Privacy in E-Commerce: Stated Preferences vs. Actual Behavior[*]

Bettina Berendt, Oliver Günther, and Sarah Spiekermann
Humboldt-Universität zu Berlin
Institute of Information Systems
Spandauer Str. 1
D-10178 Berlin, Germany
{berendt,guenther,sspiek}@wiwi.hu-berlin.de

Introduction

In times of ubiquitous electronic communication and increasing industry pressure for standard electronic authentication, the maintenance of privacy (the "right to be let alone") becomes a subject of increasing concern. The possibility of a "transparent human" appears most obvious in electronic commerce, due partly to the large amounts of data available, partly to the high payoffs expected from using this data for marketing purposes.

Questionnaire-based surveys suggest that many people strongly oppose this trend. For example, 75% of the German Internet users surveyed in [5] professed to be at least sometimes afraid that their privacy may be compromised when surfing the Internet. 60% had avoided a website it in order to protect their privacy, and 47% sometimes provided false data. Similar results have been obtained in other countries [8]: 30-40% of online users admit to regularly lying online when asked about their personal habits and preferences. The use of obvious ("Donald Duck") and less obvious aliases is commonplace. The integrity and efficiency of commercial websites' data protection measures is widely doubted.

Given these widespread concerns about personal privacy in a networked world, it is commonly assumed that people systematically draw their consequences. Privacy enhancing technologies (PETs) such as P3P build on the idea that Internet users reflect on privacy statements, are restrictive in what they reveal, to whom, and under what circumstances—in short, that they *act in accordance with their privacy preferences*. However, as our study will show, this is by far not always the case.

In this paper, we describe results from a large-scale online shopping experiment. They suggest that, given the right circumstances, online users easily forget about their privacy concerns and communicate even the most personal details without any compelling reason to do so. This holds in particular when the online exchange is entertaining and appropriate benefits are offered in return for information revelation—circumstances easily created by second-generation agent technologies and embodied interface agents. Privacy statements have no impact on most users' behavior. In concluding, we discuss some possible reasons for this discrepancy between stated preferences and actual behavior. We also suggest ways how to help users better align their actions with their goals.

[*] To appear in *Communications of the ACM*; final manuscript; 6 October, 2003

An Experimental Investigation of Privacy Attitudes and Behavior

The setting: An online store with agent recommendations

The initial goal of our study (for details, see [12]) was to investigate drivers and impediments of online interaction in general. Privacy concerns were suspected to be *one* major impediment of truthful and deep online interaction. In particular, our study focused on how self-reported privacy concerns relate to actual self-disclosing behavior, and on the impact of privacy statements.

In a laboratory experiment, 206 participants took a virtual shopping trip for cameras and jackets. As an incentive to participate, these high value goods were offered at a 60% discount compared to local store prices. The buying decision was assisted by an anthropomorphic shopping bot. Participants had to spend their own money if they decided to buy.

Before shopping, participants filled out a questionnaire. 27% of the questions were privacy-related. They addressed the willingness to reveal certain types of private data, the general trust in privacy statements, the value of privacy, and intended reactions to various privacy scenarios.

Participants were asked to sign the store's privacy statement, agreeing to the sale of their data to an anonymous project sponsor. One group received a "cordial" privacy statement which told them that their navigational data would be handed over to the sponsor, a reputable European company, and advised them of their rights under the European Union Directive on Data Protection (95/46/EC): the rights (a) to be informed about who processes the data for which purpose, (b) to inspect one's data and, (c) if incorrect, to enforce amendment, and (d) to refuse to consent to specific types of usage. The other group received a "terse" privacy statement, which did not mention the EU Directive but told them that it was unknown which use the sponsor would make of their data.

The navigation opportunities in the store were similar to those in current online shops. At the beginning, the anthropomorphic shopping bot named Luci introduced herself and her purpose to the user. Before entering the store, users were given the possibility to leave their home address, but no reason or requirement to do so.

