Privacy Perceiver: Using Social Network Posts to Derive Users' Privacy Measures

Frederic Raber
DFKI Saarland Informatics Campus
Saarbruecken, Saarland
frederic.raber@dfki.de

Antonio Krüger
DFKI Saarland Informatics Campus
Saarbruecken, Saarland
krueger@dfki.de

ABSTRACT

Current research has shown that a person's personality can be derived from written text on Facebook or Twitter, as well as the amount of information shared on their personal social network sites. So far, there has been no further investigation on whether a person's privacy measures can be extracted from these information sources. We conducted an explorative online user study with 100 participants; the results indicate that privacy concerns can be derived from written text, with a prediction precision similar to personality. At the end of the discussion, we give specific guidelines on the choice of the correct data source for the derivation of the different privacy measures and the possible applications of those.

ACM Reference Format:
Frederic Raber and Antonio Krüger. 2018. Privacy Perceiver: Using Social Network Posts to Derive Users' Privacy Measures. In *UMAP'18 Adjunct: 26th Conference on User Modeling, Adaptation & Personalization, July 8–11, 2018, Singapore.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3213586.3225228

1 INTRODUCTION

The latest privacy scandal on Facebook made it clear that it is needed to first understand users' privacy concerns, and second, also to comply to these concerns by suggesting new privacy settings. One way of measuring privacy concerns is by questionnaires like the Internet User's privacy concern scale (IUIPC) [11]. Research has also begun to facilitate these privacy measures to recommend app permission settings [14], privacy settings for intelligent retail data [15] or to recommend recipients for a new Facebook post [13]. Although the personality and privacy measures can be used in many different use-cases, the additional effort to fill in questionnaires leads users to avoid using such recommender systems, even though they could benefit from them.

Unfortunately, there is no correlation between the desired privacy and the actual privacy settings [2], making it hardly possible to infer privacy settings or privacy attitudes through observation, e.g. through extracting the privacy settings of similar users from an online social network, and propose them to users with a similar usage profile. Our work therefore tries to go a different way of deriving the privacy measures, namely through a textual analysis of posts written on Facebook or Twitter. Whereas this field is already sufficiently investigated for *personality* measures, the prediction of *privacy* measures has not been taken into account so far. With our work, we try to shed light on this aspect, and give some guidelines on how the privacy measures (in terms of the three IUIPC measures and the westin privacy index) can be predicted using social network data. These results can be used in the next step to build a recommender system that, based on the privacy measures, can recommend privacy settings for various domains like mobile app permission settings [14] or selecting the Facebook post audience [13] or inferring privacy settings for data collected inside an "intelligent retail store" [15].

The main goal of our research paper is not to derive specific privacy settings for social networks, but to derive the general privacy concern according to the IUIPC questionnaire and the westin privacy index, that can be used in a second step to derive privacy settings or recommendations for various domains like social networks or mobile app settings, as described in section 5.3 later. In this paper, we try to elaborate **which** data source (Facebook profile data or Facebook/Twitter posts) and, in the second step, which language and profile features lead to the best prediction results, and exactly how precise the prediction is with the different data sources.

We performed a user study to capture privacy measures using traditional questionnaires, and recorded three data sets as possible sources for the prediction: language features from Facebook posts and Twitter posts, as well as Facebook profile information like number of entered workplaces, number of friends, number of items liked, etc. Based on the study data, we perform a regression and machine learning analysis to determine *whether and how precisely* the personality and privacy measures can be predicted. The paper ends with a discussion about which data set should be used to predict each of the different measures using machine learning. Our results show that privacy measures *can* be predicted significantly better than a random baseline, in the best case by using either Twitter or Facebook language features.

2 RELATED WORK

For the scope of this paper, there are three major topics that are of interest in the literature: conventional methods capturing the privacy concerns of a user, the extraction of personality using social network data including posts as well as profile information, and finally, applications of the privacy measures in current research.

