Privacy-awareness information for web forums: Results from an empirical study

Conference Paper · January 2010

DOI: 10.1145/1868914.1868957 · Source: DBLP

3 authors, including:

Peter Wolkerstorfer
CURE - Center for Usability Research and Eng…
21 PUBLICATIONS **192** CITATIONS

SEE PROFILE

Cornelia Graf
CURE - Center for Usability Research and Eng…
11 PUBLICATIONS **46** CITATIONS

SEE PROFILE

Privacy-Awareness Information for Web Forums: Results from an Empirical Study

Stefanie Pötzsch
Technische Universität Dresden, Faculty of Computer Science
01062 Dresden, Germany
stefanie.poetzsch@tu-dresden.de

Peter Wolkerstorfer
Center for Usablitity Research and Engineering
1110 Vienna, Austria
wolkerstorfer@cure.at

Cornelia Graf
Center for Usablitity Research and Engineering
1110 Vienna, Austria
graf@cure.at

ABSTRACT

While interacting with others on the internet, users share a lot of personal data with a potentially large but "invisible" audience. An important issue is maintaining control over personal data and therefore, in the first place, users need to be aware to whom they are disclosing which data. Based on the cues-filtered-out theory we introduce a new feature to support the privacy-awareness of forum users and tested it with 313 users. The results of our empirical study show that the presentation of privacy-related context cues indeed increases forum users' privacy-awareness. This is an important precondition for users' willingness to modify privacy settings or to use privacy-enhancing technologies.

ACM Classification Keywords

H.1.2 Models and Principles: User/Machine Systems—*Human factors*; H.5.2 Information Interfaces and Presentation: User Interfaces

General Terms

Privacy, Privacy Awareness, User-Centred Design, Empirical Study, Forum, Social Software, Social Web

INTRODUCTION

The success of the social web is based on the active participation of users and their willingness to contribute to the creation and improving of contents on the internet by sharing data and knowledge. By using social software, a lot of personal data is disclosed either directly, e.g., real name and date of birth on social networking sites, or indirectly, e.g., through editing specific topics in a wiki, commenting on blog entries or posting statements in a forum [9, 7]. The possibilities of the social web may enrich people's life, however there are also privacy risks involed.

NordiCHI 2010, October 16 - 20, 2010, Reykjavik, Island.

A German newspaper reported about the case of a woman who runs a restaurant and was looking for help in terms of an online gambling issue. Therefore, she became member of a forum and posted a lot of personal details in this forum over time. With all this information publicly available, the woman attracted the attention of a cyberstalker who continued to spy on her and posted a lot of further allegations about her. Some of the stalker's statements may be true, others are wrong, but in neither case the woman would have wanted this details published on the internet. The story went on for four years when the owners of the restaurant, which this woman was running, found some of the negative things about her online. For fear of a negative image of the restaurant, the owners finally canceled the contract with the woman, who some months later no longer even dares to have a nameplate on her front door [3]. This is only one of the extreme examples that appear in the media from time to time and all of these examples illustrate that the sharing of (too much) personal data with possibly millions of unknown people on the internet is a critical point from a privacy perspective and may result in negative consequences, like e. g., bullying, cyberstalking, harassment or identity theft.

Yet, social software also has positive sides: It enables its users to get in contact with like-minded people anytime, anywhere, no matter which kind of issues they like to discuss. The mutual exchange of information, both personal and non-personal, is the major feature of social software and the motivation why people use it. Further, from a social perspective, the exchange of implicit and explicit personal data allows people to get an impression of the potential interaction partners and their situations. In this sense, the disclosure of personal data contributes to the success of social interactions and the forming of communities [4]. This means, that users inevitably have to give up their anonymity to a certain extent to become part of the social web. However, they should not need to give up control over their personal data. Therefore, in the first place users need to be *aware* to whom they are disclosing which data and it is an interesting challenge for the HCI community to support users in making informed decisions whether and to which extent they disclose their personal data in social software.

It needs to be considered that there is an essential difference between *actual privacy* that is realised and secured by tech-

nical means on the one hand, and *perceived privacy* as an individual feeling of users on the other hand. Individuals decide and behave based on a subjective evaluation of their environment instead of collecting and analysing precisely all objective facts (cf. the *theory of bounded rationality* [15, 2]). Thus, our research about user-centred design and human behaviour in social software primarily focuses on the individual's perception of privacy.

In the next section we point out the importance of perceived privacy in computer-mediated communication by referring to results from related research. Then we describe our approach for the integration of privacy-awareness information in web forums. This feature was empirically tested with 313 users. The hypotheses and the design of the study are explained in the following section. Descriptive statistics and findings of the user study are discussed subsequently. We conclude the paper with a brief summary of the results.

