The “Retailio” Privacy Wizard: Assisting Users with Privacy Settings for Intelligent Retail Stores

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

David Ziemann
Saarland University
david.ziemann@gmail.com

Antonio Krueger
DFKI Saarland Informatics Campus
krueger@dfki.de

***Abstract*—Intelligent retail stores like Amazon Go collect and process a large amount of shoppers’ personal data to offer their service. In this paper we present *Retailio*, privacy management software that allows the customer to select the private data that should be accessible by retail stores. A privacy wizard helps the user to set her privacy settings, by using either a small informal privacy questionnaire or privacy measures extracted out of the user’s Facebook posts for a machine learning-based prediction of user-tailored privacy settings. We conducted an expert interview to determine the different types of data that could be recorded in intelligent retail stores, and performed a user study to find out whether their disclosures correlate with shoppers’ personalities. Retailio was evaluated in a validation study, regarding accuracy of the privacy wizard and user experience of the software. Our results show that there is a strong correlation between the IUIPC questionnaire and the data disclosure choice, which allowed us to predict the privacy settings with 70% accuracy.**

I. Introduction

Since the beginning of retail business, retail data privacy has always been a big issue. Although convential retail stores already collect a lot of information like products purchased, number of customers, sales amounts and much more, the data was at least collected anonymously. The retail companies rarely had a chance to match the sales data to the individual customers. Amazon as an online shopping platform, on the other hand, has the ability to match viewed, bought, sold and returned items to individual customer accounts. The recent launch of Amazon’s first brick-and mortar retail store, called “Amazon Go”, brought the topic of retail data to a new level of privacy nightmare. Where the customer could once rely on being anonymous during the shopping process, she is now tracked throughout the complete shopping journey: Upon entering the shop, the shopper uses the NFC functionality of his smartphone to identify himself at the entrance gate. She can then browse the store, grab products, put them back again, and just leave the store without going through a checkout process or scanning the products she decides to buy. Amazon achieves this using “sensor fusion and deep learning”[1] without naming further details. The technology behind the service is most likely based on camera systems and other sensors that follow the customer from the entrance gate throughout the shop, registering products being picked up, placed back or viewed, stopping points and most likely also the exact route throughout the store. Altough the Amazon Go service saves time and is very convenient, not all shoppers are happy with the new store concept: In order to make the service work, Amazon has to record and store a large amount of private data, throughout the shopping process, that is not even anonymized. The whereabouts of the data and what it is further used for remains as unclear as the description of technologies used and what data is recorded by them.

Apart from operating intelligent retail stores like Amazon Go, there exist several research laboratories like the Innovative Retail Laboratory (IRL) [20], which investigate the capabilities of new technologies in the context of brick-and-mortar retail stores. The Innovative Retail Laboratory (IRL) is an application-oriented research laboratory of the German Research Center for Artificial Intelligence (DFKI) run in collaboration with the German retailer GLOBUS SB-Warenhaus Holding in St. Wendel. In this living lab, they conduct research in a wide range of different domains, mostly related to intelligent shopping assistance. The demonstrators range from an instrumented shopping cart employing indoor navigation to several intelligent shopping consultants, ambient information services and an automated checkout system.

Apart from being a timely topic, there are several special issues that arise within the domain of intelligent retail data:

1) The data collected in an intelligent retail store is very diverse, from rather uncritical loyalty points, to viewed products, and finally the personal data or the movement patterns of the customer inside the store, making it hard to create a system that can predict optimal privacy settings for all data items (refer to Section III-A for a more detailed discussion and a study about this topic).
2) Apart from the diversity of the different privacy items, the types of data items are also very diverse (location data, personal data, shopping data etc.), making it hard to cluster and order them by ascending sensitivity for a clear presentation in a user interface (refer to Section III-A for a discussion).
3) So far, there is only few reliable information in the literature about which data is recorded for the different assistance systems in an intelligent retail store.

There exist several approaches in other domains like social networking or mobile app privacy which facilitate users expressing their privacy desires, and help them control which

[1] http://www.self.com/story/amazon-go-grocery-store-of-the-future

European Workshop on Usable Security (EuroUSEC) 2018
23 April 2018, London, England
ISBN 1-1891562-54-1
https://dx.doi.org/10.14722/eurousec.2018.23009

third-party persons or applications may use a specific part of the private data. Although websites like Facebook already offer a privacy settings page, these are still characterized by loads of technical settings which are hard to understand for lay users. Research in the past therefore concentrated on either visualizing the complex settings in a comprehensive way, or tried to simplify the problem by reducing either the technicality or the amount of settings that have to be done, or both. In this paper, we try to tackle the problem of data privacy in intelligent retail stores by reducing the amount of needed input. In detail, we try to facilitate correlations between the personality or privacy measures of a customer and his desired retail privacy settings, in order to create a privacy wizard that is able to predict the appropriate individual privacy settings for a customer, based on personality and privacy measures using Facebook posts or a small set of questions of an easy-to-understand privacy questionnaire as a prediction input. To the best of our knowledge, we are the first to address this problem in the domain of intelligent retail data.

In detail, we try to solve the following research questions:

1) What data is collected in current or yet-to-be-built intelligent retail stores?
2) Is there a correlation between personality or privacy attitudes and customers' data disclosure preferences for an intelligent retail store?
3) Can the correlation be used to predict the data disclosure preferences using a short personality questionnaire like the IUIPC or TIPI?
4) Are customers interested in having control over their own recorded data in an intelligent retail store, or do they trust in retail companies?
5) Do customers accept a privacy UI, which helps them to monitor and tune their privacy settings for the disclosure of their private data in intelligent retail stores?

In some domains like social networks or mobile app permissions, researchers found a correlation between a user's personality, captured by the big five personality measures[5], and their privacy and posting behavior. For example on Facebook [1], *Extraverted* users have more friends, and post more statuses and likes on Facebook. Similar results could be observed for *openness*. In contrast, more conscientious subjects are less likely to "like" a post or be a member of a large number of groups. There is also a correlation between personality and mobile apps that are chosen by users, and conversely it is possible to derive the personality of a user given the installed apps on her smartphone [21].

In a first step, we used expert information from members of the Innovative Retail Lab to create a list of privacy-sensitive data that is used inside an intelligent retail store. We then conducted a larger user study including 100 participants to first check for correlations, and then to train a machine learning component doing a prediction of these.

Although the main focus of our work is on the prediction of data disclosement preferences, we present a user interface which helps the user to set his privacy settings for retail data in a centralized system. Machine learning is utilized to help the user to find his optimal settings in a *privacy wizard*, by taking the user's privacy measures as a basis for the prediction. The results of the user interface evaluation show that the UI including machine learning support is perceived as more comfortable and is significantly preferred to a standard UI without machine learning.

II. Related Work

There are two topics that are of major importance for our work: first, publications about existing personality and privacy questionnaires, either capturing a general personality or particular privacy desires concerning online companies, that can be used as a basis for the prediction. Second, we are interested in prior work that tries to implement such a privacy wizard that allows the privacy settings to be set automatically based on answers to a simple questionnaire. As there is no prior work on predicting retail privacy settings, the major related work presented here concentrates on privacy prediction in other domains, like location or mobile app privacy, and social networks.

1) Privacy and Personality Questionnaires: Privacy has always been an important research topic in the past, and since the beginning, researchers have tried to capture privacy desires using questionnaires. One of the earliest publications was actually contributed in the field of consumer privacy indices by Alan Westin[10]. Westin categorized consumers into three different categories: The *Unconcerned* hardly care about their privacy and tend to publish all information to the entire audience of a network. *Fundamentalists*, in contrast, try to disclose as little information as possible in order to preserve their privacy. The third group of persons, the *Pragmatists*, attempt to keep a balance between privacy and usability: Pragmatists believe that privacy is an important aspect, but on the other hand accept the necessity to share information in order to benefit, for example, from an additional app feature.

