

Privacy & Personalization: Preliminary Results of an Empirical Study of Disclosure Behavior¹

Evelien Perik, Panos Markopoulos

Eindhoven University of Technology, Dep. of Industrial Design,
P.O.Box 513, 5600 MB Eindhoven, The Netherlands
{e.m.perik@tue.nl, p.markopoulos}@tue.nl

Abstract. This paper describes empirical research into the privacy preferences and behaviors of individuals regarding the personalization in music recommender systems. These phenomena concern music recommendations based on two different types of user information: music preferences and personality traits. Our results indicate similar disclosure behavior by users for both types of personal information. This contradicts attitudes of users as reported in post-experiment questionnaires and interviews. Factors found to influence disclosure behavior are: information about the purpose of the disclosure and recipients of the information, the degree of confidentiality of the information involved, and the benefits people expect to gain from disclosing personal information.

1 Introduction

Personalized services rely for their operation on appropriate and sufficient information about the user. This could, for example, include information about the identity of the user, usage of the service, preferences and dislikes of the user, or even exact dates and times the service has been used by the user [5]. The acquisition and storage of such information could be regarded as intrusive. Information may be collected by explicit (and conscious) input of the user, or information may be collected implicitly, without the users' explicit intervention. Especially information that is collected implicitly can lead to privacy issues, because of unawareness and lack of control by the people concerned [6]. Little is known about the actual perception of users, their preferences and needs when they are exposed to such situations. This may be due to the fact that privacy is a difficult concept to study. For instance, it seems that the privacy concerns people report in surveys do not match their actual behavior [8].

This paper describes preliminary results of empirical research into the factors influencing the trade-off between the perceived benefits of personalization and the privacy 'costs' experienced by individuals. At the time of writing this paper data collection has finished, but the analysis phase is not yet complete. In our experiment

¹ This work is supported by Philips Research Eindhoven.

participants used a purpose-built music recommendation service over the Internet and were confronted with actual privacy dilemmas. We investigated people's disclosure behavior relating to two types of personal information: preferences for different music genres and information about their personality. In the study users were offered the choice to disclose information about themselves in return for improved recommendations. Their choices and stated opinions regarding this system provide us insight into the relative sensitivity of both types of information.

2 Method

A within-subject design was chosen for the experiment to compare the disclosure choices made by each participant in different situations. We studied the users' reactions in providing two different types of information in exchange for the music recommendation: preferences for music genres and personality traits. Research by Rentfrow and Gosling [7] has found a relation between music preferences and personality traits. This enabled us to recommend music based on the user's personality traits. For each of these information types, three different uses of the information were possible: personal use of the information only, collaborative filtering (i.e. sharing personal information with a system that matches users based on this information), and directly showing information to other users so that these users can recommend music to each other. This resulted in six consecutive situations that participants experienced. At four different points in this study each participant was offered a choice of the level of privacy he/she would choose for their personal information. The choices made by participants were logged by the system. Through post-experiment questionnaires and interviews we tried to understand the reasons behind their choices.

Participants: The participants were recruited by e-mail announcements via secretaries and on bulletin boards within Philips and Eindhoven University of Technology. Announcements were sent to over 1000 employees and students of the two organizations. In total 48 participants, of which 8 female and 40 male participants, completed the study. The ages of participants ranged from 17 to 49 with an average age of 23 years old.

Apparatus and Materials: The participants had access to the music recommender service through a web browser. This service offers personalized playlists of songs. Using streaming technology, these songs were made available for playing on their PC. The experimental recommender service is built on a database of nearly 6000 songs and covers 12 different music genres. The recommender was made especially for the purpose of the experiment to ensure the different privacy preferences were enforced at the appropriate stage of the experiment.

Measures: Several types of data were collected during the study. Participants preferences for music genres were obtained by the Short Test of Music Preferences, STOMP [7]. An inventory of personality traits was made by the Ten Item Personality Inventory, TIPI [4]. A quality rating for each recommended playlist was obtained. The titles and artists of all recommended songs were logged, as well as the time and

date the service was used by participants. And finally, in the on-line questionnaire a combination of open and multiple-choice questions were answered. Among others these questions aimed to establish the privacy attitudes of individuals, their attitudes towards taking risks, and to get explanations for their disclosure behavior.