UMAP'18 Adjunct, July 8–11, 2018, Singapore

ACM ISBN 978-1-4503-5784-5/18/07…$15.00
https://doi.org/10.1145/3213586.3225228

2.1 Capturing privacy concerns using questionnaires

Some of the earliest publications proposing a privacy questionnaire to assign consumers to three different privacy categories include Alan Westin's work on consumer privacy indices, later summarized by Kumaraguru and Cranor [10].

The newer PCS[1] questionnaire [3] is more detailed and consists of 28 questions in four categories: General Caution, Technical Protection and Privacy Concern. Although more detailed, the questionnaire still addresses the general privacy concerns of a person; furthermore it includes technical questions (for example about shredding floppy discs and CDs because of privacy issues) which seem outdated nowadays.

There also exist questionnaires that are tailored to a specific domain, like the CFIP[2] [17] and its follow-up, the IUIPC [11] questionnaire, which were designed explicitly to measure the privacy concerns of internet users. The authors found that the privacy concerns regarding online companies can be expressed well using three measures: the *control* measure, which determines how far a subject desires to have control over the disclosure of her personal information, the desired *awareness* on to whom the personal information is disclosed, and *collection*, describing how important it is for the subject to know which personal data is collected. The questionnaire consists out of seven questions for the three aforementioned questions; four for the *Collection* measure, and three for *Control* and *Awareness*, respectively. The scales range from strongly disagree to strongly agree on a seven-point scale. The IUIPC is, aside from the Westin privacy indices, one of the most frequently used questionnaires and is also used to predict privacy settings on Facebook or app permissions [13, 14].

2.2 Extraction of personality measures from social media

One of the earliest publications [1] used only a small part of the information available in a Facebook profile, like the number of friends, photos, groups etc. In contrast to our work and other recent work, the authors did not conduct an individual prediction of the big five personality traits, but rather clustered all participants into ten distinct clusters for each of the features. The evaluation compared the predicted value for each of the clusters with the average value of the personality measure of all participants inside that cluster. An evaluation of an individual prediction was not conducted.

Different approaches later used *annotation activities* [12] or the writing style on twitter [9], Facebook or Youtube [7] to predict **personality** measures with the aid of regression algorithms. Other researchers have shown that, also the "dark triangle" personality traits like narcissism, machiavellianism and psychopathy can be predicted using a crowd sourced machine learning approach [19]. However, the prediction of *privacy measures* using social network data or language features has not been examined so far.

2.3 Applications of personality and privacy measures

Therefore research has begun to use them to predict many different decisions around users, for example which followees should be recommended to Twitter users [18] or the music taxonomy for an online music streaming site [8]. The latest research has shown that even mortality and a number of specific diseases correlate with some of the big five measures, especially conscientiousness, neuroticism and openness [20]. The personality inventory is also used in several business use-cases, for example in the TWIN recommendation tool that allows one to filter for interesting hotel ratings on TripAdvisor by displaying comments of users with a similar personality profile [16]. Personality traits are also widely used in targeted advertising, for example on Facebook [6] and Twitter [5].

The IUIPC privacy measures have been used especially in the research field of privacy and security. Raber et al. showed that it is possible to infer permission settings for smartphone apps using the privacy measures provided by a questionnaire [14]. Even the audience of Facebook posts and the disclosement policies for intelligent retail data [15] can be predicted using personality, or for better results, using the IUIPC measures [13].

3 DATA COLLECTION

We conducted an online user study to find out whether the privacy measures of a user can be predicted with the amount of information items shared with friends on his or her personal social network site, as well as writing style on social network sites. We recorded the privacy measures using the IUIPC questionnaire and Westin Privacy Index, as stated in the introduction. We analyzed the Facebook and Twitter posts of the participants using the LIWC 2015[3] text analysis software, which is explicitly also suited for analyzing Facebook and Twitter posts[4].

