Stated Privacy Preferences versus Actual Behaviour in EC environments: a Reality Check

Sarah Spiekermann
Humboldt University Berlin
Institute of Information Systems
Spandauer Straße 1
D-10781 Berlin
Germany
sspiek@wiwi.hu-berlin.de

Jens Grossklags
Humboldt University Berlin
Institute of Information Systems
Spandauer Straße 1
D-10781 Berlin
Germany
jensg@wiwi.hu-berlin.de

Bettina Berendt
Humboldt Universität zu Berlin
Institute of Pedagogy and Informatics
Geschwister-Scholl-Str. 7
D-10099 Berlin
Germany
berendt@educat.hu-berlin.de

Abstract:
As electronic commerce environments become more and more interactive, privacy is a matter of increasing concern. Many surveys have investigated households' privacy attitudes and concerns, revealing a general desire among Internet users to protect their privacy. To complement these questionnaire-based studies, we conducted an experiment in which we compared self-reported privacy preferences of 171 participants with their actual disclosing behavior during an online shopping episode. Our results suggest that current approaches to protect online users' privacy, such as EU data protection regulation or P3P, may face difficulties to do so effectively. This is due to their underlying assumption that people are not only privacy conscious, but will also act accordingly. In our study, most individuals stated that privacy was important to them, with concern centering on the disclosure of different aspects of personal information. However, regardless of their specific privacy concerns, most participants did not live up to their self-reported privacy preferences. As participants were drawn into the sales dialogue with an anthropomorphic 3-D shopping bot, they answered a majority of questions, even if these were highly personal. Moreover, different privacy statements had no effect on the amount of information disclosed; in fact, the mentioning of EU regulation seemed to cause a feeling of 'false security'. The results suggest that people appreciate highly communicative EC environments and forget privacy concerns once they are 'inside the Web'.

Keywords: *privacy, automated shopping and trading, legal issues, marketing and advertising technology, social implications, user interface and interaction design*

1. Introduction

Privacy is a hotly debated issue. It is at the center of the question who will have access to one of the online economy's major assets: customer data. Long-existing dreams of one-to-one marketing are close to coming true for marketers on the Internet. Through personalization, companies hope to considerably improve customer retention, to build up stronger competitive boundaries and to increase revenue through up selling and cross selling. Researchers in marketing, computer science, psychology and many other disciplines have therefore started to work in this direction, investigating opportunities inherent in agent technology [3,14,22,27,29], data mining [6,24], and interface design [9,17,23,25]. A core assumption is, however, the availability of reliable customer data. Without a sufficient base of such data all these current marketing visions cannot be realized. The problem is that at this point a conflict arises: While companies are thirsty for ever more information they undermine the fundamental right of informational self-determination.

Three fundamental approaches have evolved over the past decade addressing the privacy issue: ensuring privacy through law, through self-regulation, or through technical standards. European countries rely very much on the force of regulation. The problem with regulation is that laws take an average of 10 years to go into effect, while the life cycle of information and communication goods is only 3-7 months [7, p.286]. So regulation risks to always be behind the technology deployed. Also, law enforcement is a huge challenge, not only because European countries have difficulties creating and financing appropriate control institutions [5,21], but also because imposing their national data practices on super powers such as the US proves rather difficult.[1] The biggest problem of EU data protection law is that it propagates data collection parsimony [12] while the Internet is inherently a medium of 'data richness'. It also restricts the free trade of user data [12], although this asset has become one of the most valuable goods of the new economy [13], around which many business models are built [10]. As a result, it is questionable to what extent EU regulation will have the power to enforce its good visions practically.

The USA has pursued its traditionally more liberal approach of self-regulation. US companies are focusing more on the use of privacy statements and privacy seals on their e-commerce Web sites. The main underlying assumptions are that people are privacy conscious and that they trust published privacy statements and -seals. Many surveys have supported this view [1, 28, 19]. It is therefore argued that market forces will lead to the 'survival' of only those online companies that abide by acceptable privacy standards. The Platform for Privacy Preferences Project (P3P) which probably represents the best-supported privacy technology, is a product of this school of thought.[2] P3P will block access to Web sites or automatically notify the online user if a Web site's privacy statement does not correspond to his or her privacy preferences. The consumer is then left to decide whether he or she still wants to use the service. As most surveys gave evidence of online users privacy concerns, it is expected (and hoped?) that consumers will stop accessing sites that do not provide appropriate policies.

The problem is that the surveys conducted to prove users' privacy consciousness have only asked for attitudes, but never measured actual behavior. In particular, no observations exist on how consumers will react to promising benefits of highly interactive Web sites that offer individualized content as well as highly communicative and entertaining value. However, this is particularly interesting, because it will make it possible to anticipate the success of current initiatives to protect privacy and to generate ideas for valuable adjustments. The empirical study presented in this paper aims to fill this research gap by asking people for

[1] The renewed discussion of the 'Safe Harbor Policies' in the US show the difficulties of agreeing on international regulation.
[2] P3P is an initiative of the World Wide Web Consortium (W3C) in conjunction with many industry partners including Microsoft. For more information see: http://www.w3.org/P3P/

their privacy preferences and contrasting these claims with subsequent behavior during an online shopping trip.

