# Improving the prediction of users' disclosure behavior… by making them disclose more predictably?

Hongchen Wu
School of Computer Science and Technology
Shandong University, China
hongchenwu88@gmail.com

Bart P. Knijnenburg
Department of Informatics
University of California, Irvine
bart.k@uci.edu

Alfred Kobsa
Department of Informatics
University of California, Irvine
kobsa@uci.edu

## ABSTRACT

Taking a step beyond segmentation, privacy researchers have recently proposed privacy personalization or adaptation as an approach to assist users in their privacy decision making. Analyzing a number of datasets of users' personal information disclosure behavior, we find an interesting phenomenon regarding privacy personalization: the order in which information is requested has an impact on prediction accuracy. We provide evidence that this happens because certain request orders cause people's disclosure behavior to be less variable and thus more predictable. This is an important phenomenon to study, because if request orders indeed influence the variability and predictability of subsequent requests, then adapting the request order to the user may result in positive feedback loops that promote prediction accuracy. We address several possible explanations for this phenomenon, and we propose a study that will help us find out which of these explanations is correct.

## 1. INTRODUCTION

Research has shown that people's privacy preferences are context-dependent [23, 24] and multi-dimensional [5, 15, 19, 30]. The complexity of these preferences means that systems like Facebook that manage large amounts of personal user data have to resort to "labyrinthian" privacy controls [7]. As a result, many users experience difficulties managing their privacy settings [10, 17, 20] and many even avoid the hassle of changing their privacy settings altogether [6].

Privacy researchers have proposed the use of personas and segmentation as a means to simplify users' privacy decisions without giving contextual factors an overly reductionist treatment (e.g. [15, 31]). And recently, scholars have taken this contextually tailored approach a step further, by proposing privacy *personalization* or *adaptation* as a highly dynamic, user-tailored, and context-aware approach to privacy decision support at an individual level [3, 11, 12, 18, 25, 32].

In the broadest sense, a "privacy adaptation procedure" entails predicting users' privacy behaviors with machine learning techniques. Within this context, our paper is concerned with the more specific case of predicting whether or not a user will disclose a requested piece of information, based on which of the last *n* pieces of requested information this person had disclosed. If the system can predict with high likelihood that the user will not answer a specific question, it may decide to skip that question in order not to burden or upset the user.

This paper explores a way to improve this prediction process. Analyzing a number of datasets, we find that the order of requests has an impact on prediction accuracy. The conventional explanation for this effect is that different request orders vary in their ability to quickly overcome the cold start problem that is common in personalized systems. However, we demonstrate that this effect may *not* be related to the cold start problem. Instead, we provide evidence for the more provocative theory that certain request orders can actually cause users to *behave more predictably and less variably* in their information disclosure.

Previous research has shown that the order of requesting personal information can change users' *level* of disclosure [2, 13], but how is it possible that the request order influences the *predictability* of users' disclosure behavior? In this paper we formulate several explanations for this effect, and propose a study that will help find out which of these explanations is correct. We conclude by arguing the importance of investigating this topic: if request orders indeed influence the variability and thus the prediction accuracy of subsequent requests, then adapting the request order to the user may result in positive feedback loops (where adaptation and behavior "perpetually" promote each other).

## 2. THEORY AND RELATED WORK

### 2.1 Predicting disclosure behavior

Predicting users' disclosure behavior can be done by machine learning techniques such as decision tree analysis [28], which is very efficient and powerful in finding logical connections between the predicted item and the known items. A growing body of work concerns itself with predicting users' disclosure behaviors, especially in the field of mobile (location-aware) systems.

For example, Sadeh et al. [28] apply a *k*-nearest neighbor (kNN) algorithm and a random forest algorithm to learn users' privacy preferences in a location-sharing system. They show that users had difficulties setting their privacy preferences, and that the applied machine learning techniques can help users in specifying more accurate disclosure preferences.

Knijnenburg and Jin [12] also explored ways to help users with location-sharing decisions, and showed that users would feel assisted by privacy recommendations, but that the input required for these recommendations would counter the positive effects on their satisfaction. A study by Xie et al. [32] tried to overcome this problem by predicting users' in-situ sharing preferences based on the context and previous disclosure behavior. They showed that although users' location-sharing behaviors were highly dynamic,



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

context-aware, audience-aware and personal, they were still able to predict users' sharing preferences in various contexts.