Besides of language features extracted from the users' Facebook and Twitter posts, we also investigated whether the amount of provided profile information can be used to derive the privacy measures of the user. Related research [9] has shown that for predicting personality, it is more important *whether* a Facebook profile is filled out or how many entries exist, rather than the actual content. We therefore recorded which information fields (later called "profile features") on the user's personal profile page are filled out and shared with the Facebook friends, in addition to the language features. The next section will give more details on which information was recorded, as well as the detailed procedure.

3.1 Methodology

3.1.1 Online questionnaire. The study was conducted as an online survey. The participants were recruited using Prolific Academic,[5] which allows us to select only participants that are actively using Facebook or Twitter. Studies in the past have shown that participants who are recruited via online services, as in our case, lead to a similar quality of results as when participants are recruited at a university [4]. To avoid overly reducing the set of participants to a too-small part of the population, we did *two* studies with the

[1] Privacy Concern Scale

[2] Scale of Concern For Information Privacy

[3] Linguistic Inquiry and Word Count

[4] https://www.receptiviti.ai/science

[5] https://www.prolific.ac/, last accessed 09-07-2017

same setup, one for gathering Facebook data, and one for gathering Twitter data. For the first study, we selected only active Facebook users, and did not require them to be active Twitter users. Nevertheless, if participants were Twitter users, they were able to provide their screen name for our analysis. In the second study, we required only active Twitter users. Therefore the amount of participants and data sets for the Facebook and the Twitter data sets slightly differ ($n_{Facebook} = 104$, $n_{Twitter} = 109$). The participants were paid a compensation of £1 upon successful participation. To motivate the subjects to fill out the questionnaire honestly, the compensation was only paid after we checked the submitted data for plausibility by us by analyzing the control questions of the questionnaire, and confirming that all questions are answered. If thereby a subject was rejected, a new participant was recruited to fill in the gap. Thus we had exactly 110 results for each of the two studies. In the Facebook group, six participants had to be filtered out, as their profile contained too few posts to do a correct language analysis (at least 300 words are required). The same happened to only one user in the Twitter group, resulting in 104 participants from the Facebook group, and 109 from the Twitter group. We had 54,4% female and 45,6% male participants aged from 18 to 72 years (average 33.02, SD 10.94).

The survey can be divided into two parts: In the first part, we posed the questions of the IUIPC (7-point scale from strongly disagree to strongly agree) and Westin privacy questionnaire (4-point scale from strongly disagree to strongly agree). For the second part, users had to befriend a test Facebook user and enter their Facebook ID (first study) or Twitter screen name (second study) into the questionnaire. All participants were informed about which data was accessed, and that the data was analyzed anonymously using a script, as described in the next section. We do not store the actual data, but only language features and *whether* the profile items were filled out or how many entries existed. The questionnaire ended with a final page where participants were able to enter comments and feedback for the survey.

3.1.2 Analysis of the social network profiles. We exported the questionnaire answers into a csv file and started a python script, which uses the selenium web automation toolkit[6] to traverse the Twitter and Facebook profiles to check which profile items are filled out and which are not. In contrast to the Facebook API which was often used in the past, we can access all data that is visible to a friend on the Facebook page, rather than the subset of information that is accessible using only the API. The script opens the profile page of each study participant, traverses all sections of their "About" page and collects the users' own posts (excluding posts made by friends) for the LIWC sentiment analysis. As earlier work has shown [9], it is more important *whether or how much* personal entries a user has provided (like *political view*, *religion* or past *workplaces*), rather than whether he is actually Catholic or Protestant, for example. We therefore only recorded whether the different entries in the about section (for more details see Table 1) were visible to friends or not and how many entries existed. The extracted post data from Facebook and Twitter was directly forwarded to the LIWC analysis tool; we only stored the language features for the statistical analysis.

Section	Observed fields
General profile information	number of status updates
Friends	number of friends
Work and Education	number of work entries, number of education entries
Places You've Lived	number of places
Contact and Basic Info	number of contact entries, number of basic information entries, birthday, gender, religion, political views
Family and Relationships	relationship status, number of family members
Life Events	number of life events
Photos	number of photos uploaded by the user, number of photos uploaded by friends, number of albums
Likes	total number of likes, number of movies, TV shows, music, books, sport teams/athletes liked
Events	number of events visited in the past
Reviews	total number of reviews

Table 1: Observed Facebook profile items. As discussed above, we either counted the number of visible entries, or *whether* an entry is visible to friends.