We begin with a description of the experimental design and set-up. In section 3, we present selected results obtained from a first questionnaire on privacy attitudes and preferences and compare these attitudes with the self-disclosure displayed in communication with an anthropomorphic 3-D shopping bot that assisted participants in an online shopping trip for winter jackets and compact cameras. In the same section, we address the question whether different privacy attitudes lead to different navigational strategies. Here, we also comment on the influence of different privacy statements on behavior. Section 4 then comprises a critical discussion of current approaches to protect privacy and some suggestions on how to render them more effective in highly communicative and interactive online environments. Section 5 concludes with a summary of major findings and limitations of the study.

2. Method

The IWA experiment was carried out in winter 2000 with the goal of investigating drivers and impediments of online interaction.[3] Privacy concerns were regarded as one major impediment of truthful and deep online interaction. In investigating privacy we focused on two issues: First, we wanted to contrast self-reported privacy preferences with actual self-disclosing behavior. Second, we wanted to find out whether different privacy statements would impact interaction and disclosure.

The experiments were designed to observe participants during an online shopping trip for a compact camera or a winter jacket. Participants had to spend their own money if they chose to buy in the shop, but they were not forced to purchase anything. Before and after the shopping trip, they filled out a questionnaire.

In order to avoid information chunks and have people investigate products 'neutrally', no brand information was displayed, neither in verbal product descriptions nor on photographs.

2.1. Participants

206 participants registered to participate in the experiments and to shop for one of two products, a compact camera or a winter jacket. Their main incentive was a

[3] For more information on the IWA ("Interaction with Agents") experiments see: http://iwa.wiwi.hu-berlin.de

60% discount on the prices of all products offered in the experimental store.[4] 95% of the participants were students from different university faculties, while the remaining 5% participants held different jobs. 152 chose to shop for a camera, and 54 for a jacket.

2.2. Materials and Apparatus

The central material for the experiment was the online store. This was complemented by some additional, printed material which will be described in the "Procedure" section.

The online store was programmed for the experiment, using Meta-HTML and Java. It offered more than 50 compact camera models and more than 100 winter jackets for sale. All participants had high-speed access to it from a computer laboratory at Humboldt University. Participants were told that the store would be hosted by an industrial partner who did not wish to be named and that all data would be directly transferred to this remote host.

The online shopping environment employed a 2nd generation e-commerce type of communication, in which an anthropomorphic 3-D shopping bot involved users in a sales dialogue and gave product recommendations. Unlike current shopping agents on the web, the bot not only focused on product attributes, but also asked 'soft questions' that can typically be found in offline sales conversations. 56 bot questions had been developed for this purpose in cooperation with human sales agents from a Berlin retail store. The goal of bot communication design was not to minimize a user's time cost, but, on the contrary, to include more questions and in particular more personal questions than one would expect customers to answer. In addition to product attribute questions like *"How strong do you want the zoom of the camera to be?"* we therefore integrated three further question categories: 1) Questions concerning the intended use of the product (e.g., *"At what occasions do you usually take photos?"*). 2) Questions that addressed the buyer personally, but would also influence product recommendation (e.g., *"How important are trend models to you?"*). And 3) personal questions independent of the product but still related to the sales context (e.g., *"What do you do with your photographs?")*. This latter category also included 'non-legitimate' extremes such as questions on how *"photogenic"* or *"conceited"* people considered themselves to be. Table 1 shows some selected bot questions for the 2 products.

[4] Since project finances did not allow us to offer the 60% discount to all buyers, the incentive structure was such that a lottery after the shopping session decided on one out of 10 participants who would have the right to buy for 60% off. The remaining participants received a small financial compensation. If someone had not bought, but won the lottery, he or she would go out empty.

4 categories of bot questions	Camera	Jacket
product attribute questions	How strong do you want the zoom of the camera to be?	What size do you need for the jacket?
usage oriented questions	At what occasions do you usually take photos?	At what occasions do you want to wear the jacket?
personal questions supporting product selection	How important are relatively cheap photo development cost to you?	How important are trend models to you ?
personal questions independent of product selection	What is your motivation when taking photographs?	How often do you buy a new jacket?

Table 1: Selected bot questions

All 56 questions had been tested in a pre-study with 39 participants who had rated their relative importance, legitimacy and difficulty [2]. This yielded, for each bot question, a mean legitimacy value and a mean importance value. On average, 32% of camera questions and 39% of jacket questions were judged as relatively non-legitimate, and 37% camera questions as well as 50% jacket questions as relatively unimportant.[5] However, Mann-Whitney test for the two product question catalogues showed non-significant differences between the distributions of mean question importance ($p=0.543$) and mean question legitimacy ($p=0.386$) in the two shops. This allowed us to pool data from the two product shopping sessions for privacy analysis.