RELATED WORK

When asked whether they care about privacy and want their personal data to be protected, a lot of people indicate a privacy-aware attitude in general. Regarding their actions in concrete situations, the majority shows another behaviour [13, 1, 16]. The complex reasons for this *privacy paradox* are object of research from psychological, social and economic perspectives. A possible explanation is that, although caring for privacy, users "blind out" their privacy concerns after a while when they enjoy their computer-mediated communication and that people overestimate their level of privacy in online interactions [18]. In an early study, Sproull and Kiesler found egalitarian and deregulating effects of computer-mediated communication in comparison to face-to-face communication, e.g., the disclosure of personal data to a large audience within a company [17]. In a meta-analysis of 39 studies on self-disclosure in interviews, it was shown that subjects disclose more personal data in computer forms than they do in face-to-face interviews [22]. The finding was explained with a higher level of perceived privacy of the participants in the first case. Similar results are reported from a comparison about spontaneous self-disclosure in face-to-face situations and computer-mediated communication scenarios [10]. The researchers confirmed a positive correlation between individuals' self-disclosure and visual anonymity as one aspect that contributes to perceived privacy. Further studies about the privacy attitude and behaviour of Facebook users reveal that they do not underestimate privacy threats in general, however, they misjudge the extent of accessibility of their personal data [1, 19]. Altogether, these results suggest that the individual's perceived level of privacy is a very important factor for the decision whether to disclose personal data on the internet. Therefore, users need to be (made) continuously aware which – actively or passively created – personal data is visible to whom.

PRIVACY-AWARENESS INFORMATION

The cues-filtered-out approach implies that individuals are more lavish regarding the disclosure of personal data when they use computer-mediated communication technology than in face-to-face settings due to a lack of social and context cues [6, 17]. Therefore, we propose to provide additional privacy-related cues in social software in order to raise users' privacy awareness, help them to better assess their potential audience and eventually enable them to make better informed decisions whether to disclose personal data on the internet. Considering that privacy is only a secondary task for users, the privacy-awareness information should be easy and quick to understand and not hinder social interactions and communication as primary tasks in social software.

Recent research about privacy and the social web has especially focussed on social networking sites [1, 5, 18, 20] where users maintain explicit profiles. Other applications where personal data is mainly implicitly included in the user-generated content have been widely neglected. However, protecting privacy is also an important issue in these applications and therefore we decided to use the example of a web forum to conceptualise and implement a feature for the presentation of privacy-awareness information. Obviously, manifold social and context cues are conceivable to enrich computer-mediated communication in general and web forums in particular. Our goal is to find and test *practical* privacy-awareness information on real forum users' perception of privacy. We do not aim to identify the most salient form of social cues. Thus, we distinguished between the following two different types of privacy-awareness information: First, the display of the potential audience replaces partly the missing context cues from communication partners and should remind users about the actual audience of their contributions. In face-to-face communication, individuals usually can see who hears them speaking, even if a large part of the audience does not actively take part in the conversation. In a forum (or other social software), it is impossible to ascertain the whole audience and users may especially forget about the mass of silent readers [5]. According to the 90-9-1 rule, 90 % of all users of an online community are lurkers who never contribute, 9 % are members who post from time to time and only 1 % are considered as regular active users [12]. This implies that the number of visitors of a post in a forum is on average ten-times higher than the number of the active forum members, which – at least in theory – could be known to the author of a post.

In our privacy-awareness information panel the display of the potential audience is realised, first, by a textual cue which says that a forum post is visible to all internet users and, second, by the indication of the concrete number of visitors of the forum within the last seven days. The second direction for privacy-awareness information aims at demonstrating users how *not anonymous* they actually are. Considering that individuals usually visit a forum by following a link on the internet, the IP address is a good context cue: it is highly individual, it reminds the user that she is less anonymous than often perceived, and it can be correctly shown to each user. For similar reasons we also included a textual cue about a user's current location (city) in addition. The location can be derived from the IP address by using a geolocation database. It is less individual identifying then an IP address, however probably more understandable for those people who are not very familiar with technical terms.