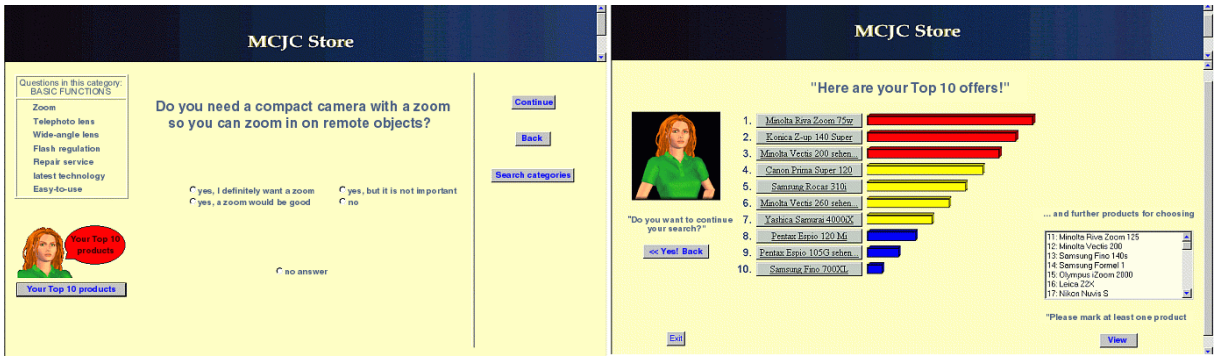

Figure 1: Agent questions and recommendations

Users were then invited to answer any of up to 56 agent questions (see Figure 1). At any time, the agent could be asked to determine a personalized top 10 list of products, based on the answers given so far. The user could request information on each product and choose to purchase it. Unlike current Web shopping agents, the bot not only focused on product attributes, but also asked "soft questions" that can typically be found in offline sales conversations. The goal was to include more questions, and 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 asked questions concerning the intended use of the product (e.g., ”At which occasions do you usually take photos?”), questions that addressed the buyer personally but also influenced product recommendation (”How important are trend models to you?”), and personal questions unrelated to the product but related to the sales context. The latter also included questions that would usually be considered inappropriate, such as how “photogenic” or “conceited” people considered themselves to be. An independent evaluation of bot questions [3] had shown that about half of them would be considered non-legitimate or unimportant with respect to the given sales context.

Analysis

The final analysis, based on 171 valid cases, compared users’ attitudes with their actual behavior. Based on the Web log data, behavior was described in terms of the information provided. The quantity of that information was measured by the proportion of bot questions the user had answered. To measure information quality, we developed an index called *personal consumer information cost (PCIC)* that considered each answered question’s legitimacy and importance in the sales context, as well as the difficulty of answering it. A PCIC of zero means that the user would have no problem at all to answer the question truthfully. A high PCIC implies that users would be greatly reluctant to give this type of information. Regression analysis confirmed that PCIC is strongly negatively correlated with legitimacy and importance, and moderately positively correlated with difficulty [3]. A participant’s PCIC index was computed as the sum of PCIC indexes of all questions s/he had answered, grouping values into “high,” “medium,” and “low.” A large number of answers in response to mostly irrelevant or non-legitimate questions thus leads to a high PCIC.

Results: Privacy attitudes and self-disclosure

A clustering of the answers to privacy-related questions revealed four different groups of users (Figure 2). We could clearly distinguish a group of *privacy fundamentalists* and another group of only *marginally concerned* users as found elsewhere [1]. We were able to differentiate the remaining participants by the focus of their privacy concerns: *Identity concerned* users are more concerned about revealing information like their name, address, or e-mail, while *profiling averse* users are more concerned about disclosing their interests, hobbies, health, etc.

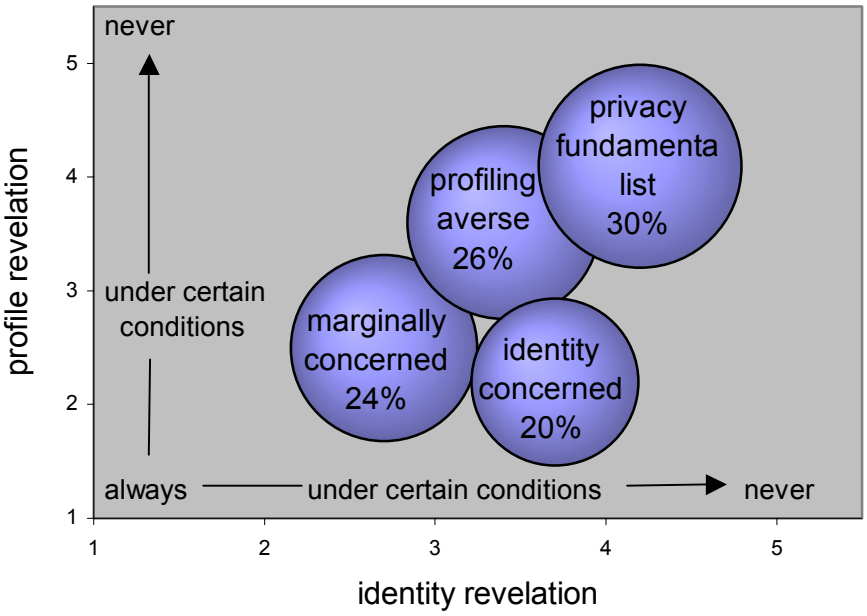

Figure 2. Four clusters of privacy attitudes

To investigate whether users' interaction behavior was consistent with their privacy attitudes, we examined (i) whether participants voluntarily gave their address to Luci before entering the question-answer cycle, and (ii) how many and what types of her questions they answered.