Pallapa et al. [25] proposed context-aware approaches to privacy preservation in wireless and mobile pervasive environments. One of their solutions leverages the history of interaction between users to determine the level of privacy required in new situations. They demonstrate that this solution can deal with the rise of privacy concern while at the same time efficiently supporting users in a pervasive system full of dynamic and rich interactions.

Finally, in a social network context, Fang and LeFevre [8] developed a privacy wizard that is able to configure users' privacy settings automatically and accurately with a machine learning model that they developed. The wizard removes the burden of setting privacy settings using tools that most users would otherwise find too difficult to understand and use.

## 2.2 Avoiding the cold start problem

A privacy adaptation procedure as presented in the introduction is essentially a conversational recommender system. In its purest form, such a recommender system presents items that it predicts the user will like, and it bases this prediction on the ratings that the user gave to previously presented items. Similarly, a privacy adaptation procedure asks questions that it predicts the user is likely to answer, and it bases each prediction on the user's willingness or refusal to answer previous questions.

It goes without saying that conversational recommenders gain a better understanding of users' preferences as feedback accumulates. In early stages, the recommendation quality is not very good due to a lack of feedback; this is the so-called "cold start problem" [29]. Research has shown that the cold start problem can be avoided by not always presenting the top predicted items to the user: although there is a high chance that users will like the top predicted items (which is good), this also means that the feedback on these items is likely to be positive, and this feedback will thus not give the system new information about the user's preferences (which is bad). To learn more about the user's preferences, it is actually better to request items that span a wide range of attribute values [21, 22, 27].

Similarly, then, if the privacy adaptation procedure wants to learn users' multidimensional privacy preferences instead of asking users only non-sensitive questions (i.e. questions they are most likely to answer), it should rather ask a mix of sensitive and non-sensitive questions, spanning all dimensions.

## 2.3 Testing the argument

In a study on users' information disclosure behavior to an Android app recommender system [13], we asked 493 Mechanical Turk participants (266 female; median age group: 25-30, range: 18 to older than 60) to disclose 19 demographic items (e.g. gender, income, which related to participants themselves) and to give their permission to track 12 context items on their smartphones (e.g. location, web browsing, which related to participants' online experiences). The dependent measure was whether or not users disclosed the requested information or allowed the tracking. The order of the requests was manipulated: the system asked half the participants the demographics questions first, and the other half the context-related questions first. Regarding our current argument, we can ask the question: "If we want to predict users' disclosure behavior in this study, which request order would result in the highest prediction accuracy?"

Our analysis of disclosure behavior (Figure 1) shows that the disclosure of demographic items is generally high (i.e. varies very little), while the disclosure of context-related items varies wildly per item. According to the aforementioned argument, it would be better to request the context-related items first, because its variability contains more information about users' preferences, and it will therefore increase the prediction accuracy of subsequent items.

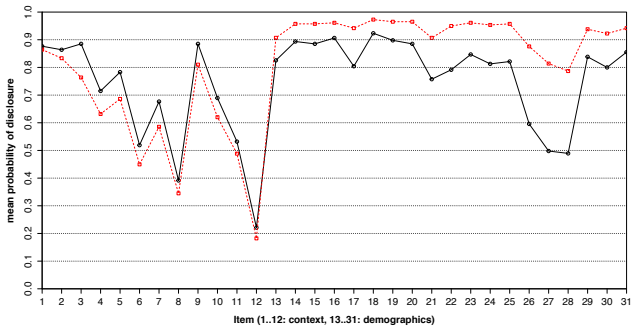


**Figure 1: Mean disclosure of each item when asking context first (black) versus demographics first (red).**

Interestingly though, when we run a number of recommender algorithms on this data to predict users' disclosures, we find that the recommender is able to predict disclosure fairly accurately even when it only uses the preceding five (instead of all) items as a basis for learning (Table 1 and Figure 2; this is similar to [26]). When asking the highly varying context items first, using these items improves the prediction accuracy of the final demographics items only by a little bit compared to not using them (see the solid and dotted black lines in Figure 2 for request time 13-31).