Other researchers continued with the idea of a general privacy questionnaire, and introduced the PCS[2] questionnaire [2] in 2007. The PCS is more detailed and consists of 28 questions in three categories: General Caution, Technical Protection and Privacy Concern. Despite the advance in granularity, the questionnaire still adresses the general privacy attitudes of a person, not the specific context of retail privacy.

In contrast, the CFIP[3] [19], and based on that, the IUIPC [14] questionnaire, were designed explicitly to measure the privacy concerns of internet users, especially in the context of online shopping companies and their data collection. The authors found that privacy attitudes regarding online companies can be expressed well using three privacy measures: the *control* measure, which determines how far a subject desires to have control over the disclosure and transfer of her personal information; the desired *awareness* of how and to whom the personal information is disclosed; and *collection*, describing how important it is for the subject to know which personal data is collected. As the IUIPC is the privacy questionnaire which best fits the goals of our paper, we used it in the survey of the main study.

The big five personal inventory [5] is currently the most widely accepted questionnaire for capturing a person's personality. The big five is a questionnaire that derives five

[2] Privacy Concern Scale

[3] Scale of Concern For Information Privacy

personality measures: *Openness to experience*, denoting general appreciation for art, emotion, adventure etc.; *Conscientiousness*, meaning the tendency to show self-discipline; *Extraversion*, meaning higher or lower social engagement; *Agreeableness* in terms of cooperation with other people and *Neuroticism* as the tendency to experience negative emotions. As the questionnaire in its original version is very long and requires up to 30 to 40 minutes for completion, we are using a shorter version to capture the big five personality traits, consisting of only ten questions [7]. The "big five" of personality can also be extracted out of written text, e.g. blog or social network entries [4]. In a paper currently under review, researchers have shown that the same is also possible for the IUIPC privacy measures. The user burden for gathering the big five personality and IUIPC privacy measures can therefore be reduced to a minimum. As stated in the introduction, there is evidence that personality correlates with the Facebook sharing behavior. We also expect some effects on the retail privacy permissions, and therefore included the TIPI questionnaire in our study.

2) Permission prediction techniques: Creating and maintaining privacy settings that reflect an optimal trade-off is very labor-intensive for the user. There have been several approaches to overcome this user burden by automatizing privacy setting maintenance. One of the most commonly used techniques is to use machine learning and a subset of labeled friends to predict the privacy settings of the remaining users [17], [18], [6]. Fang and LeFevre [6] proposed a semi-supervised machine learning technique to infer privacy settings of a user's social network (SN) friends: The user is asked to label several of her friends on the SN with privacy privileges. The decision on how many and which friends have to be labeled is made by their algorithm. After this annotation phase, the software predicts the privacy privileges for the remaining, unlabeled friends.

Ravichandran et al. [16] propose the use of privacy templates for each user, in the context of location sharing with mobile apps. They observed 30 users using a mobile phone app and asked them to annotate their privacy desires regarding location sharing (share location/do not share location) whenever they changed their context, e.g. when they came home from work. The app recorded the time when a context change appeared, as well as the corresponding privacy desires. Using decision trees and clustering techniques, they created several privacy profile templates. Their experiment showed that with only three templates, the preferences of a user are matched with 90% accuracy.

There are several publications describing the prediction of mobile application settings using different data sources for the prediction. Other approaches use machine learning to predict the settings [13], [12], [11]. Ismail et al. [9] describe an approach which facilitates crowdsourcing in order to find an optimal tradeoff between denied permissions and usability of the app, tailored to an individual user. Liu et al. [13] use a large online database of the LBE Privacy Guard app, containing the app settings of 4.8 million users, as training data for their prediction using a linear support vector machine. 90% of the user records are used for training, 10% for testing the accuracy of the prediction. When it comes to prediction, the system uses 20% of the app settings of a user to predict the remaining 80% of settings. They used only permissions, the user ID and the app ID for the prediction to achieve a precision score of 64.28% to 87.8%, depending on the features used. Privacy or personality attitudes were not taken into account.

To conclude the related work on permission prediction, there have been several approaches in other domains using crowdsourcing or machine learning techniques like clustering, based on the permission settings themselves or using comfort with the purpose of a permission. The effect of personality on the choice of retail privacy settings has, to the best of our knowledge, not been explored so far. In the next sections, we will iteratively develop a user interface for retail privacy settings, including a privacy wizard that creates an individual initial privacy profile based on the answers to a short personality questionnaire.

Unlike other related work [12], our approach does not need any knowledge about previous smartphone usage behavior, and can therefore be seen as a first step towards solving the cold start problem in this scenario.

III. User study

The main study consisted of two different stages: First we had to gather background knowledge about data usage and privacy issues in intelligent retail stores. Afterwards, we conducted an online user study to discover correlations between the personality or privacy attitudes of a person and the privacy settings for the aforementioned data. As discussed in the last section, broader questionnaires like Westin's categories or the CFIP lead to suboptimal results. Therefore, we used the more specific IUIPC questionnaire in order to capture the privacy concerns of the subjects. Personality was captured using the big five personality measure [5], more specifically the abbreviated Ten Item Personality score (TIPI) [7], which is a compressed version of the big five scale using only ten questions in total. Although it would have been possible, we did not extract the personality measures but instead used the TIPI for this study, to reduce any possible side-effects caused by the derivation of the measures. In addition to these two questionnaires, we posed two additional questions regarding privacy and privacy invasion (see Table III). In detail, we asked the subjects how frequently they had been a target of a privacy invasion (on a five point ordinal scale from very frequently to never), and how often they enter wrong information on purpose on websites (percentage as a numeric scale). The two stages of the study are described in the next two subsections.

A. Background analysis

As stated in the introduction, there are several special issues in the domain of intelligent retail data: There is no reliable information available on what data is recorded inside an intelligent retail store, and how the data can be clustered for a clear user interface design. Furthermore, the data is highly diverse regarding both type of data, as well as the sensitivity. Using a conventional list-based privacy UI like Facebook for this kind of data, where all data items are listed on top of each other, would most likely lack a clear overview. Besides of that, the attention and motivation of a user for doing privacy settings is very limited; the chance of missing a privacy setting for a highly sensitive data item is therefore high in such a UI. We therefore decided to use a card-based UI (see Section V) that clusters the data into groups of data, and that sorts them with

Variable	Description
Address	Personal information of the customer
Birthday	
Name	
(Household) income	
Nutrition	Nutrition/product preferences like vegan/vegetarian, likes fish, dislikes meat
Allergies	Customer's allergies
Recent visits	Date, time and place of the last shop visits of the customer
Wishlist	Bookmarked items/items on the customer's shopping list
Recently viewed	Items that have been recently viewed by the customer, e.g. taken from the shelf and put back
Receipt	Detailed shopping receipt, including the products bought with their exact names and product IDs
Category	The categories of the products bought, e.g. "vegetables" or "cereals"
Amount	The amount of products bought
Price	The price of each of the products bought
Loyalty	Loyalty points
Location	In-store location and movement pattern of the customer

TABLE I. PRIVATE DATA THAT IS RECORDED IN AN INTELLIGENT RETAIL STORE.

Service	Data	used by	
		IRL	Amazon
"Invisible" Checkout	- Address - Birthday - Name - Recent visits - Recently viewed - Receipt - Category - Amount - Price - Loyalty	X	X
Digital shopping list	- Wishlist	X	
Customer heatmap/ customer flow for market manager	- Location	X	?
Allergy advisor	- Allergies	X	
Product recommender	- Nutrition - Income	X	?