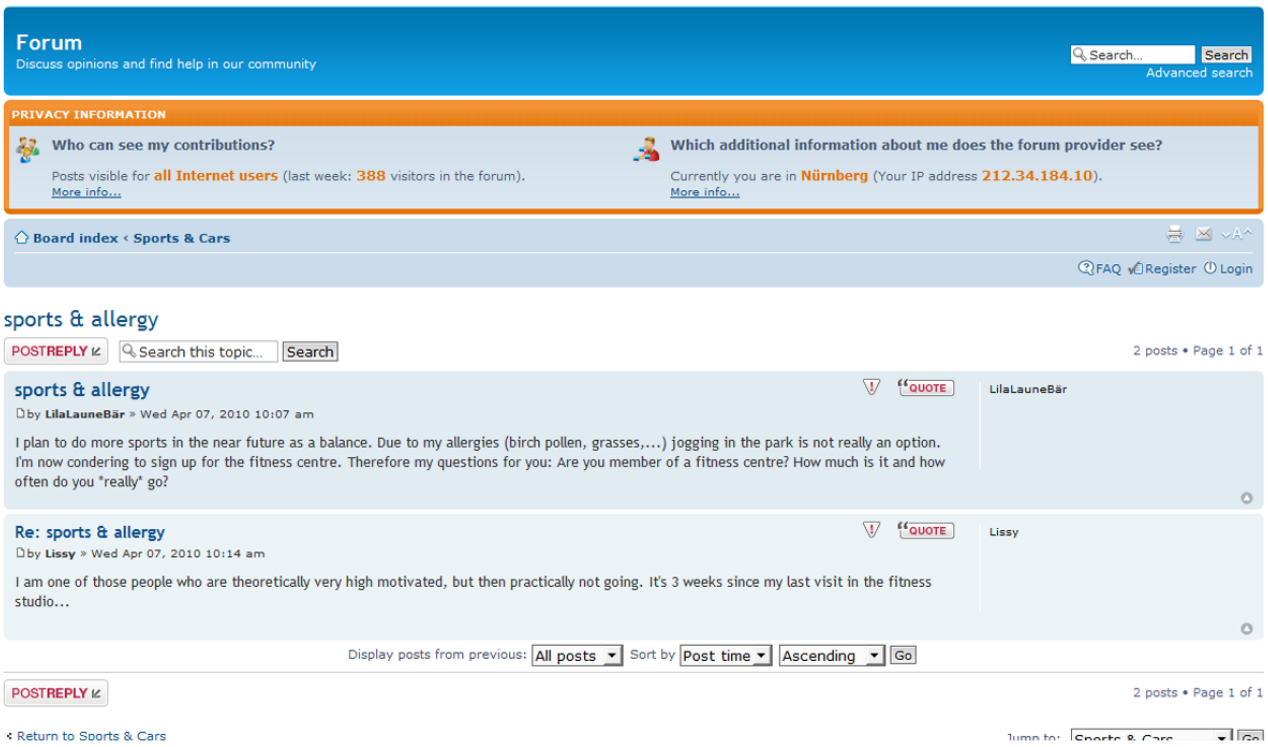


Figure 1. Forum with privacy-awareness information (also available in German)

For the graphical presentation of the privacy-awareness information we followed well-accepted design guidelines [11] and used a box in a complementary colour (orange) placed on top of the forum (blue). The cues about the potential audience are presented on the left hand and are visually separated from the cues about the additional information that the provider receives. This is displayed on the right hand. Figure 1 shows an example of a forum with our privacy-awareness information panel.

Table 1. 2 × 2 design of the study

	Numerical representation	
Textual representation	no	yes
no	CG	ExG_1
yes	ExG_2	ExG_3

EMPIRICAL STUDY

In order to test the validity of our assumption that the presentation of privacy-awareness information as additional context cues influences individuals' perception of privacy, we conducted an online study with real forum users. In line with the arguments above, we expect:

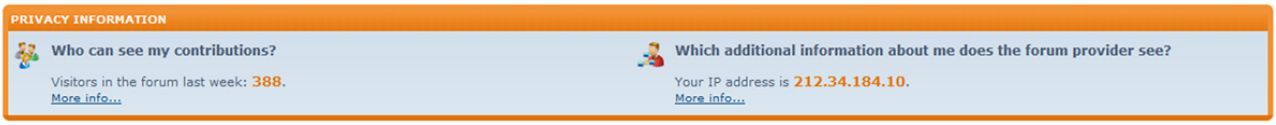
H1 – Participants who are provided with privacy-awareness information feel they have less privacy during their visit of the forum than participants from a control group.

We further wanted to study whether there are differences in the effect of single privacy-awareness information according to their form of presentation. Considering that most information in the user interface of the forum is presented as text, we assume that the numerical cues in the privacy-awareness information panel stand out and will be perceived better. Here we also have to control for participants' familiarity with technical terms, i. e., whether they actually know what an IP address is.

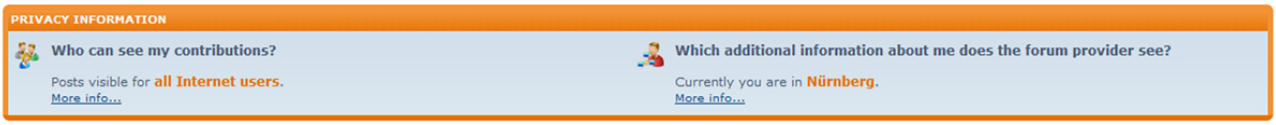
H2 – Numerical privacy-awareness information will show a stronger effect than textual privacy-awareness information.

The concrete objective of the privacy-awareness information is to continuously remind users about the potential audience if they contribute (personal) data to a forum. The privacy-awareness information should also inform users about implicitly submitted data, which is known to the forum provider in addition to the actual contribution. Since users do not actively enter such data, e. g., their IP address or current location, in a web form and confirm its submission, they may especially not be aware of the fact that it is known by the a provider anyhow.

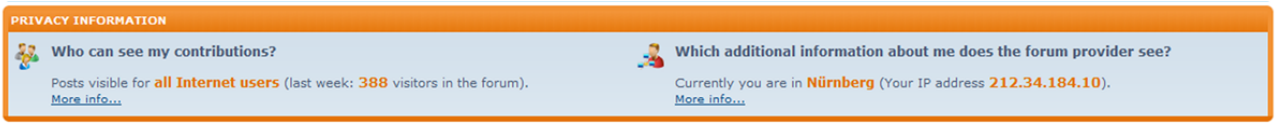
H3 – Participants who are provided with privacy-awareness information will be better informed about the potential audience of forum contributions and they will also be better informed about additional information that they implicitly transfer to the forum provider.