**Table 1: Prediction accuracy of our algorithm.**

| Data used | Algorithm | Accuracy | |
|---|---|---|---|
| | | Context first | Demographics first |
| N = ALL | Naïve Bayes | 82.65% | 88.86% |
| | LogReg | 87.66% | 92.80% |
| | J48 | 88.99% | 92.03% |
| N = 5 | Naïve Bayes | 84.25% | 89.06% |
| | LogReg | 86.70% | 91.01% |
| | J48 | 87.12% | 91.11% |

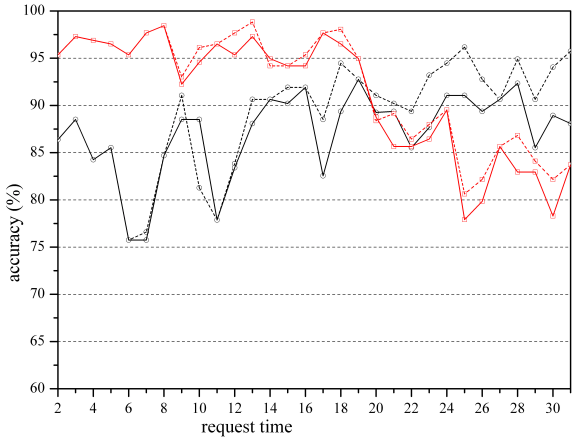


**Figure 2: The prediction accuracy of our J48 recommender for each item when asking context first (black) versus demographics first (red), based on the preceding 5 (solid) or all preceding items (dotted).**

Despite this, though, the request order still has an effect on recommendation accuracy, even when only the preceding 5 items are used. Specifically, our recommender is more accurate when we ask the demographics items first than when we ask the context items first[1] (see Table 1). This is the exact opposite of what we would expect based on the argument presented in Section 2.2.

## 2.4 An alternative theory

Since recommendation accuracy does not depend on an extensive history of previous disclosure, we argue that participants somehow behaved more *predictably* when we asked the demographics items first. Since behavior is easier to predict when it is less variable, we believe that the request order has an effect on the inherent *variability* of users' disclosure behavior.

How can request order influence the variability of users' disclosure behavior? We already mentioned that demographics items are disclosed at a more consistently high rate than context items, which vary considerably in their average level of disclosure (see Figure 1). Our hypothesis is that when demographics items are requested first, this uniform behavioral pattern "spills over" into the context items, which are then also answered more uniformly (and are consequently easier to predict). Conversely, since most users answer some context items positively and some negatively, this varied behavioral pattern also "spills over" into the demographics items, which are then answered in a more varied fashion (and are consequently harder to predict).

Figure 3 shows evidence in support of this argument: the variability of users' disclosure of both types of information is lower when demographics are requested first (red, dotted line) than when context is requested first (black, solid line). In light of the prediction results (Figure 2), it thus seems that the monotonous behavior induced by asking demographics items first is indeed easier to predict than the varied behavior induced by asking context items first; in other words, users behaved more *predictably* and less *variably* when asking demographics items first.

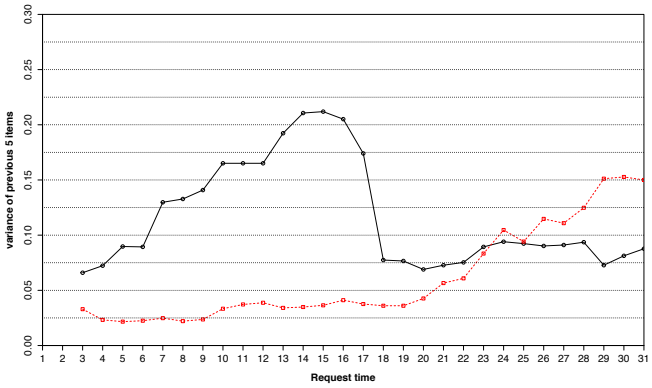


**Figure 3: Mean variability of the 5 previously requested items, asking context first (black) versus demographics first (red).**

[1] Note that we are only considering prediction accuracy here. When it comes to the disclosure rate, Acquisti et al. [2] show that asking the more risky questions (arguably those that receive a more varied response) first will increase overall disclosure.

There are three possible explanations as to how this "spill over" effect may occur. The first explanation is that if requests are all rather non-sensitive (as seems to have been the case for demographics items in the aforementioned study), users get used to the monotonous behavioral pattern in the first part, and are therefore more likely to perform the same monotonous behavior in the second part. In this case, the response pattern shows evidence of *habitual* behavior [1, 4]. Conversely, if the initial requests are mixed in terms of sensitivity, users' behavioral patterns will be more varied, and users will then not get used to a monotonous response pattern.