TABLE II. AMAZON GO AND IRL SERVICES AND PRIVATE DATA USED.

Label	Question
Falsify	Some websites ask you for personal information. When asked for such information, what percent of the time would you falsify the information?
Invasion	Have you ever been the target of a privacy invasion (e.g. your data was misused or shared without your knowledge)?

TABLE III. QUESTION TEXT AND LABEL OF THE ADDITIONAL QUESTION SET.

descending sensitivity to give a clear overview on the different data types, and to draw attention to the most sensitive data items first. As there is hardly any information about data types, clusters and sensitivity orders for intelligent retail data so far, we conducted two experiments prior to the main user study based on the results by Raber et al. [15].

First, we conducted an expert interview with an employee of the Innovative Retail Laboratory [20], to find out what data is gathered inside the IRL/Amazon Go and could be recorded in other intelligent retail stores, to create a list of privacy-sensitive data, later called *permissions* or *items* within the *retail privacy settings*. In a following pre-study, we asked a small group of participants to cluster and rank the items discovered in the expert interview. This interview and the pre-study will be described in the next two subsections.

1) Expert interview: Data collected inside an intelligent retail store: Prior to the expert interview, we brought together information of the official website of the Innovative Retail Laboratory [4], where the services of this future retail store are described. The collected information was then validated and extended (with data types that are recorded for each service) in the expert interview. The interview partner of the IRL created a list in advance of the data that was collected by the services either by doing a code review or by asking the corresponding colleague that implemented the service. On the day of the interview, the results were then discussed with the authors.

Table I contains a list of observed private retail data items together with a short description, whereas Table II shows a list of services that are present in the IRL or Amazon Go, as well as the data that is recorded or required for the service to work.

Most data is recorded for the "invisible checkout", which allows the customer to just grab products out of the shelves, and to leave the store without the need to scan and pay for the products at a checkout. Amazon Go uses "sensor fusion and deep learning", whereas the IRL relies on RFID tags inside the products, to find out which products have been placed inside the shopping cart. In addition to the viewed and bought products, the IRL also keeps track of the shoppers' route inside the store, including visited areas and stopping points. The IRL uses a Bluetooth location system called Quuppa[5] for this purpose. The data allows generating heatmaps for a "management dashboard", which allows the store manager to optimize the store layout, for example. Amazon is rather unspecific about the technology used as well as the data recorded. Nevertheless, the optical systems in the store would be capable of tracking customers' movement inside the store. Whether the data is actually stored and evaluated remains unclear. The Innovative Retail Lab offers several recommender systems to the customer, which recommend products that fit with the other products inside the shopping basket, or that match the client's typical product set. In addition, it is possible to highlight allergy information on the products inside the store. For this purpose, nutrition preferences and allergy information about the customer are stored.

2) Pre-study: Clustering and order of the data: The pilot study was conducted with five participants recruited from the university context. All of them were students aged between 21 and 48 (average 38). The study was done using a questionnaire, which was constructed as follows: In the first question, the participants were given the list of retail data types along with a set of category names (app data, personal profile, location data, sales receipt data, interests). The participants then were asked to either assign the data types to a group, or to create a new group. As our list might not be exhaustive, we asked whether there were other types of data that might be recorded that came to a participant's mind, and which data types were hard to assign to a specific group. The next question asked for the sensitivity of the different data types on a five-point scale from "I would never disclose this data" (=5) to "I would disclose this data without any concerns" (=1).

All proposed clusters were used, except for the "app data"

[4] http://www.innovative-retail.de

[5] http://quuppa.com/

		P1	P2	P3	P4	P5	Rank
Personal Data	*Address*	3	4	3	4	4	
	Birthday	3	1	2	3	2	
	Name	3	2	2	4	4	
	Income	5	3	3	4	3	
	Allergies	1	2	3	4	3	
	Nutrition preferences	3	3	3	4	4	
Location data	*Recent visits*	1	1	2	3	2	1
	Movement	2	1	3	4	2	2
Shopping Receipt	*Loyalty points*	1	2	3	3	2	1
	Items bought						
	- Price	3	2	3	3	2	2
	- Category	3	3	3	3	2	3
	- Amount	4	3	3	2	3	4
	Receipt	4	3	4	3	3	5
Interests	*Wishlist*	1	3	3	3	2	1
	Recently viewed	2	3	3	3	2	2

TABLE IV. SENSITIVITY RANKINGS AND RANK ACCORDING TO *data groups*.

Item	mean	stdev	% denied
Loyalty	4.37	1.47	15.8
Category	3.00	1.71	21.8
Amount	4.35	1.33	22.8
Nutrition	4.37	1.47	23.8
Price	2.97	1.59	24.8
Name	3.96	1.48	29.7
Receipt	3.79	1.66	31.7
Birthday	3.79	1.66	36.6
Recent visits	3.68	1.49	39.6
Allergies	3.74	1.70	40.6
Wishlist	3.78	1.62	41.6
Recently viewed	3.96	1.48	54.5
Address	3.00	1.71	60.4
Income	2.97	1.59	60.4
In-store movement	3.74	1.70	66.3

TABLE V. MEAN AND STANDARD DEVIATION FOR THE SHARING LIKELIHOOD ON OUR 5-POINT SCALE (1= VERY UNLIKELY, 5=VERY LIKELY) AND PERCENTAGES OF DENIES FOR EACH RETAIL PRIVACY SETTING.

cluster, which was perceived as too vague by most of the participants. They agreed to assign the "app data" to the data group according to the *type* of data, e.g. whether it is location data or related to the sales receipt. Apart from that, they had no problems assigning the items to the proposed group, and did not feel the need to create new groups. The clusters and severity ratings for the different data types are shown in Table IV.

Except for personal data, we were able to bring the data types inside a cluster into an ascending order regarding the reported sensitivity. The sensitivity for personal data was too varied to find a meaningful order that worked for all participants. The clusters and data orders discovered in the study have been used to group the data and order them with descending sensitivity in the user interface, as described in Section V.

B. Online study

Based on the results of the expert interview, we were able to design an online study to check for correlations between privacy attitude and data disclosement behavior. The study was conducted as an online survey using the software LimeSurvey[6]. 100 participants were recruited using Prolific Academic[7]. Studies in the past have shown that participants who are recruited via online services, like in our case, lead to a similar quality of results as when participants are recruited at a university [3]. The participants were paid a compensation of £2 upon successful participation. To motivate the subjects to fill out the questionnaire honestly, the compensation was only paid after the submitted data was checked for plausibility by us. If the result from a subject was rejected, for example if she failed to answer the control questions correctly, a new participant was recruited to fill in the gap. Therefore we have exactly 100 viable results. The age of the participants ranged from 18 to 73 years (average 33, SD 11.7). We had 46 female and 54 male participants. The recruited audience was very diverse: We recruited students, self-employed workers, employees, and also homemakers.

For the study, we gave the users an overview on the Innovative Retail Laboratory, and asked users to rate whether or not they would disclose their data in a shop like the Innovative Retail Laboratory, which is hosted by a well-known local retailer called *Globus*. The survey can be divided into two parts: In the first part, we asked the subjects to fill out the above described privacy and personality questionnaires. In the second phase, we asked, for each item of the *Retail privacy settings*, how likely she will refuse to disclose the item in the context of the IRL. We used a six-point scale (1 = very unlikely, 6=very likely) instead of a 5-point scale so the participants have to decide whether they rather disclose (score >= 4) or undisclose (score <=3) the data item. The survey ended with a short feedback question in free-text style.

