Testing Privacy Awareness

Mike Bergmann

Technische Universität Dresden, Germany

Abstract. In web-based business processes the disclosure of personal data by the user is an essential part and mandatory for the processes. Privacy policies help to inform the user about his/her rights and to protect the user's privacy. In this paper we present a test to empirically measure how the user's privacy awareness changes by presenting specific elements of the privacy policy in close proximity to the required data items. We compare an experimental group using an enhanced interface to a control group using a conventional interface regarding their capability to recall the agreed privacy-related facts. A concrete online survey has been performed. The major results are presented.

1 Introduction

Privacy has received particular attention in the media and the Internet in the last few years. Classical desktop applications were transferred into the context of the Internet and enable sharing documents over large distances. New web services were created serving various user demands and introducing a complete new application landscape. The major resource of all these services are data and in many cases user data. European legislation acknowledged the situation by defining regulations regarding user rights and privacy protection [7]. Each web site which processes user data has to present a privacy policy to declare the main facts about its data processing. There also exists a technical solution to communicate the essential facts of the privacy policy in a machine readable form [22]. However, usually the user does not read these statements. First, the texts contain lots of legal statements that are difficult to understand [2]. Second, the companies use the privacy policy to rephrase painful facts in vague and sweet words [21]. So the current legislation is rather protecting the interests of companies than the interests of the users.

The need to present privacy policies in a more effective way is obvious. Thereto the presentation should accompany the original business processes, should present the main facts regarding user data and user privacy and should not monopolize the user's attention. In [20] we proposed a solution to solve this Gordian Knot in a user-friendly manner. This paper aims to validate the proposed approach by comparing the resulting privacy awareness to the ordinary presentation approach.

In Section 2 we start with an overview about the related work on this topic. We sketch roughly the major parts of the interfaces, we have to test, while Section 3 then lists the current configuration of our experiment, discusses special aspects

of the experiment and presents the statistical methods to analyze the results of the experiment. Finally we close our paper with an outlook to further interesting topics.

2 Related Work

In this section, we elaborate the term *privacy*, some legal and technical factors of it, the term *privacy awareness* and discuss some of the existing approaches to present privacy policy in a privacy-enhancing manner. Furthermore we give a short overview about existing privacy surveys.

2.1 Privacy

There exist various privacy definitions, starting with Warren and Brandeis in 1890 [23], a more popular and general definition by Westin in 1967 [24] and for instance a definition by Fischer-Hübner [13]. According to the latter, we focus on a special branch of privacy, namely *informational privacy* as the *right of informational self-determination*. Based on these facts, we define:

> **Privacy** in the context of this paper means the right of self-determination regarding data disclosure, i.e., each user should be able to control *how much* personal information he is *willing* to give *to whom* and for *what purpose*.

This includes the following components: data minimization, purpose binding, data transfer statement, minimal data retention and informed consent (cp. also [7, 9, 3]). The European legislation acknowledged the increased importance of user's privacy protection and the necessity of secure and privacy-friendly data processing by issuing legal foundations, namely the Data Protection Directive 95/46/EC [7]. The directive defined that the purpose of data processing and the data processor itself are mandatory to state in the privacy policy.

The P3P specification [22] goes one step ahead and defines various privacy-related attributes, which are machine readable, to allow automatic evaluation of privacy policies. Based on the P3P specification, Cranor et al. developed a new approach for configuring and presenting P3P preferences [8]. Their approach visualizes the degree of the correspondence of the user's privacy preferences with the privacy policy of the web service. However the proposed approach has some drawbacks. The user's privacy concerns are strongly related to the communication partner [9]. So we miss the possibility to define dedicated privacy preferences with respect to the communication partner for the user. Besides, the visualization of a missing P3P policy[1] as less critical as a mismatch of a certain preference, we count as not appropriate for enhancing the user's privacy.

Until now in conventional data-submission forms, the corresponding privacy policy is missing. Often it is accessible via a separate link referring to the privacy

[1] what in fact means that the service could do anything with the personal data

policy. This implies additional actions for the user to get informed about the circumstances of the data disclosure. Sometimes additional marks are set to separate mandatory respectively optional data (see Figure 1). In [5] an approach is sketched to present privacy policies as a mapping of user defined privacy preferences and to emphasize the non-matching details. It allows the definition of privacy preferences per communication partner and takes the main privacy facts *data minimization, purpose binding and informed consent* into account. We also follow the suggestions to design user interfaces, proposed in the PRIME project [19].

The corresponding graphical presentation of our enhanced interface is shown in Figure 2. Using the icon in the right upper corner allows to access the full privacy policy. We include data transfer and retention statement into the presentation as we consider it as important to enhance the end user's privacy (regarding the privacy definition above). However these additional statements are not required by the EU directive 95/46/EC [7].