The procedure described in the last sections was reviewed and approved by the ethical review board of our institution.

3.1.3 Analysis of the Facebook and Twitter posts. The language analysis was done using the LIWC language analysis tool. Broadly speaking, LIWC uses a dictionary to count and categorize the words of a text to get an overview of how often the text refers to different categories like health or the household, or how often different punctuation marks are used, for example. For categorizing the words in a text, LIWC has a number of word sets, one for each category, containing words and word stems that fit to the respective category. Words starting with "vital" and "vitamin" or the names of vaccacines belong to the word set of the "health" category, for example. The measures that are delivered by LIWC (later called "language features") and that are later used as an input for the machine learning analysis, are defined as $F_{cat} = \frac{|words \in cat|}{|words|}$, where *cat* is a category word set.

4 REGRESSION AND MACHINE LEARNING ANALYSIS

We performed two different analyses: First we performed a regression analysis using SPSS to calculate the goodness of fit and standard error with our study data. Second, we used the scikit-learn python library to train a machine learning algorithm using our data, and evaluated the prediction precision in terms of the root

[6] http://www.seleniumhq.org//, last accessed 09-07-2017

mean squared error (RMSE) using a ten-fold cross validation. We tried out several algorithms, and found the support vector regression (SVR) to be most precise. As an input, we used the Facebook profile features (see 3.1.2) or the language features extracted from Twitter and Facebook posts (see 3). The output to be predicted were the IUIPC measures (control, collection, awareness) or the Westin privacy index. As recommended by related literature [7], we used a univariate instead of a multivariate regressor for our analysis. To avoid including meaningless features in the regression model, we used a *backwards elimination* heuristic using the Bayesian Information Criterion. The results for the precision with the regression and machine learning analysis are shown in Table 2. For each feature, we computed the standard error of the prediction against the correct result, the coefficient of determination (R^2), as well as the root mean squared error (RMSE) of the machine-learning prediction compared to the correct result. In addition to the regression and machine learning results, we also computed the RMSE using a random predictor. The results of the latter are independent from the input features and are thus only mentioned once in the table.

Measure	Regression anal.		ML anal.	Random
	stderr	R^2	**RMSE**	**RMSE**
Profile data				
- Control	.844	15.0	.866	2.018
- Awareness	.674	30.7	.825	2.070
- Collection	.956	18.2	1.093	2.010
- Privacy index	.633	7.1	.712	.732
Facebook data				
- Control	.635	51.9	.686	-
- Awareness	.423	71.7	.587	-
- Collection	.661	60.9	.894	-
- Privacy index	.330	74.8	.704	-
Twitter data				
- Control	.518	61.5	.816	-
- Awareness	.375	74.4	.660	-
- Collection	.643	68.3	1.17	-
- Privacy index	.453	55.4	.710	-

Table 2: Average stderr, RMSE and R^2 measures for the regression and machine learning analysis for the privacy measures (7-point IUIPC scale and 3-point westin privacy index).

The standard error and RMSE using the *profile data* are highest the privacy measures (RMSE between 0.71 and 1.09, stderr between 0.63 and 0.96). Both Facebook and Twitter language features deliver a notably better prediction for personality, whereas Facebook is slightly better (RMSE 0.55 and 0.73, stderr between 0.33 and 0.50) than Twitter (RMSE between 0.65 and 0.95, stdeer between 0.381 and 0.619). The IUIPC measures are predicted better using Facebook posts (RMSE between 0.59 and 0.89, stderr between 0.33 and 0.66) compared to Twitter (RMSE between 0.71 and 1.17, stderr between 0.38 and 0.64). Except for the the collection measure, which yields the highest prediction error with Twitter compared to all other data sources, the RMSE using Twitter data remains lower compared to a prediction using profile data or the personality traits as an input. The prediction precision of the conventional Westin privacy index remains similar for all data sources, between 0.65 using personality traits and 0.712 with the Facebook profile data as an input.