Procedure: People who were interested in participating were sent an e-mail with instructions. They were not told in advance that the research was about privacy concerns; they were only told in general terms that we were investigating their experience of using personalized music recommender. The order in which participants experienced the two recommender systems was counterbalanced. Half of the participants started using the recommender system based on preferences for music genres, followed by the recommender system based on personality traits. The other half of the participants used the recommender systems in reversed order.

On the first day of using any of the two recommender systems participants were offered no choice regarding the disclosure of their personal information. The disclosure level was set to by default to allow only personal use of their information. During the second and third day participants were offered a choice. Participants could choose between three levels of permission: no permission (no disclosure), restricted permission (disclosure of the information in an anonymous way) or full permission (disclosure of the information in identifiable way). On the second day this concerned using personal information for collaborative filtering, and on the third day it concerned showing it directly to other users. If participants would choose the 'no disclosure' option, then the music recommender would perform as on the previous day. If participants chose to disclose their information the recommendations would improve. In order to deliver benefits to users in a predictable manner the same recommender technique was used in all cases. However, the percentage of songs recommended according to the user profile would change in the various conditions. Participants were required to rate the quality of every playlist. This was done in order to verify that the quality of the playlist has been considered by the participant, and that it has indeed improved as intended.

In the questionnaire participants were asked to comment on their choices during the experiment and to indicate what level of permission they would choose if they could choose again. Some of the participants were contacted for an interview appointment to discuss their choices in more detail. In total 21 interviews were held.

3 Results

3.1 Attitudes towards the disclosure of profile information to the system

Participants were asked to provide two types of profile information: music preferences and personality traits during the experiment. In both cases all participants provided this profile information. In the questionnaire participants were asked to explain how they felt about having had to provide these types of profile information

(see Table 1). Overall, it seems that participants experienced had less reservations to disclose music preferences compared to personality. Some participants questioned the usefulness of having to provide personality traits, this was not the case for music preferences.

Table 1. Amount of participants using specific explanation to express how they felt about providing music preferences or personality traits

Explanation...	Preferences	Personality traits
OK	18	13
Expected / Logical / Useful	12	
No problem	6	8
No problem - if anonymous	2	2
Preferred	1	
Difficult	4	9
Question usefulness		9
Surprised / Unexpected		4
Interesting		2
(No explanation)	5	1
Total	48	48

Table 2. The amount of participants choosing a particular level of disclosure per situation during the experiment

	CompareMPref s			Show MPrefs			Compare Traits			Show Traits		
	No	Anon	Iden	No	Anon	Iden	No	Anon	Iden	No	Anon	Iden
M-T	0	11	13	0	13	11	0	12	12	0	11	13
T-M	1	13	10	0	11	13	0	11	13	0	12	12
Total	1	24	23	0	24	24	0	23	25	0	23	25

3.2 Chosen level of disclosure for the use of profiling information during the experiment

Table 2 displays the amount of participants that chose a particular level of disclosure per situation. The table shows the choices made by participants per experimental sequence (M-T: music preferences-traits; T-M: traits-music preferences), as well as for the total group of participants. The “no disclosure”-option is avoided (it is only chosen once by one of the participants). Furthermore, two main groups can be distinguished based on disclosure choices. One group chose anonymous disclosure in all choice situations, the other group chose disclosure including identity information in all four choice situations. These groups are about equal in size. From the observed disclosure behavior of the participants, it results that most participants stick to their initial choice throughout the 4 choice situations. Out of the 48 participants 41 participants stuck to their initial choice (either anonymous disclosure or disclosure including identity information). Only 7 participants chose different levels of

disclosure across the 4 choice situations. Out of these 7 participants who chose different levels of disclosure in the 4 choice situations, 3 participants indicated that they chose a different level of disclosure compared to the other choice situations, because they wanted to check the effect of a different choice.

3.3 Chosen level of disclosure for the use of profiling information after the experiment

In the questionnaire participants were asked what level of disclosure they would choose in each of the choice situations, if they could choose again (see Table 3). In total 29 participants would not change their opinion about the level of disclosure Table 4 shows that more participants chose the ‘no disclosure’ option after the experiment. Also more people chose a lower level of disclosure for the choice situations concerning personality traits than during the experiment. Seven participants would now choose a lower level of disclosure for the choice situations involving personality traits. Three participants would now choose a lower level of disclosure for the choice situations where information is directly shown to other users. Also 3 participants would now choose a lower level of permission in all four choice situations as compared to the choices made during the experiment. Other participants chose otherwise different levels compared to before.