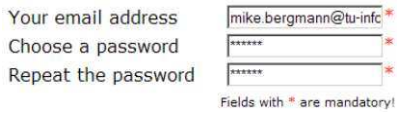

Fig. 1. Conventional interface for online forms

Fig. 2. Enhanced interface – information about the privacy policy nearby the data to disclose

According to our privacy definition, we simplify the meaning of privacy awareness for the testing and define it as:

> **Privacy Awareness** in our context is seen as the user's ability to reflect the communication partner's privacy policy statements regarding purpose binding, transfer assertion and retention period applied for a certain data disclosure.

We mention the obligation of the service provider to make the user aware of the privacy policy in Section 2.1. Usually it is taken into account by offering an omnipresent always available link to the privacy policy. However earlier and recent user tests documented that the ordinary web user neither reads nor understands the complex legal texts and has a blind spot regarding secondary information (like advertisement, banners etc.) [4, 6]. We have to make sure that we counter this by using appropriate visualization technologies. A simple text-based, not intercepting approach to present privacy policies seems to be better [11].

2.2 Privacy Surveys

Privacy studies have a long history. In the late 1960s, Alan Westin started to conduct privacy surveys [25]. He did fundamental research in creating various

general and specific indices regarding privacy. He partitioned the population into three classes, the so called *fundamentalists, pragmatists* and *unconcerned users* [26]. A description of these surveys is given in [18]. However reliable details about these surveys are not available.

Gideon et al. [15] tested the influence of information regarding the corresponding privacy policy available nearby the search results on web users' purchase decisions. They found that the simple existence of a privacy policy does not influence the purchase decisions. But the presence of a clear indicator about privacy-related facts influenced the purchase decision. A so-called privacy-bird symbol, a graphical metaphor visualizing the degree of the matching of the privacy policy with the users' privacy preferences, was displayed nearby the search result. Depending on the concrete matching details, a red, yellow or green bird was shown [8]. These results are supported by further surveys [12, 10]. However the studies do not evaluate the privacy awareness of the user. It is not verified whether the user really is aware of the privacy issue or is just afraid of the red signs. Informed consent in the sense of really informing the user about the disclosure conditions is not obtained. The privacy-policy representation, especially purpose binding and assurance evaluation (see also [1]) are insufficiently addressed by presenting just red/yellow/green indicators.

3 The Privacy Awareness Test

In this section, we motivate our approach to test the privacy awareness of the user. We describe the methodology, the global settings and the potential participants of the experiment. We continue by explaining the single experimental steps *Preparation, Application Scenario, Post Processing and Debriefing* in detail. The results based on concrete indices and statistical features are listed and discussed.

3.1 Motivation

One of the questions the test should answer is: "Does the user really perceive the privacy policy statements, presented in a superficial manner such that we could achieve an increased privacy awareness?" (see Figure 2). We need two different groups, the *Control Group* G_{NoPet} using the conventional web forms for data disclosure and the *Experimental Group* G_{Pet} using our enhanced interface presenting the details regarding the privacy policy.

In this context, a sub-question will be how the perception of the privacy policy differs among the various user classes. Our hypothesis is that privacy fundamentalists and pragmatists appreciate the enhanced presentation form, while the unconcerned users still ignore it.

By saying this we have to check, how the proposed approach increases the privacy awareness, in particular how it influences the user's knowledge about

privacy-related facts. As measurable privacy-related facts about privacy, we see e.g. the following[2]:

- Contact Partner - Various surveys have shown that the most prominent decision factor is the communication partner itself [9]. For well-known and established partners, web users are less concerned about their privacy.
- Purpose Binding - The requested data is bound to a dedicated purpose. Many users have expressed concerns about potential abuse of their personal data [14]. A clear purpose statement helps the user to understand what the requested personal data is used for, e.g., that the disclosed email address is for order confirmation only.
- Data Transfer Statements - Users are very concerned that their data is transferred to other recipients without permissions. An example is the anxiety about the abuse of email addresses to send them spam [21, 14].

3.2 The Design of the test

a) Preface: There are different methods to answer the questions, mentioned above. In a supervised setting we could just monitor (e.g., eye tracking) the test subjects during the usage of the proposed interfaces and interview them afterwards about their understanding of the interface elements. However, this interview approach is applicable only for a limited set of participants. Besides we think this test approach does not cover the usual user behaviour in the context of the Web2.0. Because of the these limitations, an interview will not deliver representative results. A more valuable approach would be a simulation of a real Web2.0 scenario with community components deployed as an online experiment, accessible for a much broader audience. An appropriate questionnaire before and after the simulation should gather the desired facts and should replace the conventional observation and interview.

b) Requirements: Existing online communities are a promising environment to recruit participants for the experiment. They do have appropriate knowledge about business processes in the Internet and they are used to disclose personal data for various purposes. Besides, they are the main target group for our enhanced interface.

