Information Disclosure Profiles for Segmentation and Recommendation

Bart P. Knijnenburg
Donald Bren School of Information and Computer Sciences
University of California, Irvine
bart.k@uci.edu

ABSTRACT

People's information disclosure does not just vary in degree, but also in kind. In fact, our recent work has identified distinct information disclosure profiles in three different datasets. In this paper I briefly iterate the findings of this work, and discuss how these profiles can be used for segmentation and recommendation purposes. I argue that segmentation is an important means to go beyond the "one-size-fits-all" approach of privacy nudges (a privacy aid that is becoming increasingly popular) towards user-tailored privacy decision support.

1. INTRODUCTION

In both social and commercial privacy settings, an important part of users' privacy management activity is deciding what information to disclose. Most existing studies investigating users' information disclosure decisions treat each piece of personal information as an independent decision [2, 9], or as a summated composite score that essentially represents a unidimensional "disclosure propensity" [7, 10, 20–22]. In our recent work [16] we conducted an analysis of three different datasets and found that people's information disclosure differed not only in degree but also in kind. In fact, in each dataset we found several distinct *information disclosure profiles*, and we made some progress towards predicting people's profile. In this paper I briefly iterate our findings, and—as a novel contribution—discuss in more detail how these profiles can be used for segmentation and recommendation purposes.

2. BACKGROUND AND RELATED WORK

Privacy segmentation is not a new idea; early research identified the three broad categories: privacy fundamentalists, pragmatists, and unconcerned [5, 6, 31, 32]. The classifications we developed are different in two ways: First, we classified people on their *behavior* rather than their attitudes. A behavior-based segmentation has more actionable consequences for the recommendation purposes we discuss in this paper. Second, our classifications demonstrated not just a difference in degree, but also a difference in kind: for example, in one of our datasets one group was willing to disclose their location but not their online activities, while another group willing to disclose their online activities but not their location. Other than these two distinctions, our work overcomes some of the shortcomings of earlier work on information disclosure profiling (cf. [18, 19, 23, 24, 26]) by applying rigorous statistical tests to validate the distinct disclosure profiles.

Symposium on Usable Privacy and Security (SOUPS) 2014, July 9-11, 2014, Menlo Park, CA.

3. DISCOVERING PROFILES

Our recent work [16] uncovered distinct information disclosure profiles in three datasets of users' disclosure behaviors and/or intentions, collected in three distinct online settings. We found these profiles by first establishing the *dimensionality* of the disclosure tendencies using Exploratory and Confirmatory Factor Analysis (EFA and CFA) and then *classifying* users into groups that show distinctly different behaviors along these dimensions using Mixture Factor Analysis (MFA) and Latent Class Analysis (LCA). Below we report the classification outcomes of the MFAs; the reader can refer to the original paper [16] for a more complete treatment of our methodology and the results of intermediate steps.

3.1 A mobile app recommender

This dataset is from a study (reported in [14]) with 493 participants (266 female; median age group: 25-30, range: 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). Participants made a decision to disclose or not disclose 12 context items and 19 demographics items. Our analysis confirmed the existence of these 2 dimensions, and 4 user profiles showing distinctly different behaviors along these two dimensions (Figure 1). Specifically, we found classes of users with low, medium and high levels of overall disclosure, but also a class of users who were likely to disclose demographics but not context data.

We also showed that users' privacy concerns and mobile Internet usage behavior (Figure 2) could be used to distinguish between the different classes.

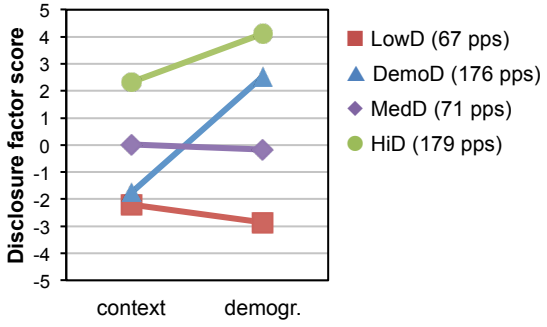

Figure 1: User profiles in the app recommender study (4-class MFA).