5 DISCUSSION AND FUTURE WORK

5.1 Comparing personality prediction with related literature

Predicting **privacy** measures has so far, to the best of our knowledge, not been evaluated previously. The most recent and most similar publication in this area was published by Farnadi et al. [7] in February 2017. The goal of the article was to predict the **personality** using only language features acquired from Facebook, Twitter and YouTube. In contrast to our work, they had access to a large database of results, containing data of 3731 Facebook users, 404 YouTubers and 44 Twitter accounts. In general, the results indicate that our **privacy measures** prediction can be performed with a higher precision than **personality** prediction in related work: Although the authors did use the uncorrected R^2 that is not normalized over the number of features and that is therefore always greater than or equal to its corresponding *adjusted* R^2 that we used, the coefficients of determination (R^2 values) for our **privacy** measures prediction are mostly larger when using Facebook or Twitter language features for the prediction, signaling a higher goodness of fit for our model. The R^2 of our regression analysis always ranged between 51.9 and 74.8 when using Twitter or Facebook language features, compared to values between 2.56 and 17.78 in related literature. Also, the standard error for the prediction using Facebook is smaller compared to the related article (0.375 to 0.643 within our **privacy** prediction compared to 0.649 to 0.776 for the **personality** prediction in related work). The authors achieved a notably lower error for the prediction using Twitter language features, but as the authors themselves stated, this might be a result of the low number of samples in their study ($n = 44$). More specifically, they achieved a standard error about 0.152 to 0.214, compared to 0.381 to 0.619 within our study. However, although we achieved good results for predicting the IUIPC measures, the prediction for the coarse-grained westin privacy scales were not notably better than random. These observations go hand in hand with related literature, where the westin privacy index was found to be too coarse-grained to predict online privacy concerns and behavior [21].

5.2 Size of the training set

The goal of this paper and the user study was not to determine how precisely the prediction of privacy and personality measures can perform, if providers like Facebook or Twitter use millions of data sets of their users as training input. The goal was to get a first impression on the possible standard error of a prediction, and to find out which data sources can/should be used for each measure. With our data set, we were able to prove that *it is* possible to perform a prediction that is notably better than random and the prediction of **personality** measures in recent literature, and to show which data can be used. Nevertheless, the prediction *might* be more precise with a large data set, which we would like to elaborate in future work.

5.3 Applications of the derived privacy measures

As stated in the introduction, the goal of this work is to derive the privacy measures, according to the westin privacy index and the Internet Users' Information Privacy Concerns (IUIPC) questionnaire. These measures are meant to give a general overview on the privacy concerns of a user, rather than directly giving a specific hint on which Facebook privacy settings should be applied, or which audience should be selected for a new post. Nevertheless, related literature has shown that there are strong correlations between those privacy measures and the desired permission settings for mobile phone apps, that can be used to build two different types of permission setting wizards for automatically deriving the settings, or actively supporting the user while doing them manually [14]. The same has shown to be possible for deriving the correct audience of a Facebook post [13], or the privacy settings for the disclosement of customer data collected inside an *intelligent retail store* [15].

We therefore see our work only as the first part in privacy assistance systems, that automates the derivation of the general privacy measures. As related literature has shown, these universal privacy measures can then be used to derive specific privacy settings for various domains, like mobile app or intelligent retail privacy settings or selecting the Facebook audience.