Because of the online experience of the prospective participants we have to deal with some top-level requirements regarding the plausibility and authenticity of the experimental scenario. We have to take existing applications as a model paragon. The awareness of the test participants about the fact that it is "just an experiment" should be lowered by simulating real Internet business processes and using the corresponding terminology and presentation styles. We have to simulate the email-confirmation mechanism and we have to place advertisements into the web sites of our online application. We will use a similar color schema like Google (www.google.com) uses to get close to a realistic and well known Web service application.

[2] cp. [7, 24, 13]

c) Focus of the experiment: Due to the characteristics of online media, we are able to attract people from all over the world and with various social and educational backgrounds. Our experiment should gather these properties in a first step. These properties will help us to recruit representative test participants to achieve representative results. We will elaborate the concrete difference between the groups of unconcerned participants, pragmatists and fundamentalists. In detail we will collect demographic facts about the test participants and capture the ability of the participants to express concrete facts about the preferences of the privacy policy of a dedicated Web service. Besides, tracking the click stream of the participants enables us to answer the question whether web users do read the privacy policies in general and what are the differences between the privacy fundamentalists, pragmatists and privacy unconcerned users in particular.

d) Methodology: For our test we combine the classical survey method with an experimental part. The survey frames the experiment, aims to collect the participants' demographic preferences, the knowledge about online business processes, common privacy concerns, and will collect our relevant feature set. The questionnaire is construed as a differential cross-sectional survey questionnaire to determine the relationship between privacy concerns and privacy-related behaviour in general and between presentation and perception of privacy-related interface elements among the different privacy-concerned groups in particular. In the pre-questionnaire we use the common Likert scaling [17]. It offers equidistant and well elaborated scaling. We do not allow neutral values to force the participant to make a dedicated yes/no decision. However we allow the refusal to a dedicated question at all.

To gather the experimental results in the post-questionnaire we use a direct scaling, listing concrete options as answers with a dedicated "not sure" option. The selection of "not sure" marks the answer as not countable for the calculation of the result. This avoids the so-called "water down" effect because of valueless data records.

Based on the questionnaire before the simulation, we select a representative sample of all participants. We use common statistical measures to apply the selection. The participants are invited using various international online-sources like mailing lists of online communities, business networks, professional survey and marketing portals.

To design the questionnaire and the experimental part, we perform pre-tests and interviews with potential participants. The found problems and issues were documented and fixed before putting the survey online.

3.3 The configuration of the test

The test is performed in three steps. At step one, we are querying items to classify the participants. As a second step, we present a typical online application. The third step aims to gather knowledge of the user about the privacy policy assigned to the disclosure process, respectively to the disclosed data items. This

step is accompanied by a debriefing section where the participant is informed that no personal data at all have been transmitted. Questions about whether the data disclosed by the user were true or faked conclude the test. A cookie-based mechanism avoids that normal users perform the test twice. However we acknowledge that experienced users may overcome this protection mechanism.

We posted messages to distribute our invitation for the survey in international and social networks (www.xing.com, www.linkedin.com) and at mailing lists of online research communities (Association of Internet Researchers, German Society for Online Research, University of Maryland, web2list.com, www.i-worker.de, genpsylab-wexlist.unizh.ch, etc.).

Step 1: Preparation The preparation step aims to motivate the test participants and to collect statistical facts about the participants with regard to demography (e.g., male/female, nationality etc.) and privacy concerns. We collect common statistical features like gender, age, nationality, general Internet skills etc. Besides, we ask questions about general privacy concerns to be able later on to cluster our participants regarding their expressed privacy concerns. This step is introduced with some moderation and the offer to participate in a draw for some material stimulation of the participants. The preparation is concluded by a short explanations of the next steps and a suggestion to print this text to have it available as a handbook for the web application.

Test	Question	T/F[3]	Valid Answer
	Please specify your age.		18-24; 25-34; ... 55-64; 65 and older
	Please specify your gender.		Male/female
	Please state your country/region.		list of countries to select
	I use the Internet for e-shopping ...		weekly or more often, monthly, rarely, never
stat	I spend most of my spare time using the Internet.		
	I speak English very well.		Strongly agree, ... strongly disagree
	I don't need help when I am using a computer.		
	I always change my browser settings to protect information about myself.	T	
	It makes sense to use different email addresses for different situations.	T	
	The probability of personal data (like credit card number, email address, online account information) misuse on the Internet is very high.	T	

[3] T=True, the scale is direct; F=False, the scale is inverted.

pc	Consumers have lost all control over how personal information is collected and used by companies.	T	Strongly agree, ... strongly disagree
	Most businesses handle the personal information they collect about consumers in a proper and confidential way.	F	
	Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.	F	
sn	Customer feedback is valuable to make decisions about products and services.		Strongly agree, ...strongly disagree
	When choosing a restaurant, I take suggestions from my personal circle into account (family, friends, colleagues).		
	Only unsatisfied customers are posting feedback about products and services on the Internet.		