The second explanation is more *cognitive* in nature. Existing research has shown that privacy-awareness can be increased or inhibited with subtle visual or interactive primes [9, 14], and this increased awareness may make users more likely to perform a "privacy calculus" (i.e. reasoned decision behavior). By the same token, a high (vs. low) variability of initial disclosure behavior may encourage (vs. inhibit) users' privacy calculus. Once users are primed (vs. inhibited) with this privacy calculus in the first part, they are more (vs. less) likely to perform it in the second part as well. Performing a privacy calculus arguably results in more reasoned but less monotonous disclosure decisions. In effect, a mixed (vs. equal) sensitivity of initial requests leads to varied (vs. monotonous) disclosures in subsequent requests.

The third explanation rests on the fact that context requests are less common than demographics requests: In general, people (and specifically Mechanical Turk participants, who answer questions about themselves all the time) do not use the privacy calculus very much for answering demographics questions, because they developed a pattern of answering most demographic questions. However, people generally do use a privacy calculus for answering context questions, with which they are usually much less familiar (this also explains why the variance in demographics disclosures is inherently lower than in context disclosures). Now, if people answer demographics items first, the frequent non-use of the privacy calculus gets carried over to the context items, and hence context items have a lower variance when they follow a batch of demographics items. By the same token, if people answer context items first, the frequent use of the privacy calculus for this type of information carries over to the demographics items, and hence demographics items have a higher variance when preceded by a batch of context items.

In section 3 we propose an experiment that can demonstrate which of these three explanations is correct. First, though, we confirm the general theory that request order influences disclosure variability and prediction accuracy on a separate dataset.

## 2.5 Confirmation of the new theory

In a study (N=390) on users' disclosure behavior to a recommender system that performs client-side personalization [16], demographics and context items were requested in an alternating fashion. Since the requested context info is generally more sensitive, this leads to requests of mixed sensitivity (explanation 1 and 2), and also accentuates the uncommon context requests (explanation 3). In effect, we can conjecture:

H1. The disclosure variability of each item type will be higher than in either condition of the Android app recommender study.

H2. The prediction accuracy will be lower than in either condition of the Android app recommender study.

And to confirm:

H3. Like in the Android app recommender study, the disclosure variability of the context data will be higher than that of the demographics data.

H4. Like in the Android app recommender study, the prediction accuracy of the context data will be lower than that of the demographics data.

Figure 4 confirms hypotheses H1 and H3: context variability is indeed higher than demographics variability, and they are both higher than in the App Recommender study.

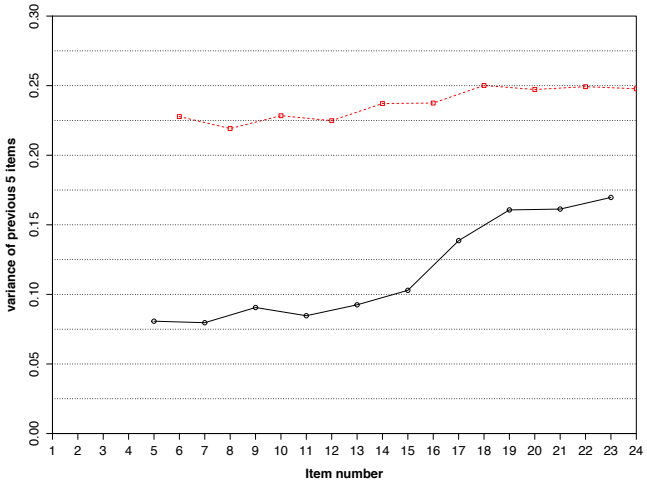


**Figure 4: Mean variability of the 5 previously requested items of the same type, for demographics (black) and context (red).**

Figure 5 confirms hypotheses H2 and H4: prediction accuracy for context is indeed lower than for demographics, and they are both lower than in the App Recommender study.

