishing different types of disclosure behaviors per type of personal information can improve the accuracy of prior research results, in which disclosures were summed up into a single "disclosure score". Our research suggests that this summation approach may fail to uncover important insights or, worse, make invalid claims on the assumption of unidimensionality.

Another important area in which qualitative distinctions in online disclosure behavior should be respected is in personalized privacy. Developers of commercial applications are increasingly aware of the fact that different people require different levels of privacy. Their solution is to tailor their privacy approach to users' needs (Kobsa, 2001; Wang and Kobsa, 2007). User modeling and personalization have gained much popularity recently under the guise of "big data". The general aim of this field is to leverage the knowledge that is being gathered about users to tailor the content and presentation of (online) services to their specific needs and preferences. Personalization has found broad practical application in commercial systems such as recommender systems (e.g. Amazon, Netflix) and intelligent agents (e.g. Apple's Siri). The increasing popularity of personalization stresses the importance of adequate privacy practices, but we suggest that these privacy practices themselves can also benefit from a personalized approach. Some of our earlier work shows the potential value (Knijnenburg and Kobsa, 2013a) and feasibility (Wang and Kobsa, 2013) of such adaptive privacy practices. We have recently begun testing (simple versions of) such systems with real users, and our results are predominantly positive (Knijnenburg and Jin, 2013).

Our current results suggest that the accuracy of such personalized systems will improve when their user models implement a multidimensional view of disclosure behaviors. As an added advantage, the identification of distinct groups of users with different "privacy-behavioral profiles" may turn the user modeling from a multidimensional preference tracking problem (Wang and Kobsa, 2007) into a simpler classification problem.

This classification can happen "on the fly" (by observing behaviors during the interaction), but also based on people's privacy attitudes. Our results show that people's collection concerns are the strongest predictor of disclosure behavior in this regard. Interestingly, control concerns either have no effect (study 1), or actually increase disclosure (study 3) in systems where people have a certain level of control over their information (cf. Nowak and Phelps, 1995; Taylor et al., 2009). Importantly, our results show that although existing measures of privacy attitudes (Malhotra et al., 2004; Smith et al., 1996) can predict the degree of information disclosure, they cannot be used to distinguish between different dimensions of disclosure behavior. Finer-grained privacy attitude scales ought therefore to be developed, which may be more accurate predictors of privacy-related behavior if they are targeted to specific types of information (e.g. location privacy attitudes, contact info privacy attitudes; cf. (van de Garde-Perik et al., 2008)).

Another way to initially classify users' information disclosure behavior is to use demographics or other user characteristics. In dataset 1, participants with high or medium disclosure were more likely to use the mobile Internet than participants with low disclosure or demographics-only disclosure. In dataset 2, participants with low disclosure, or high disclosure except contact information, were on average older than those with overall high disclosure rates. The high disclosure group also contained significantly more males. In dataset 3, there are again age differences between groups. Although classification based on these characteristics is not perfect, they could provide a useful initial prediction of class membership, which can be refined in further interaction. For research purposes, these results also indicate that participant samples should be

balanced in terms of age, gender and mobile Internet usage to achieve an adequate representation of the different disclosure behaviors.

Although people seem to be more private in their intentions than in their actual behaviors, our results seem to hold true for both behavioral intentions and actual behavior. Moreover, dataset 3 represents at least one case in which the dimensionality for intentions and for actual behavior is the same.

Our paper does not answer the important question of whether the uncovered dimensions generalize to different settings. In a recent study that used roughly the same items as dataset 3, we found that people disclose different types of information to different websites in different extent, depending on the *purpose* of the website (Knijnenburg and Kobsa, 2013c). For example, participants were more likely to disclose work items to a job search website and health items to a health insurance provider. These findings are in line with results by Khalil and Connolly (2006) and Olson et al. (2005). Importantly, even though the *levels* of disclosure differed per dimension and per website, the *structure* of the dimensions was stable between websites and similar to the dimensions uncovered in our dataset 3. This lends support to the claim that dimensions may be generalizable across different contexts. Nonetheless, not all dimensions may be generalizable to all contexts. For example, "weight" is part of the "health" factor in study 3, but it could be part of a "physical appearance" factor on an online dating site.

Due to the diverse nature of the presented studies and their limited number, the current paper cannot provide evidence for the generalizability of the discovered dimensions. This means that for now, researchers and practitioners will need to conduct their own studies to determine the dimensionality of the privacy behavior of their users. We encourage them to publish the results of their dimensionality studies since this will foster the goal of finding generalizable dimensions.