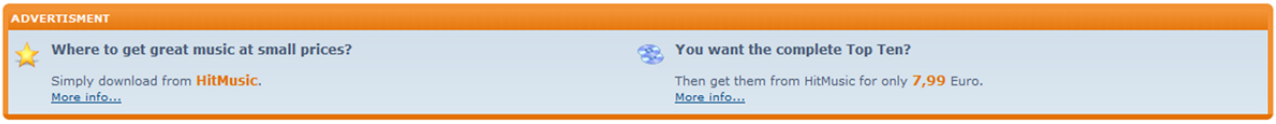
(a) Experimental group ExG_1 (numerical cues)



(b) Experimental group ExG_2 (textual cues)



(c) Experimental group ExG_3 (textual and numerical cues)



(d) Control group CG (advertisment)

Figure 2. User interface for experimental groups and control group (originally shown in German)

The study was a 2 by 2 design that distinguished two forms of presentation of comparable privacy-awareness information as independent variables: textual cues vs numerical cues (Table 1). As textual information the potential audience and the user's current location were displayed. As numerical representation we showed the exact number of visitors of the forum within the last week and the subject's IP address. Participants in the control group got displayed an advertisement instead of any textual nor numerical privacy-awareness information (Figure 2). Subjects of the study were randomly assigned either to one of the three experimental groups (ExG_i) or to the control group (CG).

We invited people to participate in the study by posting a link via mailing lists, forums and blogs on the internet. If they were interested they could follow this link and use their own computers in their familiar surroundings to do the study. To avoid bias towards privacy, we told them that we are studying usability aspects of a web forum. Further, we did not post the invitation on platforms with a special focus on privacy and data security. internet users were motivated to participate by the offer to win several vouchers for an online store after finishing the study.

The study, that was originally realised in German language, consisted of two parts. First, participants saw a realistic full-screen screenshot of the forum including the orange box on top and two posts in a discussion thread about leisure-time physical activity (see Figure 1). We instructed the subjects to imagine that they are the author of the first contribution. The orange box either contained one of the three privacy-awareness information panels or the advertisement (Figure 2). The rest of graphical user interface was identical for each study group. In the following second part, subjects filled in a well-elaborated online questionnaire. In the questionnaire we asked about their perceived level of privacy during the visit of the forum, about their knowledge of technical- and privacy-related terms and their internet use in general. We also collected demographic data.

DESCRIPTIVE ANALYSIS

327 participants completed the online study between mid of December 2009 and mid of January 2010. We excluded answers from those who ticked off that they have not seriously answered the questionnaire and also from those who needed less than four minutes to complete the whole study, because it was not possible to seriously answer all questions within such a short time frame. Since we want to test the effect of privacy-awareness information on forum users, we further did not consider answers from participants who stated that they have never even read in a forum. Altogether, 313 valid responses remain for further analysis.

Since the study was available in German language, our participants are mainly located in Austria (51 %) and Germany (44 %). The majority of 85 % has A-levels or a higher educational achievement. 45 % are employed persons and 42 % of the participants are students. An almost equal number of male (50.3 %) and female (49.7 %) subjects filled in the questionnaire.

The participants of our study are experienced internet users. The majority of 88 % indicates that they use the internet for five years or longer. Regarding their daily use, we can identify three approximately equally large groups. A first group of 29 % of all participants can be labelled *occasional users*

Table 2. Knowledge of technical aspects of the internet

Do you know...	Yes / Correct	No / Wrong
...what an *internet browser* is?	97.76 %	2.24 %
...what an *IP address* is?	86.90 %	13.10 %
Give short explanation	85.30 %	14.70 %
...how to *change your IP address*?	52.40 %	47.60 %
Give short explanation	48.88 %	51.12 %

n=313 (100 %)

Table 3. Linkability by nickname

When you contribute to more than one forum, do you use the same nickname in more than one forum?	
Always different nicknames	24.28 %
Same nickname in more than one forum	38.66 %
No nickname (as anonymous / guest user)	4.47 %
Not contributed to more than one forum	6.39 %
Never contributed to any forum	26.20 %

n=313 (100 %)

who are online two hours a day at most. Second, we have 37 % *normal users* who surf the internet three to four hours a day. A third group of 33 % can be considered as *frequent users* since they spent at least five hours a day on the internet. We further asked participants about their technical knowledge. Results in Table 2 show that nearly all subject claim to know what an internet browser and an IP address is and that almost all of them were actually able to explain the second item with a few words. When it comes to the question how to change the own IP address, the ratio of participants who claim to know that and could also explain how to do it decreases to about 50 %, which is still quite good. These numbers indicate that the participants of our study are not that clueless when it comes to key technical terms related to internet usage as someone might think.

As said previously, all participants whose answers are considered in the analysis read in web forums at least from time to time. Approximately three quarters of the subjects have also actively contributed to one or more forums and about one quarter claims to always use different nicknames for different forums which can be interpreted as very privacy-aware behaviour (Table 3). However, a huge part uses the same nickname in more than one forum which allows for linkability of the information that is provided across different applications and therefore can be considered as privacy-intrusive behaviour. When directly asked whether they care about privacy and the protection of their personal data on the internet, 89 % stated that they often or always do so. This means, among the participants of our study we see a discrepancy between the stated attitude and the actual behaviour with regard to privacy. This can be interpreted as evidence for the privacy paradox (cf. the section on related work).