5.4 Lessons learned

We compared three different data sources to evaluate which of them allows the most precise prediction of the privacy measures, according to the westin privacy index and the IUIPC privacy measures. In summary, it can be said that all observed data sets can be used to predict the IUIPC privacy measures of a user, but with differences in precision. The data items extracted from the Facebook user's "About" page allow us to predict the personality and privacy measures with an acceptable RMSE from 0.712 to 1.093. Nevertheless, if Facebook or Twitter posts and tweets are available, that kind of data source should be preferred, as the RMSE is clearly lower from 0.686 to 0.894 for privacy measures using Facebook posts. If the developer has to choose between Facebook and Twitter data, Facebook data should be used. The conservative Westin privacy can, due to our observations, not be predicted better than random within all three data sources.

5.5 Future work

We focused on the two biggest social networks, namely Facebook and Twitter, to gather the data and to perform the prediction. Nevertheless, there are several other platforms that are gaining importance in recent years, that are often neglected in research. For example, we did not try to use Youtube comments and posts to perform a prediction.

Although the study was focused on the theoretical aspect of predicting the privacy measures of a user, we would like to continue our work with a practical approach using machine learning and the techniques described in this paper to compute the privacy measures, and then propagating the values to one of the services mentioned in related work (e.g. for predicting privacy settings).

We designed our study to get a first impression on the prediction precision, and which data sets should be preferred. We cannot state how well such a prediction would work if the training set contained the data from *all* or a *large* subset of a social network like Facebook or Twitter. In a next step, we would like to first ship out a predictor as a Twitter or Facebook app, which would in the optimal case be integrated into the Facebook webpage for all users.

6 CONCLUSION

Privacy measures are gaining importance for recommender systems in several domains. Although recommender systems *can help* them to solve their tasks faster, users are *not willing* to invest the time in filling out the initial questionnaires once. Research has therefore searched for and found ways to infer the personality measures from social media content. Nevertheless, it remains unclear whether the same is possible for privacy measures. We performed a user study including more than 100 Facebook and Twitter profiles and a two-step analysis including a correlation, a regression and a machine learning analysis, to find out *how well* privacy measures can be predicted using social network data, and *which* data source should be preferred for the different variables. Our results have shown that *it is* possible to do a prediction that is notably better than random for the IUIPC measures, with a slightly better precision as for personality traits. We gave design guidelines for developers on which data should be collected for an analysis: In the best case language features should be used, although it is also possible to make use of profile information for a prediction.

REFERENCES

[1] Yoram Bachrach, Michal Kosinski, Thore Graepel, Pushmeet Kohli, and David Stillwell. 2012. Personality and Patterns of Facebook Usage. In *Proceedings of the 4th Annual ACM Web Science Conference (WebSci '12)*. ACM, New York, NY, USA, 24–32. https://doi.org/10.1145/2380718.2380722

[2] Susan B. Barnes. 2006. A privacy paradox: Social networking in the United States. *First Monday* 11, 9 (2006).

[3] Tom Buchanan, Carina Paine, Adam N. Joinson, and Ulf-Dietrich Reips. 2007. Development of Measures of Online Privacy Concern and Protection for Use on the Internet. *Journal of the American Society for Information Science and Technology* 58, 2 (2007), 157–165. https://doi.org/10.1002/asi.20459

[4] Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling. 2011. Amazon's Mechanical Turk: A new Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6, 1 (2011), 3–5. https://doi.org/10.1177/1745691610393980

[5] Jilin Chen, Eben Haber, Ruogu Kang, Gary Hsieh, and Jalal Mahmud. 2015. Making Use of Derived Personality: The Case of Social Media Ad Targeting. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10508

[6] Lillian Clark and Levent Çallı. 2014. Personality types and Facebook advertising: An exploratory study. *Journal of Direct, Data and Digital Marketing Practice* 15, 4 (01 Apr 2014), 327–336. https://doi.org/10.1057/dddmp.2014.25

[7] Golnoosh Farnadi, Geetha Sitaraman, Shanu Sushmita, Fabio Celli, Michal Kosinski, David Stillwell, Sergio Davalos, Marie-Francine Moens, and Martine De Cock. 2016. Computational personality recognition in social media. *User Modeling and User-Adapted Interaction* 26, 2 (01 Jun 2016), 109–142. https://doi.org/10.1007/s11257-016-9171-0