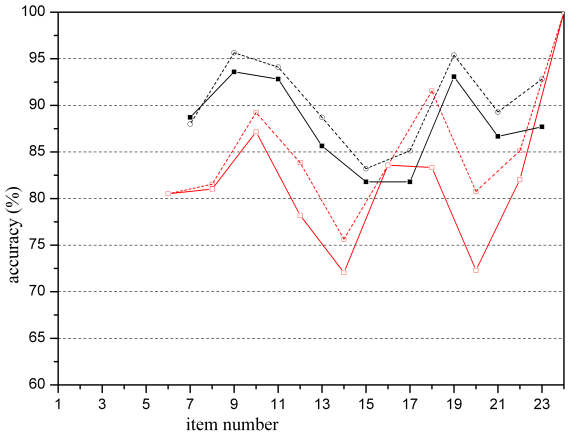


**Figure 5: The prediction accuracy of our J48 recommender for each item for demographics (black) and context (red), based on the preceding 5 (solid) or all preceding items (dotted).**

# 3. PROPOSED FOLLOW-UP STUDY

We discovered the effect of request order on the predictability of user disclosure behavior when analyzing the Android app recommender study data, and we confirmed the effect on the client-side personalization study data. Now, we propose to conduct an additional experiment that tests this effect as an *ex ante* hypothesis. The experiment will also disentangle the three possible explanations for the effect.

Explanations 1 and 2 regard the *varied sensitivity* of the presented items as the main reason for the effect. Hypothesizing along these lines, we expect the effect to occur when the first set of requests contains both sensitive and non-sensitive items, but not when it contains only sensitive or only non-sensitive items.

Explanation 3 regards the *type* of the presented items as the main reason for the effect. Hypothesizing along these lines, we expect the effect to occur when the first set of requests consists of context items, but not when it consists of demographics items only.

Finally, explanations 2 and 3 claim that users in the effect-conditions are more likely to use a *privacy calculus* than users in the other conditions. Explanation 1, on the other hand, claims that the effect is purely behavioral, and there should thus be no differences in privacy calculus between conditions. We test this by asking users to evaluate the perceived risk and benefit of disclosing each piece of information at the end of the study. If explanation 2 or 3 is correct, users' evaluation of perceived risk and benefit is a stronger predictor of disclosure in the effect conditions than in the other conditions. If this is not the case then explanation 1 makes more sense.

## 3.1 Participants and procedure

Participants will be recruited via Amazon Mechanical Turk. They will be briefed about a mobile recommender system that uses demographic information and smartphone context tracking to provide personalized recommendations. A prototype version of this system then presents them 12 requests for context and/or demographics items that they can choose to disclose or not. After the disclosure part, they will answer a set of questions about the benefits and risks of disclosing each of the 12 items, as well as some questionnaires about their system-specific privacy concerns and anticipated satisfaction with the system.

## 3.2 Manipulations

The experiment implements an orthogonal 2x3x2x2 between-subjects design. Participants are randomly assigned to one of the 24 conditions. The manipulations are:

- Type of item for first 6 requests: context or demographics
- Sensitivity of item for first 6 requests: sensitive or non-sensitive or alternating
- Type of item for the last 6 requests: context or demographics
- Sensitivity of item for the last 6 requests: sensitive or non-sensitive

The first two manipulations enable us to test our explanations; the last two are manipulations allow us to test whether of our results hold for any type of item. This leads to the following experimental conditions:

| **First 6 requests** | | **Last 6 requests** |
|---|---|---|
| Cs-Cs-Cs-Cs-Cs-Cs | X | Cs-Cs-Cs-Cs-Cs-Cs |
| Ds-Ds-Ds-Ds-Ds-Ds | | Ds-Ds-Ds-Ds-Ds-Ds |
| Cn-Cn-Cn-Cn-Cn-Cn | | Cn-Cn-Cn-Cn-Cn-Cn |
| Dn-Dn-Dn-Dn-Dn-Dn | | Dn-Dn-Dn-Dn-Dn-Dn |
| Cs-Cn-Cs-Cn-Cs-Cn | | |
| Ds-Dn-Ds-Dn-Ds-Dn | | |

Cs stands for sensitive context items, Ds for sensitive demographic items, Cn for non-sensitive context items, and Dn for non-sensitive demographics items. Note that the alternating-sensitivity conditions can either start with a sensitive or a non-sensitive item; we expect that this will have no substantial effect, but to be

certain we randomly split participants in this condition into a sensitive-start and a non-sensitive-start group.

## 3.3 Item sensitivity pre-study

The main study requires 12 items in each category (Cs, Cn, Ds, Dn). We have selected these items from a set of 96 candidate items that were developed in a collaborative effort by the researchers and their colleagues.