2.3. Procedure

On arrival at the laboratory, participants were told that the experiment's goal was to test interaction with a new product search engine developed at the Institute of Information Systems at Humboldt University. They were then asked to fill out a paper-and-pencil questionnaire. Privacy attitudes and concerns represented 27% of the pre-shopping questions.[6]

Participants were then presented with the online store's privacy statement in printed form. In the 'soft' privacy statement (type 1), participants were told that an industrial sponsor, a reputable European company which did not wish to be named, would receive all the data they left behind during their shopping trip. Also, their rights according to the EU Directive 95/46/EC were stated in this privacy

[5] Legitimacy and importance were rated on a 0-10 point scale with 0 = totally non-legitimate/totally unimportant and 10 = very legitimate/very important; ratings referred to as 'relatively non-legitimate' or 'relatively unimportant' here were all judgements < 5

[6] It is unlikely that subjects were primed on the issue of privacy in the pre-shopping questions. This was confirmed by debriefing conversations with the participants.

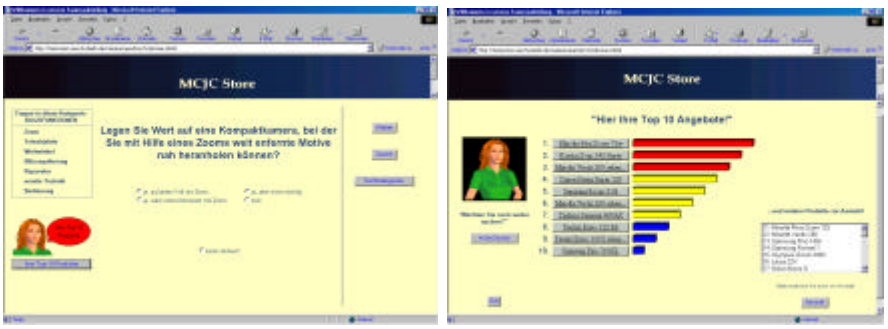
statement, including the right to know who makes use of the data, to view them and if necessary change or withdraw them. In the 'harsh' privacy statement (type 2), participants were told that their data would be handed on to an anonymous entity, and that we did not know what further use would be made of their data. 88 participants received the type 1 privacy statement, and 118 participants received the type 2 privacy statement. They had to sign that they had read and accepted this statement prior to shopping. All participants were told that it was *not* the purpose of the experiment *"to collect dummy data"*, and that we expected them to give truthful answers because the search engine we had developed would not work adequately otherwise. We added that we would *"prefer the refusal to answer to a lie"*.[7]

The navigation opportunities participants encountered in the store were similar to those in Web sites like ActiveBuyersGuide.com and PersonalLogic.com. The online store's starting page had been loaded into the Web browser by the experimenter. It displayed either a camera or a jacket storefront. After this starting page, users had the possibility to view all products one by one from a list, but could quickly find out that this way of searching was not very efficient. They were thus motivated to enter the search engine. Here, shopping bot Luci introduced herself and her purpose to the user. All users had to pass this page and were given the possibility to leave their home address. No reason was given on the page why they should enter it, but two 'proceed-bottoms' were displayed: one labelled "save address, proceed" and the second right below entitled "no address specifications, proceed". The user was thus left to decide whether to reveal the address or not without any sanctions.

Once users passed Luci's introduction, navigation occurred at two levels: a *communication* level and an *information* level. On the communication level, participants could engage in a question-answer dialogue with Luci (based on multiple-choice). Luci would ask (but not oblige) users to answer the 56 questions discussed above. On the basis of any number of answers given, she could be asked to calculate a user's product ranking with graphical emphasis given to a 'Top-10' consideration set. On the information level, participants had the possibility to view product facts, marketing text and photographs that could be enlarged. Both navigational levels where dynamically accessible from all pages. Thus, recommendations could be obtained and products inspected at any time. The shopping process could be exited at any time and a purchasing decision could be made after the request for a product information page. Figure 1 provides some screenshots of the store environment.

[7] We gave participants the option to refuse to answer any bot question by including a "no answer" button in each multiple-choice menu of answers.

Communication level: Bot questions, answers and Top-10 feedback



Information level: product photograph and description

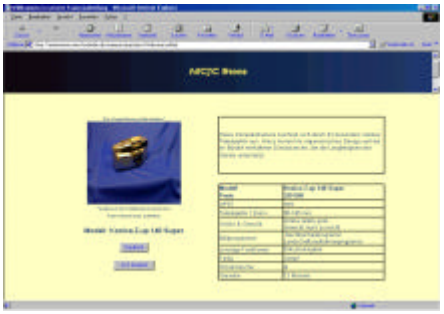


Figure 1: Screenshots of 2 navigational levels

3. Results

3.1. Data

As 6 of the 206 individual observations had missing data, analysis was based on 200 observations. Another group of 29 subjects was identified who did not see and consequently did not consciously answer or reject several bot questions. As we could not explain this behaviour and do not attribute it to any privacy concerns, we excluded these subjects from our analysis. Thus, further discussion in this paper is based on 171 observations.