C. Results

The 100 participants filled out 100 *retail privacy settings*, whose mean and standard deviation can be found in Table V together with the frequency of denied permissions (sharing likelihood < 3).

According to the shape of our data (mostly ordinal values, not normal-distributed), we decided to use a non-parametric test, and therefore performed a Spearman correlation ("Spearman's Rho") on the results of the questionnaire and the retail privacy settings. The results are shown in Table 1.

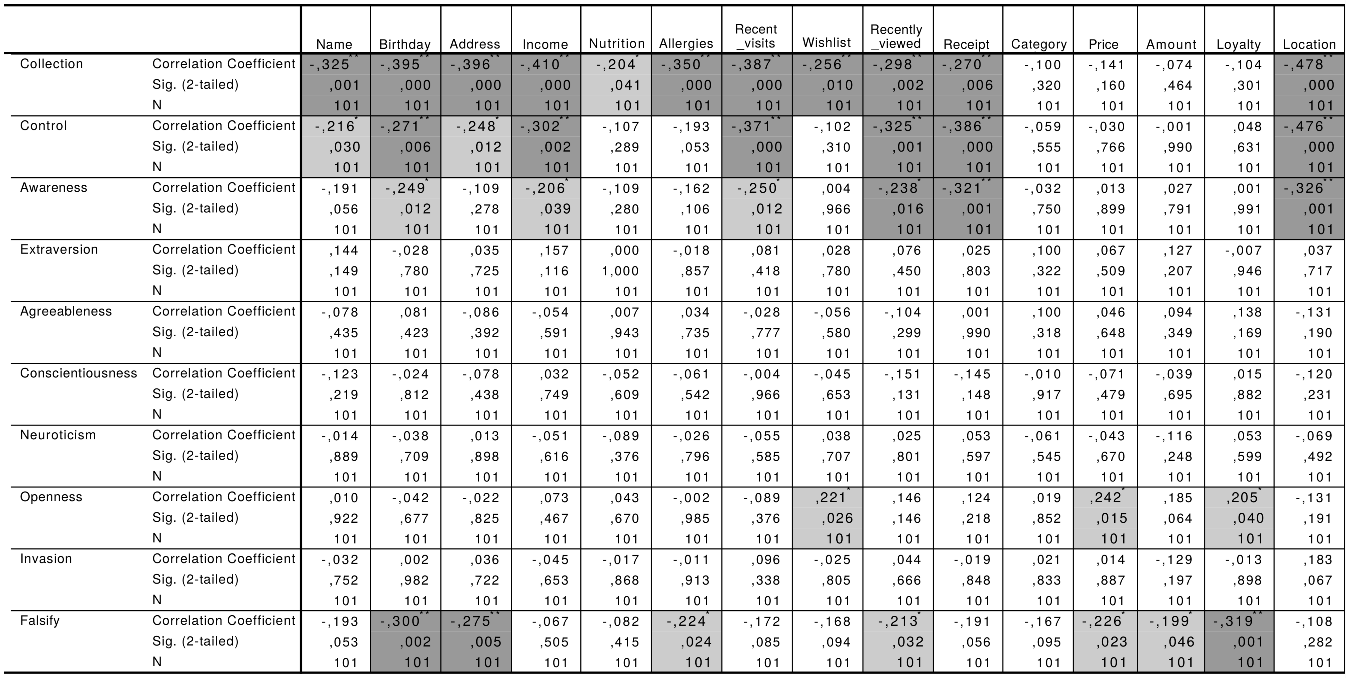
The measures of the privacy and personality questionnaires are in the rows, whereas the retail privacy settings are plotted as the columns of the table. Significant and highly significant correlations are marked with one or two asterisks, and colored in gray or dark gray, respectively. Regarding the IUIPC measures (collection, control, awareness), the collection measure yields highly significant correlations for most of the permissions (10 out of 16). Control and awareness both also correlate with several permissions. The general personality seems not to correlate with the choice of retail privacy settings and is therefore unsuitable for a machine-learning based prediction. However, the amount of falsified information given to online companies seems to correlate significantly or highly significantly with seven out of 14 items of the privacy settings. We therefore dropped the TIPI questionnaire and continued to work with the IUIPC and our additional questionnaire for the prediction in the next sections.

IV. RETAIL PRIVACY SETTING PREDICTION

Based on the results of the user study, we decided to use the results as training data to predict the retail privacy settings, based on privacy measures. Based on our data (predicting a

[6] https://www.limesurvey.org, last accessed 09-05-2016

[7] https://www.prolific.ac/, last accessed 09-05-2016



		Name	Birthday	Address	Income	Nutrition	Allergies	Recent _visits	Wishlist	Recently _viewed	Receipt	Category	Price	Amount	Loyalty	Location
Collection	Correlation Coefficient	-,325**	-,395**	-,396**	-,410**	-,204*	-,350**	-,387**	-,256**	-,298**	-,270**	-,100	-,141	-,074	-,104	-,478**
	Sig. (2-tailed)	,001	,000	,000	,000	,041	,000	,000	,010	,002	,006	,320	,160	,464	,301	,000
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
Control	Correlation Coefficient	-,216*	-,271**	-,248*	-,302**	-,107	-,193	-,371**	-,102	-,325**	-,386**	-,059	-,030	-,001	,048	-,476**
	Sig. (2-tailed)	,030	,006	,012	,002	,289	,053	,000	,310	,001	,000	,555	,766	,990	,631	,000
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
Awareness	Correlation Coefficient	-,191	-,249*	-,109	-,206*	-,109	-,162	-,250*	,004	-,238*	-,321**	-,032	,013	,027	,001	-,326**
	Sig. (2-tailed)	,056	,012	,278	,039	,280	,106	,012	,966	,016	,001	,750	,899	,791	,991	,001
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
Extraversion	Correlation Coefficient	,144	-,028	,035	,157	,000	-,018	,081	,028	,076	,025	,100	,067	,127	-,007	,037
	Sig. (2-tailed)	,149	,780	,725	,116	1,000	,857	,418	,780	,450	,803	,322	,509	,207	,946	,717
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
Agreeableness	Correlation Coefficient	-,078	,081	-,086	-,054	,007	,034	-,028	-,056	-,104	,001	,100	,046	,094	,138	-,131
	Sig. (2-tailed)	,435	,423	,392	,591	,943	,735	,777	,580	,299	,990	,318	,648	,349	,169	,190
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
Conscientiousness	Correlation Coefficient	-,123	-,024	-,078	,032	-,052	-,061	-,004	-,045	-,151	-,145	-,010	-,071	-,039	,015	-,120
	Sig. (2-tailed)	,219	,812	,438	,749	,609	,542	,966	,653	,131	,148	,917	,479	,695	,882	,231
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
Neuroticism	Correlation Coefficient	-,014	-,038	,013	-,051	-,089	-,026	-,055	,038	,025	,053	-,061	-,043	-,116	,053	-,069
	Sig. (2-tailed)	,889	,709	,898	,616	,376	,796	,585	,707	,801	,597	,545	,670	,248	,599	,492
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
Openness	Correlation Coefficient	,010	-,042	-,022	,073	,043	-,002	-,089	,221*	,146	,124	,019	,242*	,185	,205*	-,131
	Sig. (2-tailed)	,922	,677	,825	,467	,670	,985	,376	,026	,146	,218	,852	,015	,064	,040	,191
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
Invasion	Correlation Coefficient	-,032	,002	,036	-,045	-,017	-,011	,096	-,025	,044	-,019	,021	,014	-,129	-,013	,183
	Sig. (2-tailed)	,752	,982	,722	,653	,868	,913	,338	,805	,666	,848	,833	,887	,197	,898	,067
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101
Falsify	Correlation Coefficient	-,193	-,300**	-,275**	-,067	-,082	-,224*	-,172	-,168	-,213*	-,191	-,167	-,226*	-,199*	-,319**	-,108
	Sig. (2-tailed)	,053	,002	,005	,505	,415	,024	,085	,094	,032	,056	,095	,023	,046	,001	,282
	N	101	101	101	101	101	101	101	101	101	101	101	101	101	101	101

Fig. 1. Correlations between privacy awareness/personality measures and retail privacy settings.

continous privacy measure using numerical input), we chose to use a regression algorithm for this task. Just like the answers to our six-point sharing likelihood scale, the predicted values are in the interval [1,6]. We therefore mapped the prediction results halfway to allow (result > 3.5) or deny (result < 3.5).