Table 1: Pre-questions to gather statistics and privacy concerns.

We avoid any mentioning of the word *experiment* and *privacy* to not bias the user before performing the test. We add special *social networking* questions (see the *sn* rows in Table 1) to make the user believe we survey about Web2.0 topics. We will enlighten the user after the application scenario in the debriefing section.

Step 2: Application Scenario As a typical online application we present a "Foodie" web service. This service pretends to collect user recommendations about restaurants, including evaluation of food quality and price level. Therefore the participant has to perform some data disclosure, namely to register to the service submitting an email address and to assess a restaurant (details below). The privacy policy is available at any time.

To answer the question raised at the beginning of this chapter, we separate the participants randomly into two groups. The first groups is presented with the ordinary interfaces, the second group is presented with our enhanced interface (see Figure 1, 2).

Registration: We require a valid email address as user name for the purpose "registration" and a password. To increase the plausibility, we promise to send a confirmation email with an activation link. Various other data that are not really necessary for the transaction are requested. We ask for surname and city. The specified purpose is "personalization".

We apply a privacy policy according to the European Data Protection Directive 95/46/EC [7]. The privacy policy states that the data is not sent to any partner by default, that the data was requested to offer personalization, to provide validity check for the service provider and that the data is stored until objection by the participant (see Figure 3 - "Foodie" Privacy Policy).

FOODIE PRIVACY POLICY(fragment): Foodie will not share with, sell or transfer any data or personal information provided to us through usage of the Service to any third party without prior and explicit consent.
The requested personal data (email address, first name) is used to allow you to create an account at Foodie. The email address will be necessary for authentication and to contact you for product announcements and marketing information. We use your personal data to personalize our Services.
All content uploaded to the Service is your own private property. Foodie will not read, change, destroy or forward the contents of your account, unless entitled to do so by this Agreement or forced to do so by law, regulation or any extraordinary circumstance.
By making content public on your profile you give an explicit consent to show the information chosen to the audience specified for statistical purpose. Your first name and city of origin stored in your own account profile is visible to other users by default.
Foodie retains the right to temporarily or permanently discontinue any specific features at its own discretion. Foodie has no obligation to keep the uploaded data. Foodie stores your personal data until you object.
...

Fig. 3. "Foodie" Privacy Policy

Optional: A check box "use this email for product information and special offers" is presented for marketing purpose. The checkbox is selected by default. A separate page with the privacy policy for that purpose is accessible via a "privacy policy" link. The page contains information about the possible usage of the email address for possibly sending suitable special offers and news.

Input: To add a restaurant evaluation/assessment, the user has to create a new restaurant entry. The user is requested to fill out four data fields describing the restaurant (name, place, kind of, price category) and may add a free descriptive text including tags. The data entered here are public. A privacy policy informs the user about this fact. The data is not sent to the restaurant itself. For statistical purposes, the first name and city of the participant are displayed nearby the restaurant assessment entry. This quite unusual purpose helps us to avoid that the user guesses the questioned purposes in the post-question section.

Step 3: The post-questionnaire and debriefing part will be presented after the successful finish of the application scenario.

Var	Question	Valid Answer
F_A	The email address was requested for the following purpose(s):	
F_a	I accepted the following usage of the email address for the following purpose(s):	Registration,

Var	Question	Valid Answer
F_D	The additional personal data (name, city) were requested for:	personalization, marketing, statistics,
F_d	I accepted the following purpose(s) for additional data usage:	not sure[4]
F_R	Did the "Foodie" web services promise to delete your personal data if you send a deletion request?	yes, no, not sure
F_T	Does the "Foodie" web services transfer your email address to restaurants for special offers?	
F_C	Did the "Foodie" web services request your personal data via a secure https connection?	

Table 2: The Post-Questionnaire.

Post questionnaire: To answer the survey questions we raised at the beginning of Section 3, we have to find out whether or not the participant can recall the statements about the privacy policy, he/she agreed on during the test. Therefore, we ask about the stated purpose for disclosing the data item 'email' and additionally 'name' and 'city' for the restaurant-assessment entry. The corresponding questions are listed in Table 2. Depending on the participant's knowledge about the correct purposes we calculate a privacy awareness index (see Section 3.4).

Debriefing: The closing part of the experiment is introduced by a short debriefing as follows in Figure 4. It may happen that the respondents are feeling cheated at this point. By offering the chance to win a valuable technical gadget, we try to compensate this feeling.

It was just a simulation!
Sorry for being devious, but we had to make you believe it's real. Actually, the 'Foodie' service was just a simulation. We did not transfer any personal data like your name, email etc. As we promised in our privacy policy, no personal data at all was disclosed. The application just generated a pseudonymized data record only to allow further research on the survey results, but without your concrete personal data.

Please answer our last questions about the data items themselves, which you putatively disclosed to the 'Foodie' service. Please tell us honestly...