Table 3. The amount of participants choosing a particular level of disclosure per situation after the experiment

	Compare MPrefs			Show Mprefs			Compare Traits			Show Traits		
	No	Anon	Iden	No	Anon	Iden	No	Anon	Iden	No	Anon	Iden
M-T	1	12	11	1	10	13	4	12	8	5	13	6
T-M	1	14	9	0	13	11	1	12	11	3	12	9
Total	2	26	20	1	23	24	5	24	19	8	25	15

3.4 Comparison of disclosure behavior & privacy attitudes

If the observed disclosure behavior of participants is compared to their answers in the on-line questionnaire and the information provided in the interviews a mismatch is found between thoughts or feelings and actual behavior. Information about personality traits is considered more personal than preferences for music genres. 43 percent of the participants felt it was worrying if other people gets access to their information about personality traits, and 50 percent of the participants worried about this in the case of a music content provider. Yet all of these participants gave permission to the system for the comparison of their profile information to that of others, or for directly showing it to others. Many participants comment on the perceived difference between personality traits and preferences, like this participant: *“I think (...) that those personality traits are more confidential, than (...) the one with those preferences. (...). I think it tells more about yourself.”*

This discrepancy between disclosure behavior and privacy attitudes is in accordance with the experiment of Spiekermann et al [8] and is also described by Acquisti and Grossklags [2]. The experiment by Spiekermann is conducted in the context of e-commerce. The study described in this paper, found similar results in the context of personalized applications. The discrepancy is all the most striking considering that in our experiment self-report was obtained post-hoc and was referring directly to the choices offered rather than to general attitudes and opinions.

3.5 Factors influencing disclosure behavior

In the on-line questionnaire and interviews the reasons behind participants' disclosure behavior was discussed. Participants mentioned for example the influence of the available information, especially the amount and clarity of the information was mentioned. Two participants indicated in the on-line questionnaire not to be sure of the consequences of choosing disclosure including identity, and as a result chose anonymous disclosure instead.

Another factor was the purpose or usage of the information. Participants expressed worries about not knowing how their information will be used by the system. *"I think it is important to know that if I allow someone to just go ahead, what is actually going to happen. Who knows where you will all end up, where you will be associated with. And where they are going to use [your information] for."* Some participants also questioned the relevance of providing a name along with the profile information. This relates to the cluster found by Spiekermann et al [8], whose privacy concerns focused on the revelation of identity aspects such as name, address or e-mail.

Participants expressed worries about who gets access to their personal information: For example, because they don't know what other people may do with the information: *"If other people can link my name to certain personal information en I don't know what they can do with it or want to do with it, then I am careful with it."*

Participants do consider how personal the information is to them before deciding to disclose the information or not. *"Because personality traits are something, (...) I think that is fairly personal. (...) The fact that I think I am extravert, or introvert, that is something completely different than when you tell someone you like 'dance' for example. (...) That is a really different kind of information, that you release"*. Some participants indicate that other people cannot derive much from knowing music preferences, whereas for personality traits people may judge you before they actually get to know you. The three factors mentioned above are in line with the model of Adams and Sasse [3] who identify *Information Receiver*, *Information Usage* and *Information Sensitivity* as three critical factors for shaping privacy behavior.

Some participants also consider what benefits they will gain from disclosing the information. *"My general opinion in that respect is that as long as something does not get extremely personal, they are allowed to know everything about me, as long as I gain from it myself"*. Based on the analysis so far, it seems that different groups of people may be distinguished based on their perception of a certain situation. Some participants seem to focus mainly on potential benefits of disclosing information whereas others seem to focus mainly on the potential risk or cost of disclosing

information. This needs to be further analyzed based on the qualitative data available and compared to the chosen levels of disclosure.

Another influencing factor may have been the experimental context. From the online questionnaire and the interview data it appears that participants felt quite safe disclosing personal information in the context of this experiment, even though they were actually allowing the system to show their personal information to other users.

4 Discussion

We have reported on preliminary results of a study into the privacy preferences and behaviors relating to personalized music recommender systems. All participants disclosed both music preference and personality information to the system and most felt relaxed about it. In the case of personality traits at least some participants were questioning the quality of the recommendations in advance.