9 Conclusion

Research has shown increasing interest in comprehensive models of privacy, relating a multitude of antecedents via attitudes to behaviors (Knijnenburg and Kobsa, 2013a; Li, 2011; Smith et al., 2011; Xu et al., 2008). Such research is indispensible for a fundamental understanding of people's privacy preferences. Measurement of privacy attitudes has become increasingly sophisticated, comprising several related but conceptually distinct aspects. The goal of this paper is to introduce similar sophistication to behavioral measurement, and to argue that this sophistication is required because distinct groups of people behave very differently when it comes to information disclosure. Tailoring solutions to these distinct groups may be key to a more user-centric approach to privacy.

Acknowledgements

The research reported here has been supported by NSF grant CNS-0831526. We would like to thank Yang Wang for helping with parts of the data.

Vitae

Bart Knijnenburg is a Ph.D. candidate in Informatics at the University of California, Irvine. His work focuses on privacy decision-making and adaptive systems. He received his B.S. degree in Innovation Sciences and his M.S. degree in Human-Technology Interaction from Eindhoven University of Technology, The Netherlands, and his M.A. degree in Human-Computer Interaction from Carnegie Mellon University. The work described in the current paper was conducted while employed as an intern at Samsung Information Systems of America.

Alfred Kobsa is a Professor in the Donald Bren School of Information and Computer Sciences of the University of California, Irvine. His research lies in the areas of user modeling and personalized systems, privacy, and information visualization. He is the editor of User Modeling and User-Adapted Interaction: The Journal of Personalization Research, and editorial board member of three scientific journals. He edited several books and authored numerous publications in the areas of user-adaptive systems, privacy, human-computer interaction and knowledge representation.

Hongxia Jin is a principal engineer in the Advanced Technologies Lab at Samsung Information Systems of America, focusing on privacy and security research. She obtained her B.S degree in Computer Science from University of Science and Technology of China, and her Ph.D degree in Computer Science from the Johns Hopkins University. She has been a research staff member at IBM, where her primary research interest was on information privacy and security, and privacy-aware recommender systems. The key management and forensic technologies she developed have been adopted by several content protection industry standards.

References

Ackerman, M.S., Cranor, L.F., Reagle, J., 1999. Privacy in e-commerce: examining user scenarios and privacy preferences, in: Proceedings of the 1st ACM Conference on Electronic Commerce, EC '99. ACM Press, Denver, CO, pp. 1–8.

Acquisti, A., 2004. Privacy in Electronic Commerce and the Economics of Immediate Gratification, in: EC'04 ACM Conference on Electronic Commerce. New York, NY, pp. 21–29.

Acquisti, A., Grossklags, J., 2005. Privacy and Rationality in Individual Decision Making. IEEE Security & Privacy 3, 26–33.

Acquisti, A., Grossklags, J., 2008. What Can Behavioral Economics Teach Us About Privacy?, in: Acquisti, A., De Capitani di Vimercati, S., Gritzalis, S., Lambrinoudakis, C. (Eds.), Digital Privacy: Theory, Technologies, and Practices. Taylor & Francis, pp. 363–377.

Acquisti, A., John, L.K., Loewenstein, G., 2011. The Impact of Relative Standards on the Propensity to Disclose. Journal of Marketing Research 1–15.

Ajzen, I., 1991. The theory of planned behavior. Organizational Behavior and Human Decision Processes 50, 179–211.

Ajzen, I., Fishbein, M., 1977. Attitude-behavior relations: A theoretical analysis and review of empirical research. Psychological Bulletin 84, 888–918.

Bélanger, F., Crossler, R.E., 2011. Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. MIS Quarterly 35, 1017–1045.

Bennett, C.J., 1995. The political economy of privacy: a review of the literature.

Bentler, P.M., Bonett, D.G., 1980. Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. Psychological Bulletin 88, 588–606.

Buchanan, T., Paine, C., Joinson, A.N., Reips, U.-D., 2007. Development of Measures of Online Privacy Concern and Protection for Use on the Internet. Journal of the American Society for Information Sciences and Technology 58, 157–165.

Caudill, E.M., Murphy, P.E., 2000. Consumer Online Privacy: Legal and Ethical Issues. Journal of Public Policy & Marketing 19, 7–19.

De Souza, Z., Dick, G.N., 2009. Disclosure of information by children in social networking—Not just a case of "you show me yours and I'll show you mine". International Journal of Information Management 29, 255–261.

Dinev, T., Hart, P., 2004. Internet Privacy Concerns and Their Antecedents: Measurement Validity and a Regression Model. Behaviour & Information Technology 23, 413–422.

Gardner, P.L., 1996. The dimensionality of attitude scales: a widely misunderstood idea. International Journal of Science Education 18, 913–919.