Fig. 4. Debriefing

[4] Multiple choices allowed or "not sure".

Var	Question	Valid Answer
T_D	Was the "data item"[5] correct?	yes, no, not sure
T_{R1}	Did you believe the 'Foodie' website was real?	
T_{R2}	Did you answer the questionnaire seriously?	
F_S	I need more privacy-related information about usage of my data during disclosure.	Strongly agree, ... strongly disagree
F_E	I wish to see an easier-to-understand presentation of privacy related information about usage of my data during disclosure.	

Table 3: Debriefing.

The test was scheduled to be performed within a two months period. We planned to have at least 50 participants in each of the two groups, successfully passed the test, with valid test results and acknowledged that they passed the test seriously (see the question F_{R2} in Table 3). Participants, who cancelled the test before answering the post questionnaire are not counted.

3.4 Expected Results

Hypothesis: Even if the test participants do not read the privacy policy, the participants who received the enhanced interface do know the answers better than the control group with the ordinary interface.

Privacy Concerns Index (px): Based on the results of the pre-tests, we classify the two groups (experimental group and control group) each into the classes of privacy *Fundamentalists, Pragmatists* and *Unconcerned* similar to Westin's approach [26]. The participants are clustered by using the pc index (px). It is calculated as the sum of the six privacy-related answers (see Table 1, pre-questions, part 'pc').

The "strongly agree" answer gets assigned the value 4, the "strongly disagree" answer gets the value 1 assigned. In case of inverse meaning (see the true/false column in Table 1) we invert the assignment.

To assign the participants to the corresponding classes, we use the quartiles of the value of px. The first quartile represents the privacy-unconcerned participants. The following two quartiles represent the privacy pragmatists. The fourth quartile represents the privacy fundamentalists. This approach differs from the approach Westin used for classification. However, due to the huge differences in the sample size and setting, in the method the survey was performed, etc., a direct comparison seems not useful.

[5] There are three separate questions. Valid values for "data item" are email, first name, city.

Privacy Awareness Index (ax): For the first four post-test questions F_A,F_a,F_D and F_d multiple answers are allowed (see Table 2). The answer F_A is correct, if only the checkboxes for purpose "registration" and "marketing" are checked. For each correctly checked respectively not checked checkbox, F_A is increased by 1, so $0 \leq F_A \leq 4$.

To avoid that the participant guesses the answer, we introduced the purpose "statistics". The answer F_D is correct if the checkboxes for purposes "statistics" and "personalization" are checked, all others should remain unchecked. For each correct checkbox, we increase F_D by 1, so $0 \leq F_D \leq 4$.

The correct value for answer F_a depends on the status of the checkbox allowing the usage of the email address for "marketing" purpose. If this purpose was allowed, then the test participant should check the corresponding checkbox, so $0 \leq F_a \leq 4$. In this work, we ignore this answer.

The correct value for answer F_d depends on the status of the checkbox allowing the usage of the first name and the city for statistical purpose. It follows the same schema as for F_a. In this work, we ignore this answer.

For the results F_R, F_C, F_T we have only one correct answer. If the answer is correct, the corresponding value is 1. In case the value is not correct we assign -1, so allowed values for these variables are ± 1.

To take into account participants just exercising the test for scientific reason or curiosity we offer the option to invalidate the own data record by answering "no" to the question T_{R2} (see Table 3). So we are able to lave out these data records.

The coefficients F_S and F_E are representing the information requirements of the test participant. In the current experiment we ignore these answers. In the future we could use these answers to evaluate the overall result in more detail.

As an indication of privacy awareness, we use the sum $ax = F_D + F_R$. These answers allow to conclude whether the participant read respectively perceived the purpose "statistics" and the statement about the possibility to object the data disclosure afterwards. Higher values for ax represent a more correct response. Summarizing the two value ranges of F_D and F_R the awareness index ax has the value range of $-1 \leq ax \leq 5$.

The results F_A, F_C and F_T are control variables. Assuming an influence of the enhanced interface the results F_A and F_C should be similar in both groups because the answers are not depending on the enhanced interface. The answer for F_T should underline the trend, found in ax. Corresponding to our *hypothesis* we expect a higher percentage of correct answers (a higher index ax) in the group with the enhanced interface, especially in the sub-groups of the "pragmatists" and "unconcerned". We will list the results per group.

3.5 Statistics

Equidistance: The answers of the test participants have to be distributed equidistant, so instead of naming all values like *Strongly agree, agree, disagree, strongly disagree* we only offer the names for the edge values *Strongly agree, . . . , strongly disagree*. A neutral value is missing so the test participants have to

vote in a clear direction. This forces them to make clear statements about their position.

Participants classification: To classify the participants with regard to their privacy concerns we suggest to use the corresponding quartiles of the privacy concerns index. In [26] Westin proposed a different algorithm to separate survey participants into privacy fundamentalists, pragmatists and unconcerned participants. However the algorithm seems to be arbitrary with regards to the quantitative distribution. To improve this and to be more objective, we propose the separation using the quartiles of the privacy concerns index px.