Table 4. Detailed reasons to use forums

Inform myself about different topics / products	93.29 %
Ask for advice / help	56.87 %
Share my knowledge	37.38 %
Pass the time	32.91 %
Discuss own view on topics with others	31.63 %
Report about own experiences	25.56 %
Stay in touch with others	21.73 %
Follow what happens in the life of others	13.74 %
Get to know new people	9.90 %
Use anonymity of the internet for communication	8.63 %
Friends outside the internet have no time	5.11 %
Showing off	3.51 %

multiple answers have been possible n=313 (100 %)

Those 26.20 % (=82 participants) who claim to never have written a forum post were asked in a free text field about the reasons for not contributing. Besides a lack of time or interest, 20 % of the 82 indicated privacy concerns. This number underlines that forum providers should have an interest to develop privacy features for their users.

Selection of quotations from participants who indicate privacy reasons for not contributing to forums (originally posted in German):

> **Q1** (from CG): "I do not feel secure, when I post something on public web sites."
> **Q2** (from ExG_2): "I don't want to reveal so much about me. Besides, I can get the information I am looking for somewhere else. If a need advise, I ask real people."
> **Q3** (from ExG_3): "Not interesting for me, I don't want to reveal information about me all over the internet where everybody may read it."

Table 4 lists different reasons why participants of our study use forums in general. We see that informing oneself about different topics and/or products, i. e., passively consuming information is the primary reason. However also the sharing of a user's own knowledge, the discussion of own views and the reporting about own experiences are often stated motivations. All of the latter three reasons imply the disclosure of personal data.

RESULTS

To the best of our knowledge, at the time of the study there is no established scale available to measure people's individual *perceived privacy*. However, in order to make participants' perception of their own privacy in the forum – our dependent variable – comparable, we created the *Perceived Privacy Index (PPX)*. Since privacy in general is a very ambiguous term, we decided to use four specific questions, namely participants were asked to specify how public and how private their contribution in the forum is and how anonymous and how identifiable they have felt during their visit of the fo-

Table 5. Perceived Privacy Index for different groups according to the presented privacy-awareness information

	PPX			
	Min	Mean	Max	ANOVA (F-test)
Num, text or both	0	133.59	280	$F(1,311)=13.36$; $p=0.00^{***}$
others	0	161.63	292	
Num or both	0	128.27	266	$F(1,311)=12.46$; $p=0.00^{***}$
others	0	151.76	292	
Text or both	0	133.96	280	$F(1,311)=4.08$; $p=0.04^{*}$
others	0	147.58	292	
Both	0	123.83	248	$F(1,311)=7.89$; $p=0.01^{**}$
others	0	145.85	292	

sign. levels: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 6. Regression models for Perceived Privacy Index

PPX	*Model 1, n= 313*			*Model 2, n= 313*		
(dependent v.)	Est	StdEr	p	Est	StdEr	p
Intercept	159.16	5.73		161.63	6.62	
Predictors						
Num or both	−23.78	6.62	0.00^{***}	−28.86	9.49	0.00^{**}
Text or both	−14.12	6.61	0.03^{*}	−18.83	9.15	0.04^{*}
Both				9.88	13.25	0.46

sign. levels: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

rum. Each of the four items was measured on a 0 to 100% slider scale. Then, the PPX is calculated using Equation 1. The higher the PPX value, the more private a subject has felt.

$$PPX = (100 - public) + private + anonymous + (100 - identifiable) \quad \text{(Equation 1)}$$

The results, which are are listed in Table 5, clearly support hypothesis H1. The PPX is significantly lower, i. e., subjects feel they have less privacy, when textual, numerical or both types of privacy-awareness information are available. To further disentangle the effect of the numerical cues vs the effect of the textual cues, we used linear regression models. Table 6 shows that both kinds of cues significantly decrease participants' perceived privacy and that hypothesis H2 can also be confirmed since the effect is indeed stronger – for both, estimate value and level of significance – for numerical cues. Thereby, from Table 2 we know that most of the participants really have a concrete idea of what an IP address is. We further see that there is no additional effect if both kinds of information, textual and numerical, are presented together as can be seen in model 2.

Regarding hypothesis H3 the picture is less clear. In order to test whether participants are aware of the potential audience, we asked them which of the four groups in Table 7 really have access to the forum post. We further asked which of these groups they would *intend* to have access. Actually, the post that we showed to the participants was accessible for all people with access to the internet, i. e., all registered and unregistered users, the forum provider and the internet provider. The fact that the post from the example was completely public could be learnt from the privacy-awareness panel with textual information (if shown) and there was also a visual cue visible for participants of all study groups indicating that the post can be viewed without being logged in, which means it is visible for everybody with internet access. We found no statistical evidence that any of the presented privacy-awareness information leads to better informed users with regard to the question which groups really have access (see rows *expected* in Table 7). The comparison of the percentages of expected access vs intended access of different audience groups reveals that nearly all participants know about and agree with the access to all post for registered members. Also nearly all participants know that the forum provider has access and three-quarters stated that the forum provider should have access. Our results further show that a majority of participants knows that also unregistered visitors can see the post, however only about one-third would want unregistered people to view their posts. This means, there is a considerable difference between the percentage of participants who would let registered users read their posts and those who also would allow unregistered users access to their posts. This finding is very interesting considering the fact that in most forums on the internet anybody can easily become a registered member by providing a fake e-mail address and choosing a password. Thus, practically each internet user could have access in any case with no great effort.