Gruys, M.L., Sackett, P.R., 2003. Investigating the Dimensionality of Counterproductive Work Behavior. International Journal of Selection and Assessment 11, 30–42.

Harris, 2000. A Survey of Consumer Privacy Attitudes and Behaviors, Privacy and American Business Newsletter. Harris Interactive, Inc.

Harris, L., Westin, A.F., associates, 2003a. Most People Are "Privacy Pragmatists" Who, While Concerned about Privacy, Will Sometimes Trade It Off for Other Benefits. Equifax Inc.

Harris, L., Westin, A.F., associates, 2003b. Consumer Privacy Attitudes: A Major Shift Since 2000 and Why (No. 10), Privacy and American Business Newsletter. Harris Interactive, Inc.

Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 6, 1–55.

Iachello, G., Hong, J., 2007. End-User Privacy in Human-Computer Interaction. Foundations and Trends in Human-Computer Interaction 1, 1–137.

John, L.K., Acquisti, A., Loewenstein, G., 2011. Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information. The Journal of Consumer Research 37, 858–873.

Joinson, A.N., Paine, C., Buchanan, T., Reips, U.-D., 2008. Measuring self-disclosure online: Blurring and non-response to sensitive items in web-based surveys. Computers in Human Behavior 24, 2158–2171.

Joinson, A.N., Paine, C.B., 2007. Self-disclosure, privacy, and the Internet, in: Joinson, A.N. (Ed.), The Oxford Handbook of Internet Psychology. Oxford University Press.

Joinson, A.N., Reips, U.-D., Buchanan, T., Schofield, C.B.P., 2010. Privacy, Trust, and Self-Disclosure Online. Human–Computer Interaction 25, 1.

Khalil, A., Connelly, K., 2006. Context-aware telephony: privacy preferences and sharing patterns, in: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work. ACM, Banff, Alberta, Canada, pp. 469–478.

Kiel, G.C., Layton, R.A., 1981. Dimensions of Consumer Information Seeking Behavior. Journal of Marketing Research 18, 233–239.

Knapp, H., Kirk, S.A., 2003. Using pencil and paper, Internet and touch-tone phones for self-administered surveys: does methodology matter? Computers in Human Behavior 19, 117–134.

Knijnenburg, B.P., Jin, H., 2013. The Persuasive Effect of Privacy Recommendations, in: Submitted to SOUPS 2013.

Knijnenburg, B.P., Kobsa, A., 2013a. Helping users with information disclosure decisions: potential for adaptation, in: Proceedings of the 2013 ACM International Conference on Intelligent User Interfaces. ACM Press, Santa Monica, CA, pp. 407–416.

Knijnenburg, B.P., Kobsa, A., 2013b. Making Decisions about Privacy: Information Disclosure in Context-Aware Recommender Systems (No. UCI-ISR-12-1). Institute for Software Research, University of California.

Knijnenburg, B.P., Kobsa, A., 2013c. Counteracting the negative effect of form auto-completion on the privacy calculus, in: Submitted to ICIS 2013.

Kobsa, A., 2001. Tailoring Privacy to Users' Needs (Invited Keynote), in: Bauer, M., Gmytrasiewicz, P.J., Vassileva, J. (Eds.), Proc. User Modeling 2001. Springer Verlag, pp. 303–313.

Kobsa, A., 2007. Privacy-Enhanced Personalization. Communications of the ACM 50, 24–33.

Kobsa, A., Teltzrow, M., 2005. Contextualized Communication of Privacy Practices and Personalization Benefits: Impacts on Users' Data Sharing Behavior, in: Martin, D., Serjantov, A. (Eds.), Privacy Enhancing Technologies: Fourth International Workshop, PET 2004, Toronto, Canada. Springer Verlag, Heidelberg, Germany, pp. 329–343.

Koshimizu, T., Toriyama, T., Babaguchi, N., 2006. Factors on the sense of privacy in video surveillance, in: Proceedings of the 3rd ACM Workshop on Continuous Archival and Retrival of Personal Experences, CARPE '06. ACM, New York, NY, USA, pp. 35–44.

Li, Y., 2011. Empirical Studies on Online Information Privacy Concerns: Literature Review and an Integrative Framework. Communications of the Association for Information Systems 28.

Lusoli, W., Bacigalupo, M., Lupiáñez-Villanueva, F., Andrade, N., Monteleone, S., Maghiros, I., 2012. Pan-European Survey of Practices, Attitudes and Policy Preferences as Regards Personal Identity Data Management (SSRN Scholarly Paper No. ID 2086579). Social Science Research Network, Rochester, NY.

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.

Metzger, M., 2007. Communication Privacy Management in Electronic Commerce. Journal of Computer-Mediated Communication 12.