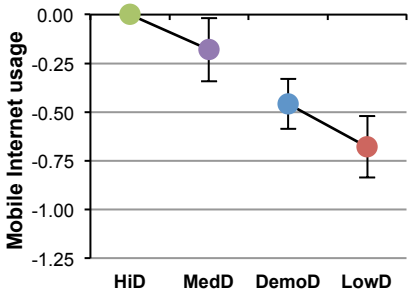

Figure 2: Differences between profiles in mobile Internet usage (standardized). Points that are not connected are significantly different from one another.

3.2 Facebook profile data

This dataset comprises the 359 US participants of Wang et al.'s [30] cross-cultural Facebook study (222 female; median age: 28, range: 18 to 75). Participants indicated on a seven-point scale their level of comfort with disclosing 16 different Facebook profile items to "everyone on the Internet". We found that these items formed 4 dimensions: Facebook activity (Act), Location (Loc), Contact info (Con), and Life and interests (Int). Our classification of participants on these dimensions resulted in 5 behavioral profiles that differed with respect to the types of information participants were willing or unwilling to disclose (Figure 3).

We also showed that users' trust in Facebook, need for consent, age, and gender could be used to distinguish between the different classes. Figure 4 shows these results for trust in Facebook and Gender.

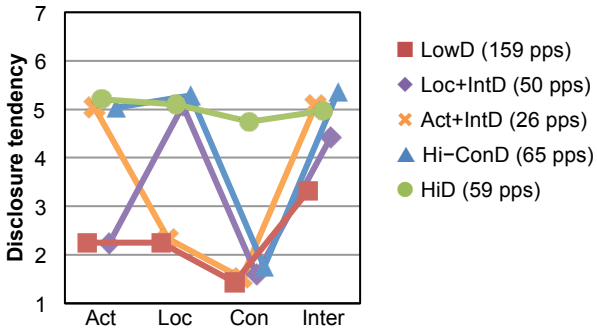

Figure 3: User profiles in the Facebook profile study (5-class MFA).

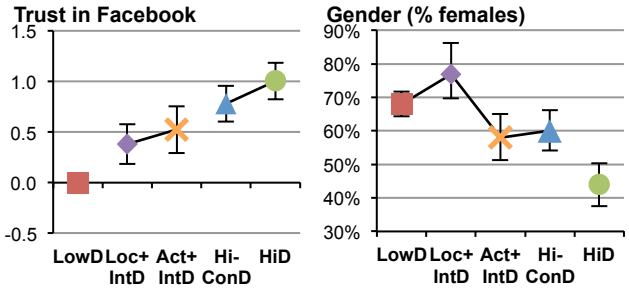

Figure 4: Differences between proflies in trust in Facebook (left, standardized) and gender (right). Points that are not connected are significantly different from one another.

3.3 An online retailer

This dataset was gathered specifically for this study. 154 people (69 females; median age: 29, range: 18 to 65) were first asked to enter the answer to 24 demographical questions into a text field, with the option to rather not disclose it. We then asked participants for each item how likely they were to provide the answer to an online retailer.

We constructed our 24 items so that 6 of them were related to health (Hlth), 6 to interests (Int), 6 to work (Wrk), and 6 to more general information including contact information (Con). These 4 dimensions were confirmed in the dimensionality analysis of the dataset. We also determined 4 behavioral profiles regarding these dimensions (Figure 5), and we found age differences between these profiles (Figure 6).

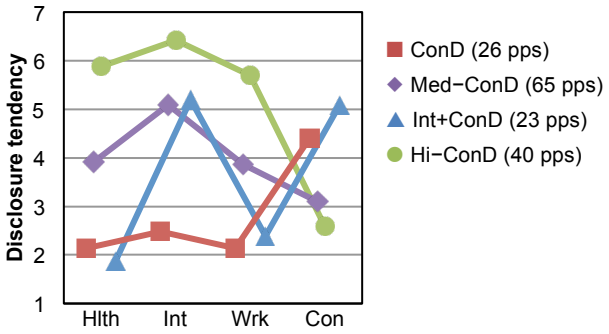

Figure 5: User profiles in the online retailer study (4-class MFA).