Similar publications predicting privacy settings in other domains [13] used a support vector algorithm for their prediction. We also tried out SVR and several other regression methods, and achieved the best results with a ridge regression, using the scikit-learn[8] machine learning library.

As described in the introduction section, one of the major special issues with intelligent retail data is the high diversity of the data items; therefore we had to find a way to train the machine learning algorithm differently from what is usual: Normally, the prediction algorithm (in our case a ridge regressor) is trained by passing the input features (type of permission and IUIPC measures of the subject) as well as the goal features (retail privacy setting for the given permission) for each user and permission. That means the study data is used to train *exactly one* prediction algorithm; the type of the permission is included as an additional feature inside the feature set, using a dummy variable for example. When a prediction of a permission has to be conducted, again the input features (type of permission and IUIPC measures of the subject) are passed to the regressor, which returns a prediction of the privacy setting. As the findings of our pre-study indicate, the sensitivity of our data is very diverse; using the whole data for one regressor would therefore not lead to optimal accuracy. We therefore used what we call a "compound regressor" which works as follows: The core of the compound regressor consists of n ridge regressors, where n is the number of permissions. Whenever the set of input features (type of permission and IUIPC measures of the subject) and goal features is passed to the compound regressor, it retrieves the ridge regressor for the permission to be trained, and applies the goal features and input features *excluding the type of permission* to the regressor. Accordingly, if a prediction with input features (type of permission and IUIPC measures) has to be done, the compound regressor retrieves the corresponding regressor according to the permission, applies the input features excluding the type of permission, and returns the result.

We followed the usual way of training, adjusting parameters, and validating the prediction of a machine learning algorithm. In order to prevent biasing of the results, we used a cross-validation method called *repeated random sub-sampling validation*, also known as *Monte Carlo cross-validation*: The data set is split into two basic parts. The first part is called the *training set*, and is composed of 75% of the data set. It is used to train and to calibrate the prediction algorithm, and select the optimal features. The second and remaining part is called the *test set*, and is used solely for the evaluation of the results later. **The data in the *test set* is never used while setting up the algorithm, neither for training/fitting, nor selecting optimal algorithm parameters, nor for finding the optimal feature set**. We performed 100 distinct runs, and used the average precision of all runs for selecting the best set of features. After each run, the data set was shuffled randomly, and its items were reassigned to one of the two subsets.

1) Validation: As stated above, we kept the data of the *test set* untouched in order to perform a validation later, which is described in this subsection. We validated the results of the prediction using the trained estimators as described in the last subsection. Neither the estimators nor the input features were changed throughout the validation. Afterwards, the feature

[8] http://scikit-learn.org

Item	Probabilistic	IUIPC	Additional
All	**57.7**	**69.1**	**67.6**
Name	60.0	73.6	71.8
Birthday	55.4	60.0	56.4
Address	48.1	61.8	59.1
Income	50.9	59.1	54.5
Nutrition	53.6	63.6	53.6
Allergies	67.2	81.8	81.8
Recent visits	61.8	78.2	78.2
Wishlist	52.7	64.6	65.5
Recently viewed	51.8	52.7	53.6
Receipt	49.0	60.9	60
Category	54.5	68.2	68.2
Price	50.9	59.1	59.1
Amount	80.0	87.3	87.3
Loyalty	60.9	78.2	77.3
Location	62.7	75.5	74.5

TABLE VI. PREDICTION ACCURACY (IN PERCENT OF CORRECT PREDICTIONS) FOR THE PREDICTION WITH THE PROBABILISTIC MODEL (PROBABILISTIC) AND PREDICTION USING THE IUIPC QUESTIONNAIRE, OR OUR ADDITIONAL QUESTIONS.

values from the *test set* are used to predict the permission settings, and compared with the actual permission settings of the *test set*. Again this procedure was repeated 100 times, and the results were averaged.

In order to get an impression of the quality of the results, we implemented a naive approach to predict the settings, which will later be called the *baseline* or *random* condition. We started with a simple random method, which randomly predicts "allow" or "deny" for each of the settings, giving a 50% accuracy. Since the percentage of allow and deny differs from setting to setting and is rarely 50% for both (see Table V), we enhanced the random approach by a probabilistic component: We first use the *training set* to calculate the probability of getting allowed or denied for each permission respectively. Based on these probabilities, we then predict the permission settings on the *test set*. If for example a setting for the item *Receipt* has to be chosen, the prediction will decide to "allow" with a probability of 63.7%, and to "deny" in 31.7% of all cases.

As before, 100 runs were conducted to evaluate the probabilistic random method, and the results were averaged. The percentage of correct predictions of this probabilistic *Random* approach, as well as the correctness using only the IUIPC or the additional questions as features, is shown in Table VI. The columns denote the feature sets, whereas the rows contain the different app permissions. The topmost row ("all") denotes the average percentage over all permissions.

Although the probabilistic approach achieves better results ($M = 57.7$) than a pure random method, it is still outperformed by the machine learning-based prediction ($M_{IUIPC} = 69.1$, $M_{Additional} = 67.6$). The prediction based on the IUIPC questionnaire (12 questions) performs best, although good results can also be achieved using our additional questionnaire (two questions). Best results can be achieved for the *Amount* permission ($M_{IUIPC} = 87.3$, $M_{Additional} = 87.3$). The disclosement setting for the *recently viewed* permission was hardest to predict ($M_{IUIPC} = 52.7$, $M_{Additional} = 53.6$). Overall, the machine learning approach outperformed the probabilistic method by about 11%.

As the results seemed promising, we created a user interface called *"Retailio"* that implements our approach. The next section will give details about the UI as well as a final evaluation study.

V. "RETAILIO" PRIVACY SETTINGS UI

In our opinion, a good privacy system always consists of two parts: A privacy UI that allows the user to have a clear overview on the settings and allows easy modifications thereof, as well as a backend that helps him choose the privacy settings, for example by automatically adjusting them based on privacy measures using a privacy wizard. As a prediction can never be 100% correct, it is very crucial to let the user have an overview of the settings and the possibility of fixing wrong predictions in the user interface. As the attention of a user is always limited, we clustered the data items into four different categories with descending sensitivity order (according to our pre-study), to make sure the most crucial sharing settings are seen and checked first by the user. The UI offers a privacy wizard (based on the machine learning algorithm presented in the last section) to set the privacy settings automatically, based on privacy measures than can either be automatically derived out of the user's Facebook or Twitter account[4], or captured directly using a short, non-technical questionnaire that can also be filled out by non-experts.

The detailed workflow of the UI is denoted in Figure 2. When the customer accesses the website for the first time after registration, she is offered a privacy wizard (see Figure 3, upper left) which asks the user to connect to Facebook or to answer the 12 questions of the IUIPC questionnaire.