Stochastic independence of the results: To prove that the data collected for ax are stochastically independent, we use Pearson's chi-square test.

4 Results and outlook

In this section, we summarize the main results of the survey. We will assess the instrument and discuss further ideas and research questions extending or reusing the developed instrument.

4.1 Common results regarding participation and classification

The survey was announced to a broad audience in various online forums. Therefore we got many participants. Counting the participants using the start button to start the survey, we got 618 participants. An overview about the participants for the experimental group G_{Pet} and control group G_{NoPet} at different stages of the survey is given in Table 4. There was no statistical selection performed towards a more representative data sample regarding demographical or social parameters like age, gender, education etc.

Participants	G_{NoPet}	G_{Pet}	Sum
Overall	not available		618 (100%)
Completed the pre-questionnaire	not available		496 (80.3%)
Entered the registration form	217 (35.1%)	214 (34.6%)	431 (69.7%)
Filled out the registration form	86 (13.9%)	86 (13.9%)	172 (27.8%)
Completed the survey	78 (12.6%)	78 (12.6%)	156 (25.2%)

Table 4. Overview about the survey participants.

The classification of the participants similar to Alan Westin's classification [26] is shown in Table 5. We applied the classification schema as described in Section 3.4. In the following we take into account only these participants completing the survey successfully. As the corresponding criterion, we use the state of the checkbox T_{R2} (see Table 3). Following this criterion, 156 participants completed the survey. For these participants we got px in the range of $12 \leq px \leq 24$. Using quartiles usually gets four equally distributed sets. Due to the discrete character

of px we had to adopt this quartile approach. We first separated our participants in four equal parts of 39 participants per part according to the quartiles. Then we looked for the transition of px to the next lower value. This point we took as the class limit. So we got as a first part of the participants a class of 28 unconcerned participants. The parts two and three we count as pragmatists (75). Participants belonging to the fourth part are counted as fundamentalists (53). Inside these classes, we show how the participants are distributed regarding assignment to the experimental group and control group. We got the classification, shown in Table 5.

Class	Unconcerned		Pragmatists		Fundamentalists	
Range of px	$12-17$		$18-20$		$21-24$	
Overall particip.	28 (17.9%)		75 (48.1%)		53 (34.0%)	
Participants	G_{Pet}	G_{NoPet}	G_{Pet}	G_{NoPet}	G_{Pet}	G_{NoPet}
per group	11 (7.1%)	17 (10.9%)	40 (25.6%)	35 (22.4%)	27 (17.3%)	26 (16.7%)

Table 5. Participants classification regarding the privacy concerns index px.

Summary: In our sample about 18% of the participants were counted as *Unconcerned* regarding privacy. About 48% of our participants were *Pragmatists* regarding privacy. About 34% we count as *Fundamentalist*. This does not ideally represent the separation into quartiles as proposed at the beginning, because the fourth class is much bigger than the first. However this may be plausible due to the omnipresent news about data leakage and data misuse in the media, the raised importance of data protection on the Internet in the last years and the need and requirement to use the Internet for daily business and private activities. This could be subject of further research.

4.2 Results regarding our hypothesis

In Section 3.4 we postulated our hypothesis. We assumed that participants belonging to our experimental group G_{Pet} do have a higher privacy awareness than the participants of the control group G_{NoPet}. This is indeed the case. Figure 5 shows the mean value for ax per group and class[6]. Figure 5 shows that in general the ability of the participants of the experimental group G_{Pet} to reflect the main preferences regarding the privacy policy, stated by the service provider, is higher. The significance of this outcome we prove with Pearson's chi-square test. The results are stochastically independent with the probability of approximately 0.995%. This fulfills our requirements formulated in Section 3.5.

Conclusion: We have shown that the proposed approach for presenting information related to the privacy policy of a certain transaction does significantly help the user to perceive the essential privacy preferences, like purpose of data

[6] For details regarding calculation of ax see Section 3.4.

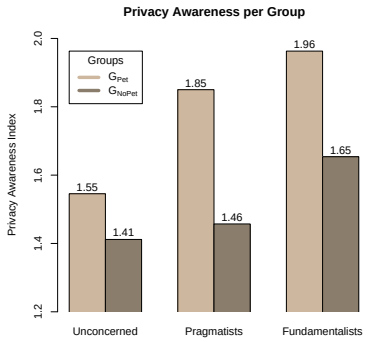

Fig. 5. Contrasting experimental and control group regarding mean ax

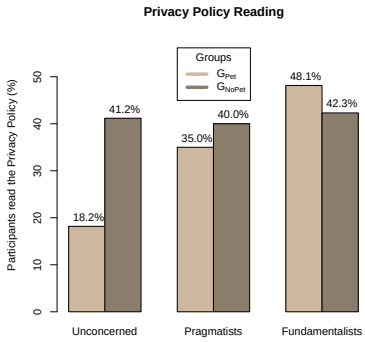

Fig. 6. Contrasting experimental and control group regarding policy reading

usage. The effect was observed for all classes, but with most success for pragmatists. Our hypothesis was confirmed. The effect is even excelling our expectations because the increase of ax was also observed for the class of fundamentalists. Initially we expected that the fundamentalists read and perceive the preferences of the privacy policy anyway. The increase of ax may be due to the mismatch between stated vs. observed behaviour [16]. So we may conclude that the usage of enhanced interface pays off for all classes of web users.