The candidate items were tested for average disclosure levels in a separate pre-study with 200 Amazon Mechanical Turk workers. Eligibility was restricted to U.S. citizens with a high worker reputation. Each participant was presented with half of the candidate items, and asked whether they would disclose each item or not. The order of the items was counter-balanced. From the set of candidate items, we removed items that were overly sensitive (disclosure rate <10%), overly non-sensitive (disclosure rate >90%) or ambiguous (disclosure rate 45-55%).

The remaining items consisted of 16 Cs items, 13 Cn items, 16 Ds items, and 25 Dn items. We selected 12 items from each set so that the averages and variability of the sensitive and non-sensitive items roughly matched between the context and demographics items (see Table 2). These are the items that will be used in the main study.

**Table 2: Items selected for the main study.**

| **Sensitive context items (Cs)** *M* = 21.4%, *sd* = 6.30% | **Sensitive demographic items (Ds)** *M* = 20.5%, *sd* = 5.61% |
|---|---|
| Email messages | Pornography preferences |
| Phone's call log | Number of sexual partners |
| Phone's microphone | Credit score |
| Friends' comments on social netw. | Work phone number |
| Contact list | Work address |
| Personal work files | Illegal drug usage |
| Online bills | Birth control usage |
| Phone's notepad | Home address |
| Calendar entries | Social services used recently |
| Phone's location record | Traffic violations |
| Phone's IP address | Name of employer/company |
| Websites browsed on phone | Ever been evicted? |
| **Non-sensitive context items (Cn)** *M* = 64.3%, *sd* = 7.86% | **Non-sensitive demogr. Items (Dn)** *M* = 67.6%, *sd* = 7.57% |
| Your mobile app usage | Email address |
| Internet recommendations | Religious belief |
| Mobile apps usage time | Political preferences |
| News items read on phone | Weight |
| Mobile browser home page | Hometown |
| Music on phone | Hours spent on work |
| Downloaded mobile apps | Number of cars |
| Phone water damage sensor status | Relationship status |
| Phone's remaining storage space | Time spent on fitness |
| Number of phone restarts | Personality (e.g. honest, reliable) |
| Mobile games and high-scores | Highest degree earned |
| Phone's remaining battery life | Field of work |

## 3.4 Measurement

The main dependent variables are the disclosure variability of the last 6 items, and the prediction accuracy of an algorithm that tries to predict users' disclosure behavior of the final item, using the 5 preceding items. Using only the last 6 items for the prediction and variability analysis will allow us to measure the spill over effect from the first 6 items.

To determine whether users' privacy calculus is also affected by the manipulations, participants will indicate their perception of the risks and benefits of disclosing each of the 12 items on a 7-point scale (cf. [14] for a similar procedure). Finally, participants will answer general questions about their system-specific privacy concerns and anticipated satisfaction with the system. These are used as control variables.

## 3.5 Expected effects and main experiment

Figure 6 shows the effects we predict to hold, depending on which explanation is correct. Explanation 1 implies that if the first 6 items have a varying sensitivity, this will lead to a more varied behavioral response and consequently lower predictability of the last 6 items (measured by the prediction accuracy of the final item, based on the 5 preceding items). Therefore, we accept explanation 1 if (a) the disclosure variability of the last 6 items is higher and (b) the prediction accuracy of the last item is lower in the "alternating first-6" conditions than in the "sensitive first-6" or "non-sensitive first-6" conditions.

Explanation 2 additionally implies that this effect is not simply behavioral, but also cognitive: if the first 6 items have a varying sensitivity, users will have a stronger tendency to perform a privacy calculus (i.e. actively weigh the risk and relevance of disclosing the information, cf. [14]) in their disclosure of the last 6 items. Therefore, we accept explanation 2 if—in addition to (a) and (b)—(c) perceived risk and relevance play a more important role in users' decisions regarding the last 6 items in the "alternating first-6" conditions than in the "sensitive first-6" or "non-sensitive first-6" conditions.

Finally, explanation 3 implies that these effects do not happen when the first 6 items vary in sensitivity, but rather when they are context requests. Therefore, we accept explanation 3 if (a) the disclosure variability of the last 6 items is higher, (b) the prediction accuracy of the last itemsis lower, and (c) perceived risk and relevance play a more important role in users' decisions regarding the last 6 items in the "context first-6" conditions than in the "demographics first-6" conditions.