After the survey is finished (typically 2-3 minutes), the customer is presented the results (mean scores) of the questionnaire along with the typical mean scores of other customers (Figure 3, lower right). When clicking on "calculate settings now", the wizard uses the ridge regression estimators (see Section IV) to predict the privacy settings tailored to the customer. From that point on, Retailio is initially set up and ready to use. The interface of Retailio (Figure 4) consists of four "index card" boxes that contain the data items of the four different clusters, which are based on the findings of the pre-study. Except for the "Personal Data" cluster, where no general sensitivity order could be found, all data items are sorted by descending sensitivity, to draw the focus to the most sensitive data items first. Shared data items are highlighted in blue, whereas undisclosed items are colored in grey. By enabling editing first by clicking on the pencil button in the respective card view, the user is able to change the setting to disclose/undisclose by a single click. Using the button at the top of the screen, the user can re-run the privacy wizard. With a click on the "show recent usage" button, a table is displayed showing which intelligent retail service (see Table II) accessed a permission in the past, together with the name of the permission as well as a timestamp of the access.

VI. EVALUATION

In order to evaluate the final design of *Retailio*, we performed a lab study with 24 participants from the university context, including students but also employees from different faculties, researchers, and employees of a research institude at the university campus. We had exactly 12 male and 12 female participants with a mean age of 28.2 years (min=21, max=54). 15 of the 24 participants were students: six computer

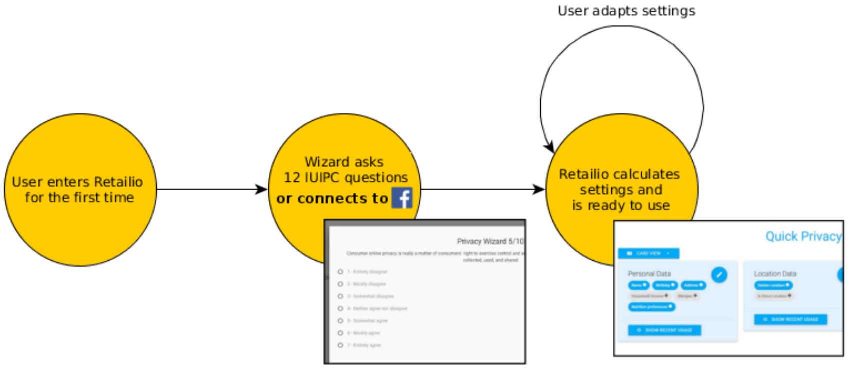
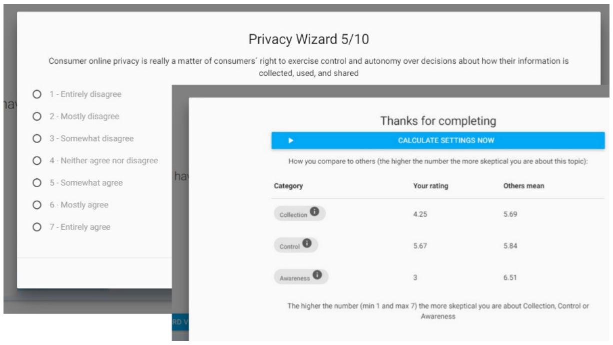

User adapts settings
User enters Retailio for the first time
Wizard asks 12 IUIPC questions or connects to
Retailio calculates settings and is ready to use
Privacy Wizard 5/10
Quick Privacy

Fig. 2. Typical workflow of Retailio during first use.



Privacy Wizard 5/10

Consumer online privacy is really a matter of consumers´ right to exercise control and autonomy over decisions about how their information is collected, used, and shared

○ 1 - Entirely disagree
○ 2 - Mostly disagree
○ 3 - Somewhat disagree
○ 4 - Neither agree nor disagree
○ 5 - Somewhat agree
○ 6 - Mostly agree
○ 7 - Entirely agree

Thanks for completing

How you compare to others (the higher the number the more skeptical you are about this topic):

Category	Your rating	Others mean
Collection	4.25	5.69
Control	5.67	5.84
Awareness	3	6.51

The higher the number (min 1 and max 7) the more skeptical you are about Collection, Control or Awareness

Fig. 3. Retailio's privacy wizard: IUIPC questions (upper left) and results page (lower right).

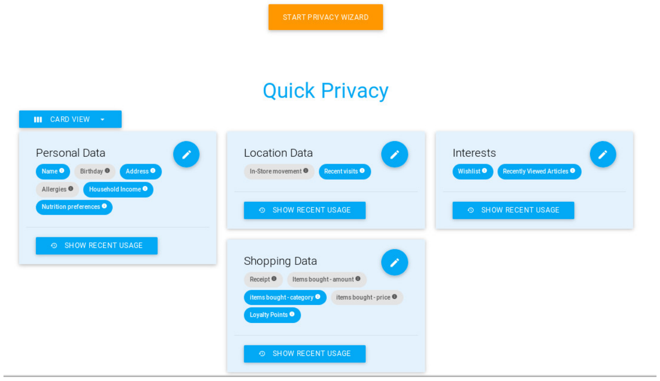


Fig. 4. Main screen of Retailio.

scientists, five in business administration, and the remaining four in other disciplines like law. The other participants were mostly employed for wages. We also asked about their frequency of shopping in general and online shopping, and excluded subjects that do online shopping less frequently than "several times a year". 12% of the participants do online shopping several times a week, 47% several times a month, and 41% several times a year. Shopping in general is performed several times a week by 77% of the subjects; the remaining 18% do it several times a month.

As stated in the introduction, there is currently no system that offers the customer a user interface to set his retail privacy settings. Therefore we could only evaluate the attractiveness of the Retailio UI as it is; there is no baseline interface that we could use as a comparison. Nevertheless, we were able to evaluate the prediction of the privacy wizard as a comparative study against the probabilistic approach as the baseline. To avoid side-effects and concentrate on the precision of our privacy setting prediction, we let the participants fill out the privacy questionnaire in the privacy wizard instead of connecting to Facebook and automatically extracting the data.

The goal of the evaluation was to check in a realistic scenario:

- Whether the idea of using a privacy wizard is accepted by users
- If the predicted settings of the privacy wizard are useful
- How well Retailio is rated in terms of user experience and performance
- If there are still some points for improvement

All studies were conducted remotely using a Teamviewer session. The Retailio website was hosted on our machine; the participants accessed it using their web browser. Although the UI was the same, we tested two different conditions *for the prediction*: The first condition used the *ridge regression estimator* in the privacy wizard; the second one used the *probabilistic estimator*. Just as in Section IV, the second condition therefore forms the baseline condition. The order in which conditions were used was shuffled using a Latin square.

The procedure was the same for both conditions: Before the start of the study, the participants were given a short introduction on Amazon Go and the Innovative Retail Lab prior to the study. If necessary, the experiment leader described the services that are present in both stores, and which data they need in order to operate. After filling out a questionnaire containing general information (gender/age etc.), the subjects were given a link and the account data for the Retailio website. After logging in, they followed the typical workflow as depicted in Figure 2: First, the privacy wizard was used to do an initial setup of the retail privacy settings. After that, the user reviewed the predicted settings on the main page, and changed incorrectly predicted settings. When the user finished the editing process,

Variable	Question
General privacy	I like the idea of privacy management in general (being able to individually set your own settings)
Privacy wizard	I like the idea of a privacy wizard that helps me to set my permissions
Prefer to Manual	I would prefer a privacy wizard over a manual setting
Prefer predefined	I prefer to use predefined privacy profiles
Trust	In general, I trust the conditions and privacy statements of companies

TABLE VII. ADDITIONAL QUESTIONS ASKING FOR GENERAL ATTITUDE TOWARD WIZARDS, PRIVACY SETTINGS AND GENERAL TRUST IN COMPANIES' PRIVACY POLICIES.

the procedure ended with a subjective rating of the prediction on a 10-point scale (1 = not at all accurate, 10 = very accurate). The participants were then invited to a second, final meeting seven days after the main experiment. We chose this timespan as we assumed that the participants' privacy preferences would not change significantly during this period of time. On the other hand, it was long enough to assume that they had forgotten which settings exactly they chose in the main experiment. The procedure was the same as in the first meeting, this time with the other condition. The second and last meeting ended with an attrakdiff questionnaire[8] as well as additional questions about the general attitude toward wizards, privacy settings and general trust in companies' privacy policies as described in Table VII on a five-point ordinal scale from strongly disagree to strongly agree.