Two data sources were used for analysis: questionnaire answers to discern privacy preferences and log files to analyze behavior.

3.2. Measures of Interaction Behavior

Self-disclosure is usually measured along two dimensions: its depth and breadth [18, p.328]. *Breadth* refers to the quantity of information exchanged and is

measured here by the number or proportion of bot questions answered. *Depth* usually refers to the quality of information disclosed. We operationalized information quality with the help of an index called "personal consumer information cost" (PCIC). The index was developed on the basis of the pre-study mentioned above and is described in more detail there [2].

Participants of the pre-study had been asked to rate a presented question's legitimacy and importance in the sales context, and the difficulty of answering it, as well as the "overall perceived information cost" of this question. This construct had been explained prior to the rating session as follows: *"Information cost denotes the 'intuitive readiness' to truthfully answer the question of the search engine, i.e. the spontaneous feeling whether you would be willing to reveal the demanded information about yourself. 'No' information cost means that you would have no problem at all to answer the question truthfully. 'Very high' information cost means that you would, under no circumstances, give this type of information about yourself to a search engine."* A regression analysis of the judgements of all 56 questions showed that PCIC decreased linearly with legitimacy and importance, and increased linearly with difficulty [2].

For the purposes of the current study, we computed $PCIC_j$, considering all the questions that participant j had answed (see Figure 2 for details).

$$PCIC_j = \sum_{i=1}^{k_n} \left(a - \boldsymbol{a} * Leg_i^n + \boldsymbol{b} * I_i^n + \boldsymbol{d} * Diff_i^n\right) + \sum_{i=1}^{k_s} \left(b - \boldsymbol{f} * Leg_i^s + \boldsymbol{g} * I_i^s + \boldsymbol{r} * Diff_i^s\right)$$

where

n = questions of *type* n are focusing either on the person or on envisaged product usage

s = questions of *type* s are questions concerned directly or indirectly with product attributes

i = a question answered by an online user j

k = total number of questions answered by user j

j = user

Leg_i^n = Mean perceived *legitimacy* of a *question i of type t*, $t \in \{n, s\}$

I_i^s = Mean perceived *importance* of a *question i* of type t, $t \in \{n, s\}$

$Diff_i^s$ = Mean perceived *difficulty* of a *question i* of type t, $t \in \{n, s\}$

Figure 2: Computing a users' PCIC

The values of *Leg, I, Diff* as well as the 8 regression parameters (a, b & a, b, d, f, g, r) were taken from the pre-study.

For this study it is important to know that PCIC aims to reflect an individual's perceived 'cost of disclosure' in a communication context. More precisely, we define it as the loss in utility a consumer perceives when disclosing a number of *truthful* information units about himself, assuming that his identity will afterwards be known to the organization hosting a site and that his data are collected for further usage. For example, when people decide to lie on the Internet, the cost of providing truthful information is obviously too high.

A user with a high PCIC answers many bot questions even though he perceives them as being rather non-legitimate, unimportant and difficult to answer. A user with low PCIC values answers few questions, most of which he perceives as legitimate and important and easy to answer.

3.3.Privacy attitudes and self-disclosure

As discussed above, privacy statements published on Web sites are an important baseline for today's advances in consumer protection. For the deployment of P3P, for example, it is assumed that users regard privacy policies as relatively trustworthy, consider their own privacy preferences and then act consciously in accordance with them.

To investigate privacy preferences, we built on earlier work presented by Ackermann et al. [1] at the ACM conference on EC in 1998. Employing standard multivariate clustering techniques (k-means), however, we found four instead of three groups with different privacy attitudes. We could identify not only privacy fundamentalists (cluster 4: 30%) and, in contrast to these, marginally concerned users (cluster 1: 24%), but also found two distinct groups that would focus their privacy concerns either on the revelation of identity aspects such as name, address or e-mail (cluster 3: 20%) or on the profiling of interests, hobbies, health and other personal information (cluster 2: 25%). We were thus able to separate the *"pragmatic majority"* identified by Ackerman et al. [1] into two more meaningful groups which we called "identity concerned" and "profile concerned". Figure 3 gives an overview over the four clusters identified.

Compared to the earlier study, a general rise in privacy concern can be recognized: While the proportion of *"privacy fundamentalists"* was larger in our study, the group of *"marginally concerned"* was smaller.