Metzger, M.J., 2004. Privacy, Trust, and Disclosure: Exploring Barriers to Electronic Commerce. Journal of Computer-Mediated Communication 9.

Metzger, M.J., 2006. Effects of Site, Vendor, and Consumer Characteristics on Web Site Trust and Disclosure. Communication Research 33, 155–179.

Muthén, B., 2007. Latent variable hybrids: Overview of old and new models, in: Hancock, G.R., Samuelsen, K.M. (Eds.), Advances in Latent Variable Mixture Models. Information Age Publishing, Inc.

Norberg, P.A., Horne, D.R., Horne, D.A., 2007. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. Journal of Consumer Affairs 41, 100–126.

Nowak, G.J., Phelps, J., 1995. Direct marketing and the use of individual-level consumer information: determining how and when "privacy" matters. Journal of Direct Marketing 9, 46–60.

Olson, J.S., Grudin, J., Horvitz, E., 2005. A study of preferences for sharing and privacy, in: CHI '05 Extended Abstracts on Human Factors in Computing Systems. ACM, Portland, OR, pp. 1985–1988.

Patil, S., Kobsa, A., 2009. Privacy Considerations in Awareness Systems: Designing with Privacy in Mind, in: Markopoulos, P., Ruyter, B. de, Mackay, W. (Eds.), Awareness Systems: Advances in Theory, Methodology and Design. Springer Verlag, Berlin, Heidelberg, New York, pp. 187–206.

Phelps, J., Nowak, G., Ferrell, E., 2000. Privacy Concerns and Consumer Willingness to Provide Personal Information. Journal of Public Policy & Marketing 19, 27–41.

Phelps, J.E., D'Souza, G., Nowak, G.J., 2001. Antecedents and consequences of consumer privacy concerns; an empirical investigation. Journal of Interactive Marketing (John Wiley & Sons) 15, 2–17.

Smith, H.J., Dinev, T., Xu, H., 2011. Information Privacy Research: An Interdisciplinary Review. MIS Quarterly 35, 989–1015.

Smith, H.J., Milberg, S.J., Burke, S.J., 1996. Information Privacy: Measuring Individuals' Concerns about Organizational Practices. MIS Quarterly 20, 167–196.

Spiekermann, S., Grossklags, J., Berendt, B., 2001. E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus Actual Behavior, in: Proceedings of the 3rd ACM Conference on Electronic Commerce. Tampa, FL, pp. 38–47.

Stewart, K.A., Segars, A.H., 2002. An Empirical Examination of the Concern for Information Privacy Instrument. Information Systems Research 13, 36–49.

Taylor, D., Davis, D., Jillapalli, R., 2009. Privacy concern and online personalization: The moderating effects of information control and compensation. Electronic Commerce Research 9, 203–223.

Van de Garde-Perik, E., Markopoulos, P., de Ruyter, B., Eggen, B., Ijsselsteijn, W., 2008. Investigating Privacy Attitudes and Behavior in Relation to Personalization. Social Science Computer Review 26, 20 –43.

Wang, Y., Kobsa, A., 2007. Respecting Users' Individual Privacy Constraints in Web Personalization, in: Conati, C., McCoy, K., Paliouras, G. (Eds.), 11th International Conference on User Modeling. Berlin - Heidelberg - New York: Springer-Verlag, Corfu, Greece, pp. 157–166.

Wang, Y., Kobsa, A., 2013. A PLA-based privacy-enhancing user modeling framework and its evaluation. User Modeling and User-Adapted Interaction: The Journal of Personalization Research 1.

Wang, Y., Norice, G., Cranor, L., 2011. Who Is Concerned about What? A Study of American, Chinese and Indian Users' Privacy Concerns on Social Network Sites, in: McCune, J., Balacheff, B., Perrig, A., Sadeghi, A.-R., Sasse, A., Beres, Y. (Eds.), Trust and Trustworthy Computing, Lecture Notes in Computer Science. Springer, pp. 146–153.

Westin, A.F., Harris, L., associates, 1981. The Dimensions of privacy : a national opinion research survey of attitudes toward privacy. Garland Pub., New York.

Westin, A.F., Maurici, D., 1998. E-Commerce & Privacy: What the Net Users Want. Privacy & American Business, and PricewaterhouseCoopers LLP.

White, T.B., 2004. Consumer Disclosure and Disclosure Avoidance: A Motivational Framework. Journal of Consumer Psychology 14, 41–51.

Xu, H., Dinev, T., Smith, H.J., Hart, P., 2008. Examining the formation of individual's privacy concerns: Toward an integrative view, in: Proceedings of the 29th International Conference on Information Systems. Paris, France, pp. 1981–1996.