In addition to the questionnaire results, we recorded the number of changes made by the user after finishing the privacy wizard in each of the two conditions.

A. Results

We first analyzed the data on the number of changes made by the user. As a test on normal distribution failed, we used a Wilcoxon signed ranks test, which is a non-parametric test to analyze interval or ordinal data of two populations. As the results show, the machine learning-based privacy wizard ($M_{changes} = 4.35, SD = 2.76$) was better accepted with high significance ($Z = 2.891, p = 0.004$) compared to the control condition using the probabilistic approach ($M_{changes} = 5.6, SD = 2.8$).

The subjective rating for each condition (10-point scale, 1=worst, 10=best) gave us two sets of ordinal data for the two conditions. Tests on normality failed; therefore, we compared the results again using a Wilcoxon signed ranks test, as it is the most favorable statistic for this kind of ordinal data. Also here, the users significantly ($Z = 2.331, p = 0.02$) preferred the machine learning-based settings ($M = 7.05, SD = 1.5$) to those of the probabilistic privacy wizard ($M = 6.1, SD = 1.48$). The rating strongly correlated with the number of errors that were made by the software: In the baseline condition, we achieved a correlation coefficient of 0.814 using a Pearson correlation ($p < 0.001$), and 0.842 for Retailio ($p < 0.001$).

The other additional questions (see Table VII) have been shown to be normal-distributed this time using a Kolmogorov-Smirnov and Shapiro-Wilk test. Therefore we were able to use a one-sample t-test with a test value of 3 (mean of the five-point scale) for the analysis. The results can be found in Table VIII. According to the results, people highly significantly like the idea of managing their privacy settings themselves ($t = 17.61, p < 0.005$) in general and also using a wizard for this task ($t = 10.672, p < 0.005$). Wizards are preferred to manual settings ($t = 4.27, p < 0.005$). In general people do not trust the privacy statements and regulations offered by companies ($t = -3.58, p = 0.002$), highlighting the need for a custom privacy management tool like Retailio. We were not able to prove any trend in whether the subjects prefer to use pre-defined privacy templates instead of setting every single permission themselves, although there is a slight lean towards individual settings rather than privacy templates ($t = -0.195, p = 0.84$).

Variable	mean	T	p
General_privacy	4.75	17.62	<0.005
Privacy_wizard	4.45	10.72	<0.005
Prefer_to_manual	3.70	4.27	<0.005
Prefer_predefined	2.95	0.195	0.85
trust	2.35	-3.58	.002

TABLE VIII. STATISTICAL RESULTS FOR THE ADDITIONAL QUESTIONS.

Retailio received a high pragmatic score ($PQ = 1.37$) which clearly attests a to above-average usability. Although we did not put much effort into the user experience or design aspect of the UI design, we still received a hedonic quality at the edge of being above average ($HQ-I = 1.04, HQ-S = 0.8$). Meanwhile, the attractiveness of the UI remains clearly above average ($ATT = 1.49$). An average user interface would have a neutral pragmatic and hedonic score (about zero). Scores > 1 or < -1 are perceived as *above average* or *below average*, respectively [8].

VII. Discussion and future work

A. Precision of the prediction vs. size of the data set

As we have seen throughout the development process of Retailio, including the correlations found (Figure 1) and the precision of the machine learning algorithm (Table VI), there is a strong correlation between the IUIPC and the retail privacy settings. Although the setting prediction led to good results (about 70% correctness), we think that it is still possible to improve the prediction. Although not fully comparable to our work, research from other areas, like mobile app settings prediction, achieved up to 80% correct predictions [13], [11] with a large online settings database containing several million data sets. In contrast to the mentioned work, there is no large online database about retail privacy settings that we could utilize as training data. We would like to see whether the performance improves with a larger training set.

B. User acceptance

According to the questionnaire results in the main study (Table VIII), customers desire to have control over the data that is collected, shared and used by intelligent retail stores. They generally distrust the companies and their privacy regulations, emphasizing the urgent need for a privacy management system like Retailio. Privacy wizards are perceived as a reasonable approach to support them while making their settings. Shoppers dislike doing all the settings manually, and prefer a wizard to do all the work for them. The predictions performed by Retailio were significantly more precise than those from a simple probabilistic method. Still, we cannot state for sure whether fine-grained individual settings are the best solution for all customers. As the question *Prefer_predefined* shows, opinions are diverse: Some of the subjects stated they preferred pre-defined privacy setting templates, while some liked to be able to adapt every single setting, as offered by Retailio. A different approach to Retailio could use a finite set of privacy profile templates, and use (maybe smaller) questionnaires to select one of the questionnaires, as is done in related work on Facebook privacy settings [16]. To sum up, we can say that a concept like Retailio is accepted. Still, there is another promising approach that could be followed in future work.

C. How to motivate retailers to use Retailio

On one hand, retailers can only collect a limited amount of data (Table 1), although they might be interested in more (like shoppers' experiences at other stores, which food they liked on their vacations etc.). On the other hand, some customers might stay away from a brick-and-mortar retail store that records their data, and prefer conventional stores. The mydata concept (http://mydata.org) is a user-centric privacy approach that allows the customers to collect their personal data at a central point, and to offer parts of it to companies, depending on the purpose it is used for. Using mydata, the customer has the control over his own data and which intelligent retail services she wants to use. The retailer profits from the additional data she is offered, and can offer intelligent services while still maintaining the privacy of the customers. Retailio would perfectly fit into mydata as a frontend for setting privacy preferences.

D. User interface design and user experience

We took an iterative approach when designing Retailio, starting from background research, doing a user study, checking for correlations that could be utilized as a basis for machine learning, and ending up with a proof-of-concept UI that implements our approach. Although a lot of effort was put into the implementation of the UI to make it as convenient as possible and to fit the special needs of the domain of intelligent retail data (see Section III-A), we did not conduct an in-depth design process, including design thinking and the design and evaluation of several layouts. As the results show, the UI is indeed perceived as convenient; on the other hand, there is some space for improvement in the hedonic quality, e.g. the user experience as such when using the interface. Especially the stimulation measure ($HQ-S$) could be improved, meaning the interface could be designed to be more eye-catching and interesting. We would like to go through this process of designing an advanced UI, involving all the steps that are needed for a design process in future work.

In a second step, we want to bring Retailio to customers, connecting it with an intelligent retail store like Amazon Go. We would like to explore in an in-the-wild study whether Retailio will be used in practice, how well the prediction performs with a large user base, and how useful such an approach is perceived to be by the customers.

E. Lessons learnt

The study results show that shoppers *distrust companies* and therefore significantly *prefer to set the privacy settings on their own* rather than using standard settings or privacy templates. Using a *privacy wizard like Retailio is the preferred way* of doing so.

As we learnt from the background analysis, the field of *retail privacy settings* has some *special issues* that make it hard to implement a privacy wizard for that domain. The *type of data* as well as the *sensitivity* is very diverse. We therefore implemented a different type of regressor, called the *compound regressor* to be able to predict an initial set of privacy settings to the customer.