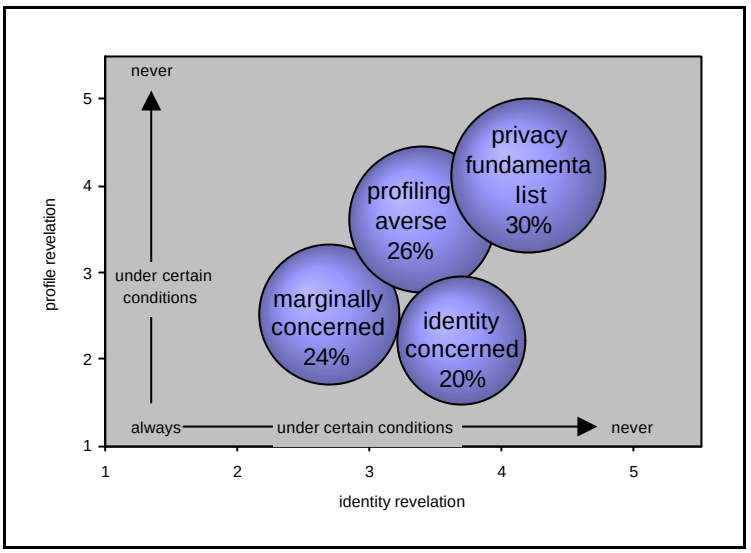


Figure 3: 4 clusters of privacy attitudes identified

We then investigated whether interaction behavior was consistent with the attitudes stated. Two aspects of interaction behavior were considered: (a) whether participants voluntarily communicated their address to Luci before entering the question-answer cycle, and (b) how many and what types of questions participants answered when communicating with Luci. The first variable is a measure of the willingness to satisfy an information request *separated* from the sales dialogue and linked to identification. We expected that 'identity concerned' users (cluster 3) would react particularly averse to this type of information provision. The second variable is a measure of the willingness to provide information *embedded* in a sales dialogue. As many personal and profile-sensitive questions are asked in this communication context, one would expect that here 'profiling averse' users (cluster 2) would be particularly reserved.

3.3.1. Address Provision

As expected from the nature of the cluster, marginally concerned users (cluster 1) had the lowest refusal rate in providing their home address for both privacy statements (30% for PS type 1 and 41% for PS type 2). Surprisingly, 24-28% of privacy fundamentalists voluntarily provided their address before interacting with the search engine. Identity concerned participants (Cluster 3) showed unexpected behavior. While under the condition of the first privacy statement 93% refused to provide their home address, only 65% did so under the even "harsher" conditions of PS type 2. Due to the particularly small size of this group, one should probably

not generalize. More research is needed to investigate this finding. All observations are summarized in table 2.

Notably, across privacy statements there was an average of 35-40% of participants who gave their home address without any reason to do so. This raises the question how privacy conscious online users really are. In particular, the mentioning of the 'security providing' EU law, led to an increase in voluntary address provision, as can be seen for most clusters in table 2. The average difference of 5% more address provision with EU law citation (11% without the inconsistent group of cluster 3) was interesting, though not significant ($\chi^2_{(1)}=0.33$, $p > 0.5$ one-sided).

Cluster	PS type 1 (voluntary address provision)	PS type 1 (*no* voluntary address provision)	PS type 2 (voluntary address provision)	PS type 2 (*no* voluntary address provision)	sum of participants
CL1: marginally concerned	14	6	13	9	42
% of cluster	70%	30%	59%	41%	
CL2: profiling averse	9	10	7	19	45
% of cluster	47%	53%	27%	73%	
CL3: identity concerned	1	13	7	13	34
% of cluster	7%	93%	35%	65%	
CL4: fundamentalists	7	18	6	19	50
% of cluster	28%	72%	24%	76%	
sum tot	31	47	33	60	171
% of sum	40%	60%	35%	65%	

Table 2: Contrasting privacy attitudes with voluntary address provision

3.3.2. Revelations during the sales dialogue

To represent the depth of interaction with the sales bot, we used the PCIC index described above. The 171 PCIC index values where split into terciles, contrasting individuals with low, medium and high disclosure. Table 3 summarizes the findings. Table 3 shows that participants from all clusters had a strong tendency to self-disclose. 87% of users were in the group with maximum PCIC values. This

behaviour could be observed across both product types, with 84% of camera shoppers and 98% of jacket shoppers in the highest PCIC group.

Averaging across clusters, an average of 85.84% of bot questions were answered (85.77% for cameras and 86.05% for jackets). As expected, however, the distribution of PCIC was different across clusters ($\chi^2_{(6)}$=16.57, p<0.05 two-tailed).

Cluster	low PCIC	medium PCIC	high PCIC	sum
CL1: marginally concerned	0	0	42	42
row %	0%	0%	100%	100%
total %	0%	0%	24%	24%
CL2: profiling averse	3	7	35	45
row %	7%	15%	78%	100%
total %	2%	4%	20%	26%
CL3: identity concerned	0	1	33	34
row %	0%	3%	97%	100%
total %	0%	1%	19%	20%
CL4: fundamentalists	3	8	39	50
row %	6%	16%	78%	100%
total %	2%	5%	23%	30%
Sum	6	16	149	171
total %	4%	9%	87%	100%

Table 3: Contrasting privacy attitudes with online communication behaviour

An investigation of cluster details showed that especially privacy fundamentalists (cluster 4) do not live up to their expressed attitude. 78% of them display high PCIC values and answered an average of 86% of the bot questions. With this, they only answered 10 percentage points fewer questions than marginally concerned participants (cluster 1). Comparing behaviour for the two product groups, we found that for cameras only 83% of privacy fundamentalists had a high PCIC value, while for jackets 95% of fundamentalists were in this group. A difference of 7% in self-disclosure between the two products can also be observed for cluster 2. The findings hint at the possibility that the product category may have an influence on the extent of information revelation.
Consistent with the expectations, profiling averse participants (cluster 2) gave less information during the shopping dialogue than identity concerned participants

(cluster 3). With 'only' 78% of people being in the high PCIC group, cluster 2 and 4 turned out to be the groups with the most reserved behavior.