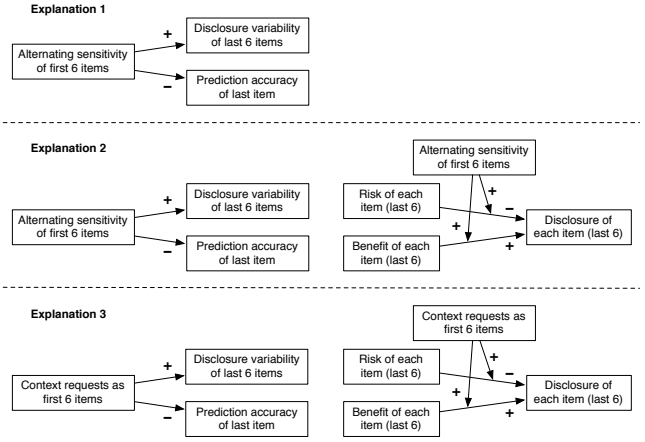


**Figure 6: Predicted effects of the proposed study, depending on which explanation is correct.**

# 4. CONCLUSION

In this paper, we set out to improve the prediction of participants' disclosure behavior, only to discover an interesting phenomenon regarding the impact of the request order on their disclosure behavior: the order of requesting items increases the *variability* and *predictability* of users' disclosure pattern and, consequently, reduces the accuracy of our prediction algorithms.

We provided three possible explanations for this effect: 1) users' disclosure *habitually* becomes more monotone when the initial

requests are all sensitive or non-sensitive (as compared to when these request are more variable in sensitivity); 2) users react *cognitively* to more requests that are more variable in sensitivity by employing a privacy calculus; and 3) users react by employing a privacy calculus when they encounter the less frequent context-related rather than the more frequent demographic items.

Finally, we described a study in this paper that will help us to disambiguate these three explanations. Our initial analyses of the data from this described study indicate that a combination of the mentioned hypotheses could be correct: we find that both the "alternating first-6" and the "context first-6" conditions lead to lower the prediction accuracy. At the workshop we will present more detailed analyses of our initial results.

The outcomes of our study will have important implications for the accurate prediction of users' disclosure behavior in any system that employs a privacy adaptation procedure. Our future work will develop such adaptive systems as a means to alleviate the burden of enacting complex privacy settings in various online environments.

# 5. REFERENCES


[1] Aarts, H., Verplanken, B. and van Knippenberg, A. 1998. Predicting Behavior From Actions in the Past: Repeated Decision Making or a Matter of Habit? *Journal of Applied Social Psychology*. 28, 15 (Aug. 1998), 1355–1374. DOI= http://dx.doi.org/10.1111/j.1559-1816.1998.tb01681.x.

[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] Benisch, M., Kelley, P.G., Sadeh, N. and Cranor, L.F. 2011. Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs. *Personal Ubiquitous Computing*. 15, 7 (Oct. 2011), 679–694. DOI= http://dx.doi.org/10.1007/s00779-010-0346-0.

[4] Betsch, T., Haberstroh, S., Glöckner, A., Haar, T. and Fiedler, K. 2001. The Effects of Routine Strength on Adaptation and Information Search in Recurrent Decision Making. *Organizational Behavior and Human Decision Processes*. 84, 1 (Jan. 2001), 23–53. DOI= http://dx.doi.org/10.1006/obhd.2000.2916.

[5] Buchanan, T., Paine, C., Joinson, A.N. and 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, 2 (2007), 157–165. DOI= http://dx.doi.org/10.1002/asi.20459.

[6] Compañó, R. and Lusoli, W. 2010. The Policy Maker's Anguish: Regulating Personal Data Behavior Between Paradoxes and Dilemmas. *Economics of Information Security and Privacy*. T. Moore, D. Pym, and C. Ioannidis, eds. Springer US. 169–185.

[7] Facebook & your privacy: Who sees the data you share on the biggest social network?: 2012. *http://www.consumerreports.org/cro/magazine/2012/06/facebook-your-privacy*. Accessed: 2012-05-13.

[8] Fang, L. and LeFevre, K. 2010. Privacy Wizards for Social Networking Sites. *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, NC, 2010), 351–360. DOI= http://dx.doi.org/10.1145/1772690.1772727.

[9] 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.