The precision results show that this machine-learning approach can *correctly predict about 70%* of all settings. Nevertheless, on average 30% of the settings remain that are not correctly predicted. We therefore think that a privacy system that is accepted by users always has to *consist out of two parts*: The *privacy wizard* that automatically does some of the privacy settings for the user, and a *privacy UI* that helps the user to get an overview on the actual settings and to help him *correct prediction errors*. Retailio offers both a wizard to automatically derive initial settings, as well as a user interface based on a card metaphor, to offer an overview on the settings and allow an easy adjustment of those.

VIII. Conclusion

New intelligent retail stores like Amazon Go make it clear that we are on the verge of brick-and-mortar stores becoming more comfortable, intelligent, customer-sensitive and individualized. On the other hand, the increasing comfort and individualization comes with a need for a higher amount of individual customer data. Although some accept giving away their data for advanced customer services, not all customers want to share all their data with retail companies; sometimes they want to share only a part of it. We implemented a system called Retailio, which gives shoppers control over, and an overview on, their personal shopping data, and offers a privacy wizard to automatically set up an individual initial privacy profile. We pointed out difficulties that arise especially in the domain of intelligent retail data, did some background research on what data is recorded in intelligent retail stores, and did an online user study to capture how far the personality and privacy awareness of a customer correlates with the desired data disclosure settings (retail privacy settings). Machine learning has been used to build a privacy wizard for Retailio, which creates the initial privacy settings profile after answering some simple questions. The study results show that customers have a strong mistrust of retail companies' privacy settings and a need for control over their personal data. The wizard concept used in Retailio was accepted and the results of the prediction were perceived as useful. Nevertheless, our research brought some different promising approaches as well as chances for further improvements to light, which could make Retailio even stronger in a future version.

References

[1] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell, "Personality and patterns of facebook usage," in *Proceedings of the 4th Annual ACM Web Science Conference*, ser. WebSci '12. New York, NY, USA: ACM, 2012, pp. 24–32. [Online]. Available: http://doi.acm.org/10.1145/2380718.2380722

[2] T. Buchanan, C. Paine, A. N. Joinson, and U.-D. Reips, "Development of measures of online privacy concern and protection for use on the internet," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 2, pp. 157–165, 2007. [Online]. Available: http://dx.doi.org/10.1002/asi.20459

[3] M. Buhrmester, T. Kwang, and S. Gosling, "Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data?" *Perspectives on Psychological Science*, vol. 6, no. 1, pp. 3–5, 2011.

[4] J. Chen, E. Haber, R. Kang, G. Hsieh, and J. Mahmud, "Making use of derived personality: The case of social media ad targeting," in *International AAAI Conference on Web and Social Media*, 2015. [Online]. Available: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10508

[5] P. Costa, R. McCrae, and I. Psychological Assessment Resources, *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Psychological Assessment Resources, 1992. [Online]. Available: https://books.google.co.in/books?id=mp3zNwAACAAJ

[6] L. Fang and K. LeFevre, "Privacy wizards for social networking sites," in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 351–360. [Online]. Available: http://doi.acm.org/10.1145/1772690.1772727

[7] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the big-five personality domains," *Journal of Research in Personality*, vol. 37, no. 6, pp. 504–528, December 2003. [Online]. Available: http://dx.doi.org/10.1016/S0092-6566(03)00046-1

[8] M. Hassenzahl, M. Burmester, and F. Koller, "Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualitaet," in *Mensch & Computer 2003: Interaktion in Bewegung*, G. Szwillus and J. Ziegler, Eds. Stuttgart: B. G. Teubner, 2003, pp. 187–196.

[9] Q. Ismail, T. Ahmed, A. Kapadia, and M. K. Reiter, "Crowdsourced exploration of security configurations," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: ACM, 2015, pp. 467–476. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702370

[10] P. Kumaraguru and L. F. Cranor, "Privacy indexes: A survey of westin's studies," *ISRI Technical Report*, 2005.

[11] J. Lin, B. Liu, N. Sadeh, and J. I. Hong, "Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings," in *Symposium On Usable Privacy and Security (SOUPS 2014)*. Menlo Park, CA: USENIX Association, Jul. 2014, pp. 199–212. [Online]. Available: https://www.usenix.org/conference/soups2014/proceedings/presentation/lin

[12] B. Liu, M. S. Andersen, F. Schaub, H. Almuhimedi, S. A. Zhang, N. Sadeh, Y. Agarwal, and A. Acquisti, "Follow my recommendations: A personalized privacy assistant for mobile app permissions," in *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. Denver, CO: USENIX Association, Jun. 2016, pp. 27–41. [Online]. Available: https://www.usenix.org/conference/soups2016/technical-sessions/presentation/liu

[13] B. Liu, J. Lin, and N. Sadeh, "Reconciling mobile app privacy and usability on smartphones: Could user privacy profiles help?" in *Proceedings of the 23rd International Conference on World Wide Web*, ser. WWW '14. New York, NY, USA: ACM, 2014, pp. 201–212. [Online]. Available: http://doi.acm.org/10.1145/2566486.2568035

[14] N. K. Malhotra, S. S. Kim, and J. Agarwal, "Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model," *Info. Sys. Research*, vol. 15, no. 4, pp. 336–355, Dec. 2004. [Online]. Available: http://dx.doi.org/10.1287/isre.1040.0032

[15] F. Raber and N. Vossebein, "Uretail: Privacy user interfaces for intelligent retail stores," in *Human-Computer Interaction – INTERACT 2017*, R. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O'Neill, and M. Winckler, Eds. Cham: Springer International Publishing, 2017, pp. 473–477.

[16] R. Ravichandran, M. Benisch, P. G. Kelley, and N. Sadeh, "Capturing social networking privacy preferences: Can default policies help alleviate tradeoffs between expressiveness and user burden?" in *Proceedings of the 5th Symposium on Usable Privacy and Security*, ser. SOUPS '09. New York, NY, USA: ACM, 2009, pp. 47:1–47:1. [Online]. Available: http://doi.acm.org/10.1145/1572532.1572587

[17] M. Shehab, G. Cheek, H. Touati, A. Squicciarini, and P.-C. Cheng, "User centric policy management in online social networks," in *Policies for Distributed Systems and Networks (POLICY), 2010 IEEE International Symposium on*, July 2010, pp. 9–13.

[18] M. Shehab and H. Touati, "Semi-supervised policy recommendation for online social networks," in *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, ser. ASONAM '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 360–367. [Online]. Available: http://dx.doi.org/10.1109/ASONAM.2012.66

[19] H. J. Smith and S. J. Milberg, "Information privacy: Measuring individuals' concerns about organizational practices," *MIS Q.*, vol. 20, no. 2, pp. 167–196, Jun. 1996. [Online]. Available: http://dx.doi.org/10.2307/249477

[20] L. Spassova, J. Schoening, G. Kahl, and A. Krueger, "Innovative retail laboratory," in *Roots for the Future of Ambient Intelligence. European Conference on Ambient Intelligence (AmI-09), 3rd, November 18-21, Salzburg, Austria*. o.A., 2009. [Online]. Available: https://www.dfki.de/web/forschung/publikationen/renameFileForDownload?filename=AmI-Landscape-InnovativeRetailLab.pdf&file_id=uploads_338

[21] R. Xu, R. M. Frey, D. Vuckovac, and A. Ilic, "Towards understanding the impact of personality traits on mobile app adoption - a scalable approach." in *ECIS*, J. Becker, J. vom Brocke, and M. de Marco, Eds., 2015. [Online]. Available: http://dblp.uni-trier.de/db/conf/ecis/ecis2015.html#XuFVI15