4.3 Further interesting results

Besides we may have a look at the question "Does the Internet user read the privacy policy?" (see Section 3.1). Based on the 156 valid data records, we may state that in general there is no correspondence of stating privacy concerns and acting privacy-concerned, respectively privacy-aware, as shown in Figure 6. The frequency for reading the privacy policy in the control group is nearly uniformly distributed. So even fundamentalists do not read the privacy policy more often than the unconcerned in the control group. This corresponds to the findings of Jensen et al. [16].

However in the experimental group it looks different. Except the fundamentalists all other participants of the experimental group G_{Pet} did read the privacy policy less often. This may be due to the fact that in general, the presence of the privacy-related information satisfied the desires regarding privacy information of the participants belonging to the classes of unconcerned and pragmatists. For the fundamentalists it increased the policy-reading frequency. This may be due to the fact that the presence of the privacy-related information 'remembers the user to have a look at the privacy policy'. Due to the thin result set, the findings however are not very reliable and have to be validated.

A further topic of interest could be how the validity of the submitted data varies through the different classes of web users. However we did not elaborate

this relationship. This is definitely a topic of further research. As a hypothesis we may state that we expect more valid results within the well-informed group using the enhanced interface.

4.4 Outlook

The results of this survey are promising and may contain further interesting facts. Using direct and indirect survey outputs, we could elaborate the following questions:

- Is there a difference between the Personal Data Validity Index (vx) within the different privacy concerned user groups? To estimate the validity of the answers given by the test participants, we introduced post-test questions after the debriefing of the users (see Section 3.3). The index may show whether the enhanced interface increases the quality of the disclosed data or not. It could also be taken as an implicit measure of the applied privacy protection of the participants. If the participant disclosed incorrect data (e.g. a wrong city, a misspelled name etc.) we may assume more privacy awareness. It could also be taken as a measurement to assess how users vary their privacy protection requirements based on available privacy policy information.
- What part of participants disabled java script and/or cookie functionality? Are these participants related to the fundamentalists? This may be a measure for the consistency of stated and observed behaviour.
- Is there a difference in the percentage of participants who do not continue to complete the tests between the different user groups?
- Is there any difference between different demographic and ethnic groups (e.g. between young/old, male/female people)?

Acknowledgement

Thanks to our colleagues for the helpful comments. Special thanks to the team of the Technische Universität Dresden, namely Andreas Pfitzmann, Rainer Böhme, Stefanie Pötzsch, Sandra Steinbrecher, Sebastian Clauß, Stefan Köpsell, Stefan Berthold and Wiltrud Kuhlisch, to the Karlstad University team, namely Simone Fischer-Hübner, Maria Lindström, John-Sören Pettersson and Erik Wästlund, and to Diane Whitehouse.

The research leading to the results presented in this paper has received funding from the European Community's Sixth and Seventh Framework Programme (FP6/2002-2006 resp. FP7/2007-2013) for the projects FIDIS, PRIME and PrimeLife. The information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The PrimeLife consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