Furthermore, in an additional free text field, a dozen of the subjects said that they would like to authorise particular readers based on properties (e. g., others with more than ten posts) or based on their relationship to them (e. g., friends from other communities). An approach that addresses the idea of user-controlled, property-based access control for forums is discussed in [14]. The authors further argue that access control only based on relationships would not be suitable for forums in general since this requires that the author of a post and the users she wants to give access have to know each other before. This assumption does not hold for web

Table 7. Expected and intended groups of people to access forum posts

	Advert. n=78	Num, text or both n=235	Num or both n=149	Text or both n=161	Both n=75
All registered users					
expected	96.15 %	96.60 %	97.32 %	96.27 %	97.33 %
intended	89.74 %	97.45 %	97.99 %	96.27 %	96.00 %
Unregistered users					
expected	69.23 %	73.19 %	74.50 %	74.53 %	78.67 %
intended	28.21 %	31.08 %	33.56 %	31.68 %	36.00 %
Forum provider					
expected	98.72 %	95.32 %	95.30 %	95.03 %	94.67 %
intended	66.67 %	75.68 %	73.15 %	73.29 %	70.67 %
Internet provider					
expected	47.44 %	50.21 %	48.99 %	51.55 %	50.67 %
intended	7.69 %	11.91 %	11.41 %	12.42 %	12.00 %

forums, where people with similar interests can meet and discuss without knowing each other in person.

We further asked subjects to name a concrete number how many people *theoretically* could access their post. Obviously there is no clearly correct answer to this question, however we were especially interested to see whether the display of privacy-awareness information about the audience leads to an increased estimated number compared to the number named by the control group. Considering a possible *anchor effect* [21] for the two experimental groups with the numerical privacy-awareness information, we further wanted to check whether the single information that posts in the forum are "visible for all internet users" (ExG_2) leads to a higher variance in the answers in comparison to the cases when the information that the forum had 388 visitors last week is given (in addion) (ExG_1, ExG_3). In the former case, the formulation leaves it completely up to the user to imagine a number of how much people "all internet users" might be, whereas in the latter case a number is already presented and may function as rough orientation. The answers, which are graphically shown in Figure 3 for each study group and on a logarithmic scale, support the assumption about an anchor effect. Though the medians (black lines through the boxes) are roughly on the same level for all groups, in the box plot for ExG_2 (*Text only*), we see that the actual box, which depicts the interquartile range[1], is considerably longer and also the range between minimum and maximum values (without outliers) is much greater compared to all other groups. We cannot see this phenomenon if no privacy-awareness information is presented and also not if a concrete number is given. This means, if no concrete number is provided but a textual information about all internet users, at least some forum users do really imagine that not only hundreds but millions ($> 2^{20}$) of people theoretically can visit their contribution.

[1]The interquartile range is the difference between the third and first quartiles, i. e., the "middle" fifty percent of all answers.

Finally, we tested whether participants are aware of the information which they directly and indirectly disclose to the forum provider. Therefore we showed participants a list with specified information items. We asked them to decide for each of the items whether a provider knows this piece of data about them after they have visited the forum. An overview about the answers is given in Table 8. Actually all of the listed facts are known to the provider. Since the IP address is the item that most of the participants are aware of anyway, we found no significant difference between the experimental groups and the control group. The knowledge that the provider can infer the location of the user is lower in general and it increases significantly if privacy-awareness information are shown. The according regression model in Table 9 shows that really only the textual cue increases the knowledge, i. e., users who only were informed about the IP address (ExG_1) do not conclude that this also conveys information about their location. Among all study groups, the majority of users is sure that the forum provider knows which posts in the forum they have visited, i. e., they are aware that their behaviour in the forum is monitored. Interestingly at least half of the participants of each group were sure that the provider does not know anything about them which can be learnt from the content of their contributions, which in our example was something about the author's pollen allergy, i. e., health information. Yet, this percentages have to be regarded with caution since some participants may just have forgotten what was written in "their" post. On the other hand, users may also forget about real forum posts after a while and which personal data they have disclosed there.

We also asked participants whether a forum provider knows which browser and operating system they have on their computers. This hint was not included in the presented privacy-awareness information and the answers indicate that there is a considerable share of participants who believe that the provider does not know this kind of information or who are not sure about this question. In fact, forum providers - like all other web site operators - are able to read out the so-called