Mann-Whitney tests for different PCIC distributions across the two privacy statements generally ($p=0.969$) and for both products separately (camera: $p=0.526$; jackets: $p= 0.227$) showed no significant differences in this obvious readiness of users to self-disclose. This is a surprising result as we would have expected the privacy statement to have a greater impact on disclosure.

The readiness of participants to reveal most of or even all of the information demanded from them during the sales dialogue with the shopping bot, and the widespread willingness to also provide their address, are alarming findings. The degree of inconsistent behavior found in the data among 'privacy aware' clusters 2-4 appears particularly problematic. The results are even more relevant when one considers the experimental conditions: after all, bot questions were designed to include many non-legitimate and unimportant personal questions. Participants also had to sign that they agreed to the selling of their data to an anonymous entity. During the experimental briefing, in which instructions were read aloud to all participants, one of our goals was to minimize sympathy with us as experimenters. The conditions under which participants 'revealed themselves' were therefore probably even more unfavorable in terms of privacy than a regular Internet interaction would be. This indicates that even though Internet users have some view on privacy, they do not act accordingly. Mostly, they are willing to reveal themselves.

4. Discussion of results with a view to privacy technologies and privacy regulation

A majority of persons who participated in the shopping experiment disclosed so much information about themselves that a relatively revealing profile could be constructed on the basis of only one shopping session. This result is not only alarming in itself, but even more so given that for many participants this behavior stands in sharp contrast to their self-reported privacy attitude (especially for the profiling averse and fundamentalist participants in clusters 2 and 4). It raises the question of how privacy can be protected effectively while at the same time avoiding tutelage.

4.1. Privacy statements and P3P

As was outlined above, privacy statements play an important role in addressing privacy through P3P. It was shown, however, that while people do tend to provide less *identification* information they do not alter their *communication* behaviour significantly in response to them, neither in disclosing their profile nor in navigation. Even privacy conscious users (clusters 2 and 4) seem to be 'drawn to

reveal themselves' to the sales bot, and only 3 out of 171 avoided the dialogue offered (which is similar to blocking or avoiding promising communication in a P3P scenario).

Still, P3P has the potential to considerably enhance privacy standards: first, it may enhance user trust in privacy statements, because companies, by taking the burden of encoding their website practices, signal their willingness to respect their users' privacy. Privacy can thus become a *recognized* means of differentiation. Second, P3P is able to correspond to the different privacy preferences of users. For example, with P3P identity concerned users have the possibility to effectively exclude sites that demand information in the categories <physical>, <online> or <uniqueid>.[8] Third, P3P is a relatively open platform standard. It could easily be extended to prohibit or at least warn of communication processes such as the one we used in the current experiment. In order to address data categories dominant in interactive EC Web sites, P3P has so far only provided for the overall data categories <interactive> and <preferences> to signal the deployment of interactive features on a site. These 2 categories, however, provide marketers with the opportunity of implementing the very type of privacy-invading communication we offered in the experiment. Since online users do have a strong incentive to generally accept interactive and preference-demanding websites (because this is basically what makes e-commerce sites interesting), there is a considerable risk for online users to sacrifice their privacy as they did in the experiment presented above. Moreover, P3P (similar to legislation) would signal the trustworthiness of the site without really living up to this standard.

On this background, an extension of the P3P protocol would appear desirable that takes this important type of application into consideration. For example, and thanks to the openness of XML (the basis of P3P) it would be relatively easy to break down the data category <interactive> into several sub-categories that signal the 'true nature' of interaction implemented in a site. One possible way to characterize a question-answer process with a sales bot would be to distinguish between the types of questions asked by the agent. For example, a differentiation could be made between product attribute questions, usage oriented questions, personal questions supporting the selection process, and finally personal questions that have no impact on product selection. Extending the data category <interactive> by this kind of sub-categories (which we also used to design the shopping bot) would give users a 'meaningful' choice to administer privacy when they interact with sales bots, because they gain an idea of what is hidden behind the term 'interactive'. For marketers who wish to inform users of their data collection practices, these proposed sub-categories are also a cost efficient way to signal the nature of communication, as the alternative would be to encode all information demanded into separate data entities.