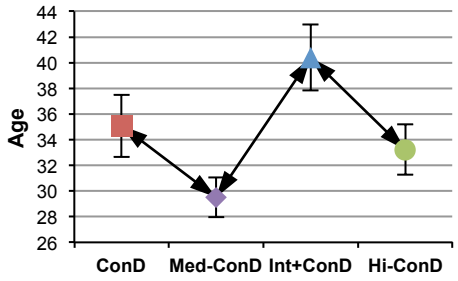

Figure 6: Age differences between profiles. Arrows indicate significant differences.

4. SEGMENTATION

Privacy profiles can be used to segment the users of a service into distinct categories. This segmentation can either happen "on the fly" (by observing behaviors during the interaction), or based on people's privacy attitudes or demographics. For example, in our Facebook dataset, users who trusted Facebook had higher disclosure tendency, and in the app recommender dataset users who scored low on mobile Internet usage were more likely to belong to the low or demographics-only disclosure classes. Although segmentation based on these characteristics is not perfect, they could provide a useful initial prediction of class membership, which can be refined in further interaction.

Segmentation can be used in user research to better understand the effect of new features or privacy policy changes on different types of users. For example, in the Facebook dataset we found a profile with high intentions to disclose location but low intentions to

disclose activity, and another profile with opposite intentions (as well as a profile with high and another with low intentions on both). A change in Facebook's location-tagging feature will thus have an impact on a different user-segment than a change in the status update sharing policy. Similarly, an online retailer who gathers contact information will have to deal with the privacy issues of a different segment than a retailer who (anonymously) gathers health-related information. User researchers thus have to carefully select the segment of users to study when testing the effects of new features or policies.

5. RECOMMENDATION

A recent approach to support privacy decisions is to introduce subtle yet persuasive *nudges* [1, 3, 29]. Carefully designed nudges make it easier for people to make the right choice, without limiting their ability to choose freely [27]. Defaults (such as framing a disclosure decision as either opt-in or opt-in, or changing the order of information requests) are one type of nudge that strongly impact disclosure [2, 8, 14].

The problem with nudges is that they take a one-size-fits-all approach to privacy: They assume that the "true cost" [7] of disclosure is roughly the same for every user, piece of information, and situation. But our results show that decisions are highly user-dependent: what users with a certain profile may consider beneficial to disclose, users with a different profile may see as a privacy threat. In other words, privacy nudges need to be *tailored* to the user's profile [4, 11–13, 17].

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 Loc+IntD group, it knows that this user is okay with the disclosure of location information and opinions, but not of activities. The system can subsequently restrict the audience of her posts by default, but reveal her current city on her profile page. If user Y, e.g., belongs to the Act+IntD group, the system knows that the user does not want to disclose location information but is okay with disclosing her opinions and activities. The system can then refrain from "geo-tagging" her status updates, but reveal her political preference on her profile page.

Similarly, an online retailer could use our finding to provide a series of shopping experiences tailored to the different profiles. For users with an Int+ConD profile, it could provide a limited "based on your interests"-type product recommender. For users with a Hi–ConD and Med–ConD profile, it could provide a more advanced product recommender using all kinds of personal data, but with an anonymous (third-party) checkout feature. Finally, it could retain a more "traditional" online shopping experience for users with a ConD profile.

Although privacy recommendation practices can be implemented without the use of profiles, the identification of profiles turns the user modeling from a multidimensional preference tracking problem [28] into a simpler classification problem. A similar classification can occur for the *recipient* of the information: for example, my recent work shows that people are more likely to disclose information that matches the purpose of the website requesting the information [15]. Segmenting websites by purpose would thus add another adaptation layer to the idea of privacy recommendation.

The ultimate goal of this work is to develop a Privacy Adaptation Procedure to support people's privacy decisions. The procedure predicts the profile of the user, the recipient, and any other contextual variables and then provides automatic "adaptive default" settings in line with these profiles [25]. These smart defaults move beyond traditional privacy nudges in that they reduce the burden of control, but at the same time respect users' inherent privacy preferences. In effect, the Privacy Adaptation Procedure puts users in control of their own privacy decisions without being overwhelming (a problem of the traditional approach) or misleading (a potential problem of nudges). It thus enables users to make privacy-related decisions within the limits of their bounded rationality.

6. CONCLUSION

In this paper I provided an overview of the information disclosure profiles discovered in three datasets, covering a mobile app recommender, a social networking service, and an online retailer scenario. These profiles demonstrate that people's information disclosure does not just vary in degree but also in kind.