Table 8. Users' knowledge about data items known to forum providers

	Advert. n=78	Num, text or both n=235	Num or both n=149	Text or both n=161	Both n=75
IP address					
known by forum provider	85.90 %	91.49 %	92.62 %	91.30 %	93.33 %
not known	7.69 %	2.55 %	4.65 %	3.11 %	1.33 %
unsure / no answer	6.41 %	5.96 %	6.04 %	5.59 %	5.33 %
Contingency test $\chi^2(2)$=4.26, p=0.12					
Location					
known by forum provider	41.03 %	57.45 %	53.02 %	65.22 %	65.33 %
not known	33.33 %	21.70 %	23.49 %	18.63 %	18.67 %
unsure / no answer	25.64 %	20.85 %	23.49 %	16.15 %	16.00 %
Contingency test $\chi^2(2)$=6.79, p=0.03$*$					
Posts visited					
known by forum provider	70.51 %	68.94 %	69.80 %	67.08 %	66.67 %
not known	15.38 %	11.91 %	10.74 %	13.66 %	13.33 %
unsure / no answer	14.10 %	19.15 %	19.46 %	19.25 %	20.00 %
Contingency test $\chi^2(2)$=1.41, p=0.50					
Content of own post *(health information in the example)*					
known by forum provider	41.03 %	41.70 %	36.91 %	43.48 %	36.00 %
not known	50.00 %	53.62 %	57.05 %	51.55 %	56.00 %
unsure / no answer	8.97 %	4.68 %	6.04 %	4.97 %	8.00 %
Contingency test $\chi^2(2)$=2.03, p=0.36					
Browser					
known by forum provider	65.38 %	59.15 %	59.06 %	56.52 %	53.33 %
not known	19.23 %	19.57 %	19.46 %	22.36 %	25.33 %
unsure / no answer	15.38 %	21.28 %	21.48 %	21.12 %	21.33 %
Contingency test $\chi^2(2)$=1.41, p=0.50					
Operating system					
known by forum provider	47.44 %	44.68 %	44.97 %	43.48 %	42.67 %
not known	34.62 %	31.06 %	30.87 %	33.54 %	36.00 %
unsure / no answer	17.95 %	24.26 %	24.16 %	22.98 %	21.33 %
Contingency test $\chi^2(2)$=1.36, p=0.51					

sign. levels: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

User-Agent string and therefore know which browser and operating system users run on their computers. Furthermore, considering details such as version numbers or installed plug-ins, the fingerprint of the web browser can be enough to re-identify a user [8]. There are settings and tools available to blur the information of the User-Agent string and to increase users' anonymity. However, before users will apply this privacy-enhancing technologies, they need to be aware about the facts.

CONCLUSION

In this paper we have shown with empirical evidence that the presentation of privacy-related context cues promotes forum users' privacy-awareness. This effect is found regardless of whether numerical, textual or both kinds of privacy-awareness information are presented and is even stronger for numerical cues. We also showed that privacy-awareness information enhances users' knowledge about personal data that can be inferred by forum providers and that textual cues tend to stimulate users' imagination of how many visitors potentially can see their contribution. Future research about privacy-awareness and user-centred design will include the transfer of the concept to other social software, such as wikis or (micro)-blogs. Then, further user studies will contribute to a more general understanding of the effect that single privacy-related context cues have on the perceived privacy of users and their self-disclosing behaviour across different types of social software. Having gained a better understanding, it will then be reasonable and very interesting to deploy a privacy-awareness tool in a real social software application and study the effect of the privacy-awareness cues on the long-term.

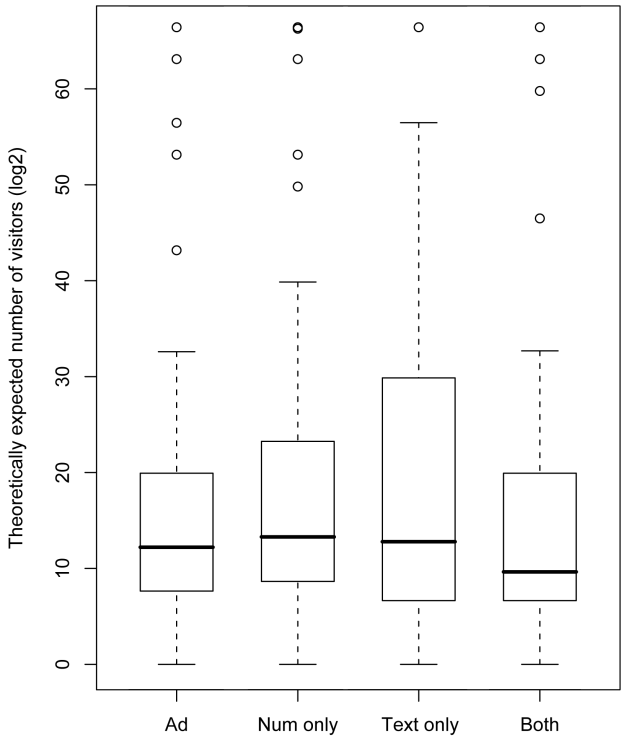


Figure 3. Comparision of theoretically expected number of visitors by the four study groups