[8] For more detail on the meaning of data categories in P3P please consult the description of the latest public version of P3P (paragraph 3.4) on http://www.w3.org/TR/P3P/

4.2. EU regulation

The effect of citing EU regulation in privacy statement 1 revealed a potential drawback of this approach to protect privacy: it seemed to make people feel 'secure', leading between 5 and 11% more users to reveal their home address than in the less protected environment of privacy statement 2. Also, EU regulation would probably have impeded a 'real-world' implementation of the type of communication we proposed in the experiment, because it would probably not comply with the principle of 'purpose limitation' or data collection parsimony. However, most subjects indicated in a debriefing questionnaire that they appreciated this very type of soft communication. Even those individuals who had expressed privacy concerns in the first questionnaire and were not too fond of the recommendation quality wrote that they felt supported by agent Luci in *"getting a feel"* for the product, that the questions were not *"too technical"* and *"easily comprehensible"* and that they *"felt personally addressed"* in their concerns. These judgements are interesting when one considers the question design and content described above. They suggest that online users would appreciate the kind of 'personal' online communication that is either prohibited by law today or avoided by marketers due to their fear of intruding on users' privacy too much. An important question for the current privacy debate and research initiatives in this field is therefore how e-privacy could be guaranteed to people while still allowing them to benefit from 'rich' and 'soft' online communication.

4.3. Pseudonymity, identity management systems and private credentials as a way out of the privacy dilemma?

Assuming that people want rich communication (as current results suggest) and are willing to 'chat' about themselves, EC environments should provide for this desire and offer more soft communication and interaction than is currently the case [23]. Fears of 'intruding' on users' privacy by asking them more personal questions online seem unfounded on the background of survey results presented above. However, even if consumers did not display a particularly privacy-conscious behaviour in our study, there are still many reasons for companies to care about the subject. Not only do they confront EU regulation (even if they are in the US), but in order to leverage the true benefits of 'e-loyalty' they should not build on the long-term persistence of their customers' current ignorance, but ensure that consumers feel free to communicate frankly and truthfully even as their privacy concerns are on the rise.

The way to realize both marketer benefits through data-intensive personalization on the one hand and e-privacy on the other may lie in the concept of pseudonymity [20]. As long as pseudonyms cannot be linked to the identity of a user (at least not for regular EC transactions) he or she can remain relatively anonymous in the online world and feel more at ease to interact. Personalization could then be applied to these pseudonyms while still reaching the customer in

person. Of course, pseudonymity is not a new phenomenon to the online world as companies such as e-bay or Yahoo! already employ it in their Web sites. However, current use of pseudonyms is still in its infancy. Not only is pseudonymity sacrificed at the moment of buying when users reveal their true identity, but also the management of pseudonymity is cumbersome. Users have to manage the complexity of an ever-rising number of virtual identities, and marketers employing pseudonyms have to maintain a database of a rising number of 'lifeless' (unused) user equivalents. The approach of currently proposed, simple-to-use identity management systems may therefore represent an important privacy technology of the future [15,16]. They would be able to assist online users in controlling their virtual identities and also ensure that customers revisit sites under the same virtual identity if they wish to (situational pseudonyms). More importantly, however, they are envisaged to include anonymous authentication methods and private credentials [8] so that transactions are supported while users are not left alone with complex technology they do not understand [11].

5. Conclusion

We conducted an experiment in which we compared self-reported privacy preferences of 171 participants with their actual self-disclosing behaviour during an online shopping episode. Our initial hypothesis that users' privacy concerns impede the depth and breadth of truthful online interaction was not confirmed. In contrast, participants displayed a surprising readiness to reveal private and even highly personal information and to let themselves be 'drawn into' communication with the anthropomorphic 3-D bot.

The results obtained are important for the current privacy debate. Not only does the study in itself represent the biggest empirical observation of actual privacy behavior, but in its set-up it was also adapted to the 2nd generation E-commerce type of sales environment lying ahead. More importantly, it revealed a major misconception of the current privacy debate: that people behave in the way they say they will. This result suggests that the development of privacy technologies needs to take a twist into a new direction: they need to be designed in such a way that they allow even moderately computer-literate online users to protect themselves from the degree of self-disclosure they are afraid of.

Acknowledgements

We wish to thank Karstadt Quelle New Media (KQNM) for the financial sponsoring of the IWA experiments, Artificial Life for lending us their 3-D bot figure, Martin Strobel <Infonomics, The Netherlands> for assisting the experimental set-up and programming of the store, Oliver Günther <Institute of Information Systems> and Marit Köhntopp <Independent Centre for Privacy

Protection Schleswig-Holstein, Germany> for helpful suggestions on earlier drafts. We are also grateful to Humboldt University Berlin as well as the NaFög scholarship programme for financially supporting the authors.