References

1. Christer Andersson, Jan Camenisch, Stephen Crane, Simone Fischer-Hübner, Ronald Leenes, Siani Pearson, John Sören Pettersson, and Dieter Sommer. Trust in PRIME. *Proceedings of the 5th IEEE Int. Symposium on Signal Processing and IT*, December 18-21, 2005. Athens, Greece.
2. Annie I. Antón, Julia B. Earp, Davide Bolchini, Qingfeng He, Carlos Jensen, and William Stufflebeam. The lack of clarity in financial privacy policies and the need for standardization. In *IEEE Security and Privacy*, volume 2, pages 36–45, March 2004.
3. Annie I. Antón, Julia B. Earp, Matthew W. Vail, Neha Jain, Carrie Gheen, and Jack M. Frink. An Analysis of Web Site Privacy Policy Evolution in the Presence of HIPAA. Online available at http://www4.ncsu.edu/~njain/publications/hipaa_7_24_submit.pdf.
4. Jan Panero Benway. Banner Blindness: The Irony of Attention Grabbing on the World Wide Web. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 1998. Rice University Research, Houston, Texas.
5. Mike Bergmann. Generic Predefined Privacy Preferences for Online Applications. In Simone Fischer Hübner, Penny Duquenoy, Albin Zuccato, and Leonardo Martucci, editors, *The Future of Identity in the Information Society: Proceedings of the Third IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School*, International Summerschool Karlstad, Sweden, May 15 2008. Springer.
6. Moira Burke, Anthony Hornof, Erik Nilsen, and Nicholas Gorman. High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12, December 2005. Rice University Research, Houston, Texas.
7. Council of Europe. Data Protection Directive 1995/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995. Official Journal L No. 281, 23.11.1995.
8. L. Cranor. P3P: Making privacy policies more useful. *IEEE Security and Privacy*, pages 50–55, 2003.
9. L. Cranor, J. Reagle, and M. Ackerman. Beyond Concern: Understanding Net Users Attitudes About Online Privacy. *AT&T Labs-Research Technical Report, TR 99.4.3, April 1999.*, 1999. Online at http://citeseer.ist.psu.edu/cranor99beyond.html.
10. Lorie F. Cranor. What do they "indicate?": Evaluating Security and Privacy Indicators. *Interactions*, XIII.3:45–57, 2006.
11. D. Diaper and P. Waelend. World Wide Web working whilst ignoring graphics: good news for web page designers. *Interacting With Computers*, 13, December 2000.
12. Serge Egelman, Janice Tsai, Lorrie Cranor, and Alessandro Acquisti. Studying the Impact of Privacy Information on Online Purchase Decisions. *Workshop on Privacy and HCI: Methodologies for Studying Privacy Issues at CHI 06*, 2006. Online available at http://cups.cs.cmu.edu/pubs/chi06.pdf.
13. Simone Fischer-Hübner. *IT-Security and Privacy – Design and Use of Privacy-Enhancing Security Mechanisms.* LNCS 1958. Springer Scientific Publishers, May 2001.
14. Susannah Fox, Lee Rainie, John Horrigan Amanda Lenhart, Tom Spooner, and Cornelia Carter. Trust and Privacy Online: Why Americans want to rewrite the

Rules. *The Pew Internet & American Life Project*, August 20, 2000. Online available at http://www.pewinternet.org/pdfs/PIP_Trust_Privacy_Report.pdf, last visited Nov. 11, 2008.
15. Julia Gideon, Lorrie Cranor, Serge Egelman, and Alessandro Acquisti. Power Strips, Prophylactics, and Privacy, Oh My! *ACM International Conference Proceeding Series, Proceedings of the second symposium on Usable privacy and security, Pittsburgh, Pennsylvania*, 149:133–144, 2006. ISBN:1-59593-448-0; url: http://portal.acm.org/citation.cfm?id=1143120.1143137.
16. Carlos Jensen, Colin Potts, and Christian Jensen. Privacy practices of Internet users: Self-report versus observed behavior. *International Journal of Human-Computer Studies*, 63:203–227, 2005. DOI 10.1016/j.ijhcs.2005.04.019.
17. A. Kallmann. Skalierung in der empirischen Forschung, 1979. München.
18. P. Kumaraguru and L. Cranor. Privacy Indexes: A Survey of Westin's Studies. *ISRI Technical Report, CMU-ISRI-05-138, Carnegie Mellon University*, December 2005.
19. John Sören Pettersson. (Ed.) HCI Guidelines, PRIME Deliverable D6.1.f. *PRIME Deliverable*, 2008. version 1 February 2008, https://www.prime-project.eu/prime_products/reports/arch/.
20. John Sören Pettersson, Simone Fischer-Hübner, Ninni Danielsson, Jenny Nilsson, Mike Bergmann, Sebastian Clauß, Thomas Kriegelstein, and Henry Krasemann. Making PRIME Usable. In *Symposium on Usable Privacy and Security*, Carnegie Mellon University, Pittsburgh, PA, USA, July 2005. Carnegie Mellon University.
21. Irene Pollach. What's Wrong with Online Privacy Policies? In *Communications of the ACM archive*, volume 50, pages 103 – 108. ACM Press, New York, NY, USA, September 2007.
22. W3C. Platform for Privacy Preferences, April 2002. Online available at http://www.w3.org/TR/P3P/.
23. Samuel D. Warren and Louis D. Brandeis. *The Right to Privacy*, volume Vol. 4 of *pp. 193-220.* Harvard Law Review, Dec. 15 1890. doi:10.2307/1321160.
24. Alain Westin. *Privacy and Freedom.* Atheneum, New York, 1967.
25. Alan F. Westin and Center for Social and Legal Research. Bibliography of surveys of the U.S. Public, 1970-2003. *Hackensack, NJ: CSLR*, June 2003.
26. Alan F. Westin and HARRIS LOUIS & ASSOCIATES. Harris-Equifax Consumer Privacy Survey. *Tech. rep.*, 1991. Conducted for Equifax Inc. 1,255 adults of the U.S. public.