I then discussed the implications of these findings for privacy segmentation and recommendation. Segmentation is an important tool for researchers who want to study the effect of their privacy changes or interventions on users: users with different profiles will react to these changes in a different way. Recommendation goes beyond traditional static approaches by helping users with their decision while at the same time respecting their inherent privacy preferences.

7. ACKNOWLEDGEMENTS

Special thanks to Alfred Kobsa and Hongxia Jin, the coauthors of the original paper. Additional thanks to Yang Wang for his help with some of the data, and Samsung Research America for financial support of this research.

8. REFERENCES

[1] Acquisti, A. 2012. Nudging privacy: The behavioral economics of personal information. *Digital Enlightenment Yearbook 2012*. (2012), 193–197.

[2] Acquisti, A., John, L.K. and Loewenstein, G. 2012. The Impact of Relative Standards on the Propensity to Disclose. *Journal of Marketing Research*. 49, 2 (2012), 160–174. DOI= http://dx.doi.org/10.1509/jmr.09.0215.

[3] Balebako, R., Leon, P.G., Mugan, J., Acquisti, A., Cranor, L.F. and Sadeh, N. 2011. Nudging users towards privacy on mobile devices. *CHI 2011 workshop on Persuasion, Influence, Nudge and Coercion Through Mobile Devices* (Vancouver, Canada, 2011), 23–26.

[4] Biehl, J.T., Rieffel, E.G. and Lee, A.J. 2012. When privacy and utility are in harmony: towards better design of presence technologies. *Personal and Ubiquitous Computing*. (Feb. 2012). DOI= http://dx.doi.org/10.1007/s00779-012-0504-7.

[5] Harris 2000. *A Survey of Consumer Privacy Attitudes and Behaviors*. Harris Interactive, Inc.

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

[7] John, L.K., Acquisti, A. and Loewenstein, G. 2011. Strangers on a Plane: Context-Dependent Willingness to Divulge

Sensitive Information. *Journal of consumer research*. 37, 5 (Feb. 2011), 858–873. DOI= http://dx.doi.org/10.1086/656423.

[8] Johnson, E.J., Bellman, S. and Lohse, G.L. 2002. Defaults, Framing and Privacy: Why Opting In ≠ Opting Out. *Marketing Letters*. 13, 1 (2002), 5–15. DOI= http://dx.doi.org/10.1023/A:1015044207315.

[9] Joinson, A.N., Paine, C., Buchanan, T. and Reips, U.-D. 2008. Measuring self-disclosure online: Blurring and non-response to sensitive items in web-based surveys. *Computers in Human Behavior*. 24, 5 (Sep. 2008), 2158–2171. DOI= http://dx.doi.org/16/j.chb.2007.10.005.

[10] Joinson, A.N., Reips, U.-D., Buchanan, T. and Schofield, C.B.P. 2010. Privacy, Trust, and Self-Disclosure Online. *Human–Computer Interaction*. 25, 1 (2010), 1. DOI= http://dx.doi.org/10.1080/07370020903586662.

[11] Knijnenburg, B.P. 2013. Simplifying Privacy Decisions: Towards Interactive and Adaptive Solutions. *Proceedings of the Recsys 2013 Workshop on Human Decision Making in Recommender Systems (Decisions@ RecSys'13)* (Hong Kong, China, 2013), 40–41.

[12] Knijnenburg, B.P. and Jin, H. 2013. The Persuasive Effect of Privacy Recommendations. *Twelfth Annual Workshop on HCI Research in MIS* (Milan, Italy, Jan. 2013), Paper 16.

[13] Knijnenburg, B.P. and Kobsa, A. 2013. Helping users with information disclosure decisions: potential for adaptation. *Proceedings of the 2013 ACM international conference on Intelligent User Interfaces* (Santa Monica, CA, Mar. 2013), 407–416. DOI= http://dx.doi.org/10.1145/2449396.2449448.

[14] Knijnenburg, B.P. and Kobsa, A. 2013. Making Decisions about Privacy: Information Disclosure in Context-Aware Recommender Systems. *ACM Transactions on Interactive Intelligent Systems*. 3, 3 (2013), 20:1–20:23. DOI= http://dx.doi.org/10.1145/2499670.