Bibliography

[1] Ackerman, M.S., L.F. Cranor and J. Reagle, "Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences", in: *Proceedings of the ACM Conference on Electronic Commerce EC'9*, 1999

[2] Annacker, D., Spiekermann, S., Strobel, M., "Private Consumer Information: A new search cost dimension in online environments", 14th Bled Conference of Electronic Commerce, June 2001, download: http://www.wiwi.hu-berlin.de/~sspiek/phdresearch.html

[3] Ansari A., Essegaier, S., Kohli, R., "Internet Recommender Systems", in: Journal of Marketing Research", Vol. 37, August 2000, pp. 363-375

[4] Berendt, B. (2000). Web usage mining, site semantics, and the support of navigation. In: Working Notes of the Workshop "Web Mining for E-Commerce Challenges and Opportunities." 6th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, August 20-23, 2000. Boston, MA. pp. 83-93

[5] Bäumler, H., "Datenschutz im Internet", in: E-Privacy, ed. by Helmut Bäumler, Wiesbaden, 2000, pp. 1-8

[6] Berry, M., Linoff, G., "Data Mining Techniques for Marketing, Sales and Customer Support", Wiley, NY, 1997

[7] Borking, J., "Erwartungen an die Datenschutzbeauftragten im Internet", in: E-Privacy, ed. by Helmut Bäumler, Wiesbaden, 2000, pp. 280-290

[8] Brands, S., Rethinking Public Key Infrastructure and Digital Certificates – Building in Privacy, thesis, 1999, 2nd edition: The MIT Press, August 2000

[9] Cassell, J., "Embodied Conversational Interface Agents", in: Communications of the ACM, Vol. 43, No. 4, April 2000, pp. 70-78

[10] Chang, A., Kannan, P., Whinston, A., "The Economics of Freebies in Exchange for Consumer Information on the Internet: An Exploratory Study", in: International Journal of Electronic Commerce, Vol. 4, No. 1, Fall 1999, pp. 85-101

[11] Clauß, S., Köhntopp, M., "Identity Management and Its Support of Multilateral Systems", accepted for publication in the Special Issue on 'Electronic Business Systems' of Computer Networks; until publication available directly from marit@koehntopp.de

[12] Directive 95/46/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: http://europa.eu.int/comm/internal_market/en/media/dataprot/law/index.htm

[13] Hagel, J., Rayport, J., " The Coming Battle for Customer Information", in: Harvard Business Review, January-February 1997, pp.53-65

[14] Häuble, G., Trifts,V., "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids", in: Marketing Science, April 2000
[15] Köhntopp, M., "Wie war noch gleich Ihr Name? – Schritte zu einem umfassenden Identitätsmanagement", accepted at the conference *VIS – Verlässliche IT-Systeme*, Kiel, Germany, September 2001
[16] Köhntopp, M., Pfitzmann, A., "Datenschutz Next Generation", in: E-Privacy, ed. by Helmut Bäumler, Wiesbaden, 2000, pp. 316-322
[17] Moon, Y.: The Interface Project: http://www.people.hbs.edu/ymoon/Interface/home.html
[18] Moon, Y., Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers, in: Journal of Consumer Research, Vol.27, No.4, March 2000
[19] Pew Internet & American Life Project, Trust and Privacy Online: Why Americans Want to Rewrite the Rules, 2000-8-20, http://pewinternet.org/reports/toc.asp?Report=19
[20] Pfitzmann, A., Köhntopp, M., "Anonymity, Unobservability, and Pseudonymity – A Proposal for Terminology", in: Designing Privacy Enhancing Technologies, *Proceedings of WS on Design Issues in Anonymity and Unobservability*, LNCS 2009, Heidelberg, 2001, revised version at http://www.koehntopp.de/marit/pub/anon/Anon_Terminology.pdf
[21] Schaar, P., "Die Möglichkeiten der Datenschutzaufsichtsbehörden", in: E-Privacy, ed. by Helmut Bäumler, Wiesbaden, 2000, pp. 69-76
[22] Schafer J., Konstan, J., Riedl, J., "RS in E-Commerce", in: *Proceedings of the ACM Conference on Electronic Commerce EC'9,* 1999, pp. 158-166.
[23] Spiekermann, S., Corina, P., "Motivating Human-Agent Interaction : Transferring Insights from Behavioral Marketing to Agent Design", in: *Proceedings of the 3rd International Conference on Telecommunications and Electronic Commerce, ICTEC3*, 2000, pp. 387-402
[24] Spiliopoulou, M., "Web Usage Mining for Web Site Evaluation – Making a site better fit its users", in: Communications of the ACM, No. 8, Vol. 43, August 2000, pp. 127-134
[25] Urban, G., F. Sultan and W. Qualls, "Design and Evaluation of a Trust Based Advisor on the Internet", MIT, December 1999
[26] Wells, N. and Wolfers, J., "Finance with a Personalized Touch", in: Communications of the ACM, No. 8, Vol. 43, August 2000, pp. 31-34
[27] West P., D.Ariely, S.Bellman, E.Bradlow, J.Huber, E.Johnson, B.Kahn, J.Little and D.Schkade, "Agents to the Rescue?", HEC Invitational Choice Symposium, February 1999
[28] Westin, A., "Harris-Equifax Consumer Privacy Survey", Atlanta, GA: Equifax Inc. (1996)
[29] Vulcan, N., "Economic Implications of Agent Technology and E-Commerce", in: The Economic Journal, February 1999, pp. 67-90