On The Dimensionality Of Information Disclosure Behavior in Social Networks

Bart P. Knijnenburg
School of Information and Computer Sciences
University of California, Irvine
Irvine, CA 92697 USA
bart.k@uci.edu

Abstract
In studies of networked social privacy, disclosure is typically seen as either a separate behavior for each informational item, or measured by an aggregate disclosure tendency score. Our recent work shows that people’s disclosure behaviors are instead 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”.

Author Keywords
Information disclosure; privacy behavior; privacy attitude; measurement.

ACM Classification Keywords

General Terms
Experimentation, Measurement, Standardization, Theory.

Copyright is held by the author/owner(s).

Introduction
Research on networked social privacy spans a broad range of disciplines, including human-computer interaction [9], information systems [5], personalization [11], behavioral economics [7], and social psychology [1]. Although each discipline has its own conceptualization of the notion of “privacy” [15], they seem to agree on the distinction between privacy attitudes and behaviors. Several scales have been developed [4,12,16,18] for the measurement of attitudes, although these scales primarily focus on information privacy rather than interpersonal privacy [14]. Comparatively little research, however, covers privacy-related behavior and specifically information disclosure (or sharing) behavior.

Related work
Research on disclosure behavior can be divided into two approaches. The first approach regards the disclosure of each item of personal information (e.g. location, gender) as a separate decision (e.g. [3,6,7,19]). In the absence of a theory of how different behaviors are related, this work does not define an overall measure of a person’s rate of disclosure. This work typically also does not try to explain how disclosure behaviors come about, or how they can be influenced.

The other approach sums individual disclosures into an overall measure of disclosure tendency (e.g. [13,17]),
thereby make an implicit assumption of unidimensionality (i.e., they assume that all items belong to the same 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.

**Our findings**

Using three datasets of online information disclosure intentions and behaviors, our recent work [10] demonstrates that information disclosure behaviors are in fact **multidimensional**, i.e., that disclosure tendencies may differ for different types of information. 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 only differ in degree of disclosure (as in the trichotomy of privacy fundamentalists, pragmatists and unconcerned [8]), but they also differ in the **kind** of information they tend to disclose or to withhold.

Among our three datasets, one dataset is in the realm of networked social privacy. This dataset, originally reported by [20], considers 359 Facebook users' intentions to publicly disclose 16 profile items. We present a summary of our results for this dataset (cf. [10] for details of all three analyses).

Using Exploratory and Confirmatory Factor Analysis, we uncovered a model (Table 1) of four dimensions underlying 14 of these 16 items: "Facebook activity", "location", "contact info", and "life and interests". We then conducted a Mixed Factor Analysis, which classified study participants into five classes with respect to their factor scores (Figure 1):

- a class of 159 participants with rather low disclosure tendencies on all dimensions (LowD);
- a class of 59 participants who are very likely to disclose any type of information (HiD);
- a class of 65 participants with a low intention to disclose contact information ("Hi−ConD", which stands for "high-minus-contact");
- a class of 50 participants who have a low intention to disclose contact info and activity, but a high intention to disclose location and interests ("Loc+IntD");
- a class of 26 participants with a low intention to disclose contact info and location, but a high intention to disclose activity and interests ("Act+IntD").

We cross-validated this solution with a Latent Class Analysis to confirm that these classes emerge even without considering the factors. Finally, we tested for differences between the five classes in terms of Trust in Facebook, Need for consent when Facebook changes its policy, age, and gender (Figures 2 and 3).

**Multidimensionality: why bother?**

Distinguishing qualitatively different types of disclosure behaviors 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. A multidimensional approach to privacy measurement may reveal that groups of people with the same amount of overall disclosure in fact show very different "disclosure profiles". For example, in the presented Facebook dataset we found one group having high intentions to disclose location but low intentions to disclose activity, and another group with opposite intentions (as well as a group with high
Another important area in which qualitative distinctions in online disclosure behavior should be respected is in personalized privacy. Developers of commercial applications, who are increasingly aware that different people require different levels of privacy, propose to tailor their privacy approach to users’ needs [2,11]. A multidimensional measurement of information disclosure behaviors could greatly improve the accuracy of this personalization. For example, Facebook could make use of our finding that its users fall into five groups with fundamentally different information disclosure behaviors along four dimensions. If the system determines that user X, e.g., belongs to the Act+IntD group, it can deduce that the user does not want to disclose location information but is okay with disclosing her opinions and activities. The system can subsequently refrain from “geo-tagging” her status updates, but publicly display her political preference on her profile page. If user Y, e.g., belongs to the Loc+IntD group, the system knows that this user is okay with the disclosure of location information and opinions, but not of activities. The system can then protect all her posts by default, but always geo-tag them.

**How to classify users**

The classification of users can happen “on the fly” (by observing behaviors during the interaction), but also based on people’s privacy attitudes. Note however that while attitudes can predict the degree of information disclosure, they typically cannot 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).

Another way to initially classify users’ information disclosure behavior is to use demographics or other user characteristics. In our Facebook dataset, participants with low disclosure, or high disclosure except contact information, were on average older than those with high disclosure rates. The high disclosure group also contained significantly more males. 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 age and gender to achieve an adequate representation of the different disclosure behaviors.

**Conclusion**

In this paper we introduced a multidimensional measure of information disclosure behaviors, and argued that such measurement is required because distinct groups of people behave very differently when it comes to information disclosure. Moreover, tailoring solutions to these distinct groups may be key to a more user-centric approach to privacy.

**Acknowledgements**

Special thanks to Yang Wang for his help with some of the data, and Samsung US R&D Center for financial support of this research.

**References**

1. Altman, I., Vinsel, A., and Brown, B.B. Dialectic Conceptions In Social Psychology: An Application To


---

**Figure 3:** 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.