[10] Kelley, P.G., Cranor, L.F. and Sadeh, N. 2013. Privacy As Part of the App Decision-making Process. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France, 2013), 3393–3402. DOI= http://dx.doi.org/10.1145/2470654.2466466.

[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. 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.

[14] 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).

[15] 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.

[16] Kobsa, A., Knijnenburg, B.P. and Livshits, B. 2014. Let's Do It at My Place Instead?: Attitudinal and Behavioral Study of Privacy in Client-side Personalization. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Canada, 2014), 81–90. DOI= http://dx.doi.org/10.1145/2556288.2557102.

[17] Lipford, H.R., Besmer, A. and Watson, J. 2008. Understanding Privacy Settings in Facebook with an Audience View. *Proc. of the 1st Conference on Usability, Psychology, and Security* (Berkeley, CA, USA, 2008).

[18] Liu, B., Lin, J. and Sadeh, N. 2014. Reconciling Mobile App Privacy and Usability on Smartphones: Could User Privacy Profiles Help? *Proceedings of the 23rd International Conference on World Wide Web* (Republic and Canton of Geneva, Switzerland, 2014), 201–212. DOI= http://dx.doi.org/10.1145/2566486.2568035.

[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] Madden, M. 2012. *Privacy management on social media sites*. Pew Internet & American Life Project, Pew Research Center.

[21] McGinty, L. and Smyth, B. 2006. Adaptive Selection: An Analysis of Critiquing and Preference-Based Feedback in Conversational Recommender Systems. *International Journal of Electronic Commerce*. 11, 2 (Dec. 2006), 35–57.

[22] Mirzadeh, N., Ricci, F. and Bansal, M. 2005. Feature selection methods for conversational recommender systems. *Proceedings of the 2005 IEEE International Conference on e-Technology, e-Commerce and e-Service* (2005), 772–777. DOI= http://dx.doi.org/10.1109/EEE.2005.75.

[23] Nissenbaum, H. 2004. Privacy as Contextual Integrity. *Washington Law Review*. 79, (2004), 119–157.

[24] Nissenbaum, H.F. 2009. *Privacy in context : technology, policy, and the integrity of social life*. Stanford Law Books.

[25] Pallapa, G., Das, S.K., Di Francesco, M. and Aura, T. in press. Adaptive and context-aware privacy preservation exploiting user interactions in smart environments. *Pervasive and Mobile Computing*. (in press). DOI= http://dx.doi.org/10.1016/j.pmcj.2013.12.004.

[26] De Pessemier, T., Dooms, S., Deryckere, T. and Martens, L. 2010. Time dependency of data quality for collaborative filtering algorithms. *Proceedings of the fourth ACM conference on Recommender systems* (Barcelona, Spain, 2010), 281–284. DOI= http://dx.doi.org/10.1145/1864708.1864767.

[27] Rashid, A.M., Albert, I., Cosley, D., Lam, S.K., McNee, S.M., Konstan, J.A. and Riedl, J. 2002. Getting to know you: learning new user preferences in recommender systems. *Proceedings of the 7th international conference on Intelligent user interfaces* (San Francisco, CA, 2002), 127–134. DOI= http://dx.doi.org/10.1145/502716.502737.

[28] Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M. and Rao, J. 2009. Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing*. 13, 6 (2009), 401–412. DOI= http://dx.doi.org/10.1007/s00779-008-0214-3.

[29] Schein, A.I., Popescul, A., Ungar, L.H. and Pennock, D.M. 2002. Methods and Metrics for Cold-start Recommendations. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Tampere, Finland, 2002), 253–260. DOI= http://dx.doi.org/10.1145/564376.564421.

[30] Wisniewski, P., Knijnenburg, B.P. and Richter Lipford, H. 2014. Profiling Facebook Users' Privacy Behaviors. *SOUPS2014 Workshop on Privacy Personas and Segmentation* (Menlo Park, CA, 2014).

[31] Workshop on Privacy Personas and Segmentation (PPS): Call for Papers: 2014. *https://cups.cs.cmu.edu/soups/2014/workshops/privacy.html*. Accessed: 2014-04-12.

[32] Xie, J., Knijnenburg, B.P. and Jin, H. 2014. Location Sharing Privacy Preference: Analysis and Personalized Recommendation. *Proceedings of the 19th International Conference on Intelligent User Interfaces* (Haifa, Israel, 2014), 189–198. DOI= http://dx.doi.org/10.1145/2557500.2557504.