Participants chose equal levels of disclosure across the 4 choice situations during the experiment: Possibly, the difference between the privacy sensitivity of the 4 different choice situations is negligible. However, the interview and questionnaire data do not support this interpretation. Further analysis of this data is needed. Another explanation may be that the participant did notice the difference between the choice situations, but were lead by their tendency to cooperate with the research as much as possible (resulting in disclosure for the use of their information throughout all choice situations). It may be, that the mere thought of participating in a research changes one's ideas about privacy. By the choosing this specific experimental set up, in which participants actually used a music recommender service, and were even asked to show their information directly to other people, we tried to prevent this from happening. If actual, this phenomenon should cast some doubt to experimental studies on the topic of privacy creating the need for triangulation with field survey data regarding actual disclosure behavior. After the experiment participants were asked to indicate the desired level of disclosure in each of the 4 choice situations again. This time, more people choose a lower level of disclosure in the choice situations concerning personality traits after the experiment than during. Indicating towards a higher sensitivity of personality traits compared to music preferences.

The discrepancy between disclosure behavior and privacy attitudes is in accordance with other sources [8], [2]. It proves again that for a thorough understanding of people's attitudes towards privacy, extensive field studies are required (based on actual behavior). For the design of personalized service, it is very important to take this discrepancy between privacy concerns and disclosure behavior into account. If people do use a personalized service, this does not necessarily imply that they feel comfortable using it.

In the analysis so far, it seems that different groups of people may be distinguished based on their perception of a certain situation, either focusing on the potential benefits of disclosing information or focusing on the potential risk or cost of disclosing information. This will be further analyzed based on the qualitative data available. It will be interesting to see how this segmentation relates to that found by Ackerman et al [1] and its refinement by Spiekermann et al [8].

5 Conclusion

This paper only described preliminary results. Further analysis will be performed, especially on the qualitative data that has been collected. The questionnaire data shows that participants had fewer problems providing music preferences compared to personality traits. Based on questionnaire and interview data it the impression arises that personality traits are considered more sensitive information.

Basically two groups of people can be distinguished based on participants' disclosure behavior, namely one group choosing disclosure in an anonymous way, and the other choosing disclosure including identity information. Further investigation is needed to see in what ways these groups can be identified. Throughout the study participants persisted with their initial choice regarding the level of disclosure. This indicates that especially the initial level of trust users feel towards these systems is crucial to the success of such systems.

The found discrepancy between disclosure behavior and attitudes of users is in accordance with studies in other domains such as e-commerce. With this study it is shown to hold for the domain of personalization as well. In the questionnaire and interviews participants mentioned various factors influencing their disclosure behavior, such as purpose or usage of the information, the recipients of the information, the sensitivity of the information involved and the expected benefits in return for disclosure.

References

1. Ackerman, M., Cranor, L.F., Reagle, J.: Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. In: Proceedings of the ACM Conference on Electronic Commerce (EC'99), 3-5 November 1999, Denver, Colorado, 1-8.
2. Acquisti, A., Grossklags, J. Losses, gains, and hyperbolic discounting: An experimental approach to information security attitudes and behaviors. In 2nd Annual Workshop of on Economics and Information Security, 2003.
3. Adams, A., Sasse, M. A.: Privacy in multimedia communications: protecting users not just data. Proceedings of IMH HCI'01. (2001) 49-64.
4. Gosling, S. D., Rentfrow, P. J., Swann, W. B., Jr.: A very brief measure of the Big Five personality domains. *Journal of Research in Personality*. 37 (2003) 504-528.
5. Kobsa, A.: Pseudonymous yet Personalized Interaction with Websites that Utilize Network-wide User Modeling Services. In: 2003 HCIC Winter Workshop, Winter Park, CO, 2003.
6. Kobsa, A., Schreck, J.: Privacy through Pseudonymity in User-Adaptive Systems. In: *ACM Transactions on Internet Technology*, vol. 3 (2), 2003, pp. 149-183.
7. Rentfrow, P. J., Gosling, S. D.: The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*. 84 (2003) 1236-1256.
8. Spiekermann, S., Grossklags, J., Berendt, B.: E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior. Proceedings of the 3rd ACM conference on Electronic Commerce. ACM Press. (2001) 38-47.