[15] Knijnenburg, B.P., Kobsa, A. and Jin, H. 2013. Counteracting the Negative Effect of Form Auto-completion on the Privacy Calculus. *ICIS 2013 Proceedings* (Milan, Italy, 2013).

[16] Knijnenburg, B.P., Kobsa, A. and Jin, H. 2013. Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies*. 71, 12 (2013), 1144–1162. DOI= http://dx.doi.org/10.1016/j.ijhcs.2013.06.003.

[17] Kobsa, A. 2001. Tailoring Privacy to Users' Needs (Invited Keynote). *User Modeling 2001*. M. Bauer, P.J. Gmytrasiewicz, and J. Vassileva, eds. Springer Verlag. 303–313.

[18] Koshimizu, T., Toriyama, T. and Babaguchi, N. 2006. Factors on the sense of privacy in video surveillance. *Proceedings of the 3rd ACM workshop on Continuous archival and retrival of personal experences* (New York, NY, USA, 2006), 35–44. DOI= http://dx.doi.org/10.1145/1178657.1178665.

[19] Lusoli, W., Bacigalupo, M., Lupiáñez-Villanueva, F., Andrade, N., Monteleone, S. and Maghiros, I. 2012. *Pan-European Survey of Practices, Attitudes and Policy Preferences as Regards Personal Identity Data Management*. Technical Report #ID 2086579. Social Science Research Network.

[20] Metzger, M.J. 2007. Communication Privacy Management in Electronic Commerce. *Journal of Computer-Mediated Communication*. 12, 2 (2007), 335–361. DOI= http://dx.doi.org/10.1111/j.1083-6101.2007.00328.x.

[21] Metzger, M.J. 2006. Effects of Site, Vendor, and Consumer Characteristics on Web Site Trust and Disclosure. *Communication Research*. 33, 3 (2006), 155–179.

[22] Metzger, M.J. 2004. Privacy, Trust, and Disclosure: Exploring Barriers to Electronic Commerce. *Journal of Computer-Mediated Communication*. 9, 4 (2004). DOI= http://dx.doi.org/10.1111/j.1083-6101.2004.tb00292.x.

[23] Olson, J.S., Grudin, J. and Horvitz, E. 2005. A study of preferences for sharing and privacy. *CHI '05 Extended Abstracts* (Portland, OR, 2005), 1985–1988. DOI= http://dx.doi.org/10.1145/1056808.1057073.

[24] Phelps, J., Nowak, G. and Ferrell, E. 2000. Privacy Concerns and Consumer Willingness to Provide Personal Information. *Journal of Public Policy & Marketing*. 19, 1 (2000), 27–41.

[25] Smith, N.C., Goldstein, D.G. and Johnson, E.J. 2013. Choice Without Awareness: Ethical and Policy Implications of Defaults. *Journal of Public Policy & Marketing*. 32, 2 (Fall 2013), 159–172. DOI= http://dx.doi.org/10.1509/jppm.10.114.

[26] Spiekermann, S., Grossklags, J. and Berendt, B. 2001. E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus Actual Behavior. *Proceedings of the 3rd ACM conference on Electronic Commerce* (Tampa, FL, 2001), 38–47.

[27] Thaler, R.H. and Sunstein, C. 2008. *Nudge : improving decisions about health, wealth, and happiness*. Yale University Press.

[28] Wang, Y. and Kobsa, A. 2007. Respecting Users' Individual Privacy Constraints in Web Personalization. *User Modeling 2007*. C. Conati, K. McCoy, and G. Paliouras, eds. Springer Berlin / Heidelberg. 157–166.

[29] Wang, Y., Leon, P.G., Acquisti, A., Cranor, L.F., Forget, A. and Sadeh, N. 2014. A Field Trial of Privacy Nudges for Facebook. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (Toronto, Canada, 2014), 2367–2376. DOI= http://dx.doi.org/10.1145/2556288.2557413.

[30] Wang, Y., Norice, G. and Cranor, L. 2011. Who Is Concerned about What? A Study of American, Chinese and Indian Users' Privacy Concerns on Social Network Sites. *TRUST* (Pittsburgh, PA, 2011), 146–153.

[31] Westin, A.F., Harris, L. and associates 1981. *The Dimensions of privacy : a national opinion research survey of attitudes toward privacy*. Garland Publishing.

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