Table 9. Regression model for location

Location known by forum provider = *true* (dependent var.)	*n= 235* Est	Std Er	p
Intercept	0.41	0.06	
Predictors			
Numbers or both	−0.00	0.08	0.95
Texts or both	0.24	0.08	0.00**
Both	0.01	0.11	0.95

sign. levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Even without privacy-awareness information users may know and accept that all information which they explicitly contribute to a forum or other social software application is public, potentially for a broad audience and for a long time period. However, the question about implicitly submitted data, which are at least known to the application providers, remains. Since users do not actively enter this data and confirm the transmission, they are not aware of the fact, that providers know, for instance, their current location or may reidentify them because of their web browser details. Privacy-awareness is important since it is an essential precondition for the use of privacy-enhancing technologies. Using these technologies will help to protect users' privacy, however it is also accompanied with extra costs. Only if users are (made) privacy-aware, they will see a clear need for these technologies and they finally might be willing to spend associated additional costs. These costs range from extra time and cognitive effort that is needed to pay attention to privacy hints, over accepting possible usability drawbacks and give up the convenience of personalisation when surfing on the internet in a privacy-friendly way, to extra money that needs to be paid for enhanced privacy and security features. It remains object of further research to precisely analyse and quantify the trade off between privacy and extra costs that privacy-aware users are willing to accept in different use cases.

ACKNOWLEDGEMENTS

Many thanks are due to Christina Köffel for her excellent research assistance. We also thank all participants of our study. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 216483.

REFERENCES

1. ACQUISTI, A., AND GROSS, R. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *Proceedings of 6th Workshop on Privacy Enhancing Technologies* (Cambridge, UK, 2006), pp. 36–58.

2. ACQUISTI, A., AND GROSSKLAGS, J. Privacy and rationality in individual decision making. *IEEE Security and Privacy 3*, 1 (2005), 26–33.

3. BURGER, J. Lügnerin! Betrügerin! `http://www.zeit.de/2009/53/Internetmobbing?page=all`, December 2009.

4. CUTLER, R. H. Distributed presence and community in cyberspace. *Interpersonal Computer and Technology 3*, 2 (1995), 12–32.

5. DANAH BOYD. Why youth (heart) social network sites: The role of networked publics in teenage social life. In *MacArthur Foundation Series on Digital Learning - Youth, Identity, and Digital Media Volume*, D. Buckingham, Ed. MIT Press, Cambridge, MA, 2007.

6. DÖRING, N. Reduced social cues / cues filtered out. In *Medienpsychologie. Schlüsselbegriffe und Konzepte* (Stuttgart, 2008), N. C. Krämer, S. Schwan, D. Unz, and M. Suckfüll, Eds., Kohlhammer, pp. 290–297.

7. EBERSBACH, A., GLASER, M., AND HEIGL, R. *Social Web*, vol. 3065 of *UTB*. UVK, Konstanz, 2008.

8. ECKERSLEY, P. A primer on information theory and privacy. `https://www.eff.org/deeplinks/2010/01/primer-information-theory-and-privacy`, January 2010.

9. GROSS, R., ACQUISTI, A., AND HEINZ, III, H. J. Information revelation and privacy in online social networks. In *WPES '05: Proceedings of the 2005 ACM workshop on Privacy in the electronic society* (2005), pp. 71–80.

10. JOINSON, A. N. Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology 31* (2001), 177–192.

11. KOYANI, S. J., BAILEY, R. W., AND NALL, J. R. *Research-Based Web Design & Usability Guidelines*. Computer Psychology, 2004.

12. NIELSEN, J. Participation inequality. `http://www.useit.com/alertbox/participation_inequality.html`, October 2006.

13. NORBERG, P. A., HORNE, D. R., AND HORNE, D. A. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs 41*, 1 (2007), 100–126.

14. PÖTZSCH, S., AND BORCEA-PFITZMANN, K. Privacy-respecting access control in collaborative workspaces. In *Privacy and Identity, IFIP AICT 320* (Nice, France, 2010), M. B. et al., Ed., Springer, pp. 102–111.

15. SIMON, H. A. *Empirically Grounded Economic Reason*, vol. 3 of *Models of Bounded Rationality*. MIT Press, Cambridge, MA, 1997.

16. SPIEKERMANN, S., GROSSKLAGS, J., AND BERENDT, B. E-privacy in 2nd generation e-commerce: privacy preferences versus actual behavior. In *EC '01: Proceedings of the 3rd ACM conference on Electronic Commerce* (New York, NY, USA, 2001), ACM, pp. 38–47.

17. SPROULL, L., AND KIESLER, S. Reducing social context cues: Electronic mail in organizational communications. *Management Science 32*, 11 (1986), 1492–1512.

18. STRATER, K., AND LIPFORD, H. R. Strategies and struggles with privacy in an online social networking community. In *BCS HCI (1)* (2008), D. England, Ed., BCS, pp. 111–119.

19. STRATER, K., AND RICHTER, H. Examining privacy and disclosure in a social networking community. In *SOUPS '07: Proceedings of the 3rd symposium on Usable privacy and security* (New York, NY, USA, 2007), ACM, pp. 157–158.

20. STUTZMAN, F., AND DUFFIELD, J. K. Friends only: examining a privacy-enhancing behavior in facebook. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems* (New York, NY, USA, April 2010), ACM, pp. 1553–1562.

21. TVERSKY, A., AND KAHNEMAN, D. Judgment under uncertainty: heuristics and biases. *Science 185* (1974), 1124–1131.

22. WEISBAND, S., AND KIESLER, S. Self disclosure on computer forms: meta-analysis and implications. In *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Common Ground.* (New York, 1996), ACM, pp. 3–10.