[8] Bruce Ferwerda, Emily Yang, Markus Schedl, and Marko Tkalcic. 2015. Personality Traits Predict Music Taxonomy Preferences. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 2241–2246. https://doi.org/10.1145/2702613.2732754

[9] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. 2011. Predicting Personality from Twitter.. In *SocialCom/PASSAT*. IEEE, 149–156. http://dblp.uni-trier.de/db/conf/socialcom/socialcom2011.html#GolbeckRET11

[10] Ponnurangam Kumaraguru and Lorrie Faith Cranor. 2005. Privacy Indexes: A Survey of Westin's Studies. *ISRI Technical Report* (2005).

[11] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Info. Sys. Research* 15, 4 (Dec. 2004), 336–355. https://doi.org/10.1287/isre.1040.0032

[12] Nizar Omheni, Omar Mazhoud, Anis Kalboussi, and Ahmed HadjKacem. 2014. The Annotation: A Track of Reader's Personality Traits on Paper. In *Proceedings of the 2014 ACM Southeast Regional Conference (ACM SE '14)*. ACM, New York, NY, USA, Article 10, 6 pages. https://doi.org/10.1145/2638404.2638469

[13] Frederic Raber, Felix Kosmalla, and Antonio Krueger. 2017. Fine-Grained Privacy Setting Prediction Using a Privacy Attitude Questionnaire and Machine Learning. In *Human-Computer Interaction – INTERACT 2017*, Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler (Eds.). Springer International Publishing, Cham, 445–449.

[14] Frederic Raber and Antonio Krueger. 2017. Towards Understanding the Influence of Personality on Mobile App Permission Settings. In *Human-Computer Interaction – INTERACT 2017*, Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler (Eds.). Springer International Publishing, Cham, 62–82.

[15] Frederic Raber and Antonio Krueger. 2018. The 'Retailio' Privacy Wizard: Assisting Users with Privacy Settings for Intelligent Retail Stores.

[16] Alexandra Roshchina, John Cardiff, and Paolo Rosso. 2011. A Comparative Evaluation of Personality Estimation Algorithms for the Twin Recommender System. In *Proceedings of the 3rd International Workshop on Search and Mining User-generated Contents (SMUC '11)*. ACM, New York, NY, USA, 11–18. https://doi.org/10.1145/2065023.2065028

[17] H. Jeff Smith and Sandra J. Milberg. 1996. Information Privacy: Measuring Individuals' Concerns About Organizational Practices. *MIS Q.* 20, 2 (June 1996), 167–196. https://doi.org/10.2307/249477

[18] Antonela Tommasel, Alejandro Corbellini, Daniela Godoy, and Silvia Schiaffino. 2015. Exploring the role of personality traits in followee recommendation. *Online Information Review* 39, 6 (2015), 812–830. https://doi.org/10.1108/OIR-04-2015-0107

[19] R. Wald, T. Khoshgoftaar, and C. Sumner. 2012. Machine prediction of personality from Facebook profiles. In *2012 IEEE 13th International Conference on Information Reuse Integration (IRI)*. 109–115. https://doi.org/10.1109/IRI.2012.6302998

[20] Sara J. Weston, Patrick L. Hill, and Joshua J. Jackson. 2015. Personality Traits Predict the Onset of Disease. *Social Psychological and Personality Science* 6, 3 (2015), 309–317. https://doi.org/10.1177/1948550614553248

[21] A. Woodruff, V. Pihur, A. Acquisti, S. Consolvo, L. Schmidt, and L. Brandimarte. 2014. Would a Privacy Fundamentalist Sell their DNA for $1000... if Nothing Bad Happened Thereafter? A Study of the Westin Categories, Behavior Intentions, and Consequences. In *Proceedings of the Tenth Symposium on Usable Privacy and Security (SOUPS)*. ACM, ACM, New York, NY. https://www.usenix.org/conference/soups2014/proceedings/presentation/woodruff IAPP SOUPS Privacy Award Winner.