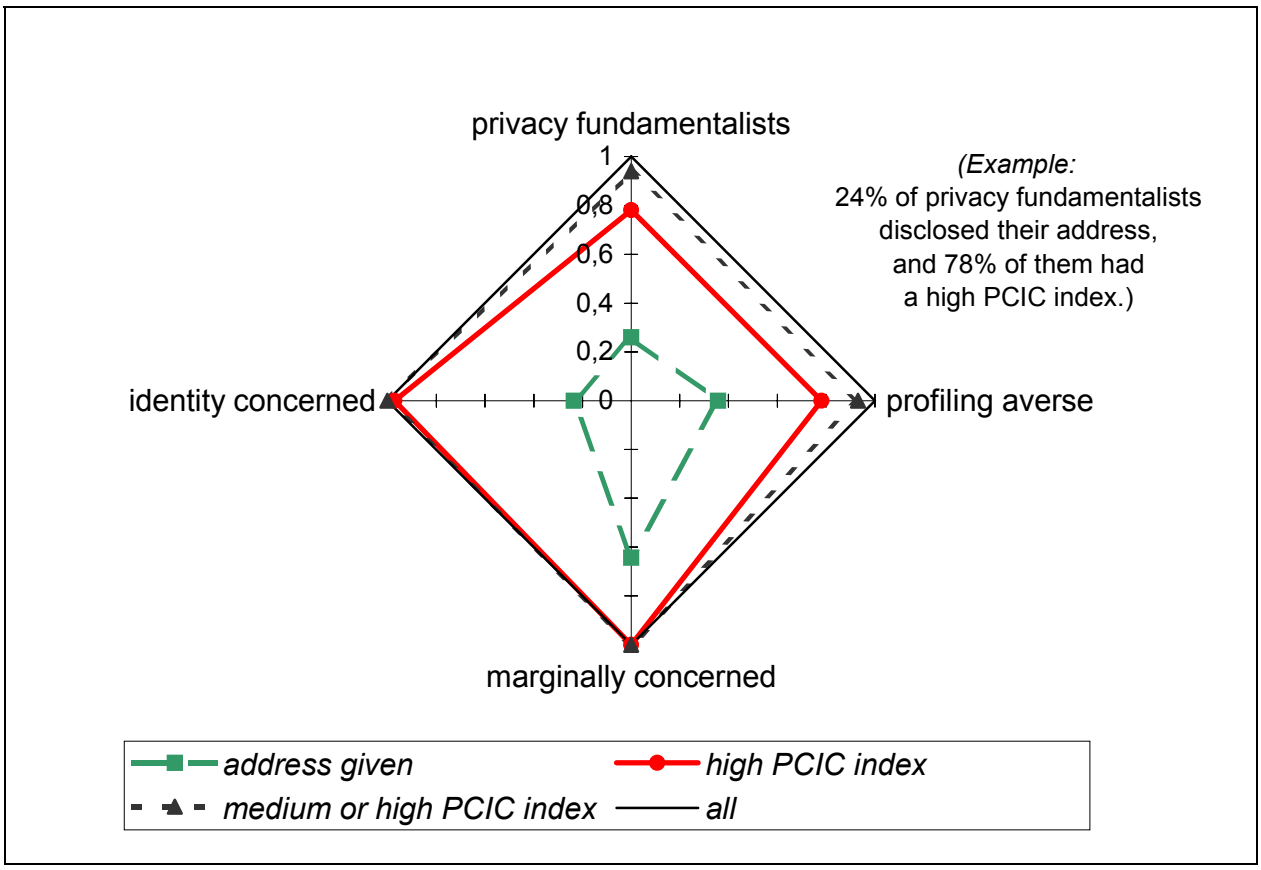

Figure 3: Attitude clusters and disclosing behavior

Figure 3 shows (i) the proportion of participants who disclosed their address (dashed line), and (ii) the distribution of PCIC index values: high (solid red line), medium (distance solid red line – dotted gray line), and low (distance dotted gray line – outermost diamond). The differences between proportions across clusters were significant at the 5% level.

As expected, disclosure rates increased from privacy fundamentalists to marginally concerned users. Identity concerned and profiling averse users showed intermediate disclosure rates and acted in *relative* accordance with their stated preferences: The former withheld their address more often, and the latter had lower PCIC index values. However, contrary to our expectations, the absolute level of disclosure is alarmingly high across all clusters, belying the previously-expressed reluctance to disclose information online.

Neither the product category nor the type of privacy statement had a statistically significant impact. However, the "cordial" privacy statement (which referred to the EU Directive) induced slightly more participants to provide their address. This is a cause for concern: It suggests that the more people believe in the effectiveness of existing jurisdiction, the less they control their personal behavior.

In the debriefing questionnaire, most participants indicated that they appreciated the communication employed and that they felt "personally addressed" and "supported" by agent Luci. This was stated even by those individuals who had previously expressed privacy concerns and were not too fond of the quality of Luci's recommendation quality. In the debriefing discussions, participants showed no sign of recognizing any link of the experiment to privacy research, and did not comment on a discrepancy between privacy preferences and behavior.

Why Does Behavior Diverge From Attitudes?

Our study has shown that Web users welcome a rich interactive environment. In such an environment, they are willing to talk about themselves, thus creating the basis for efficient customer relationships. The other important news is that they do not always act in line with their stated privacy preferences, giving away information about themselves without any compelling reason to do so.

This disparity may be disadvantageous not only for customers, but also for the e-commerce companies that may welcome the additional data at first: If customers are later confronted with the discrepancy between their actions and their ideals—e.g., because the company uses its knowledge about certain of the customers' preferences—they may react with resentment, which can damage the customer relationship [2].

Inconsistencies between people's behavior and their self-reports are a well-known phenomenon, with explanations emphasizing cognitive and/or social aspects of decision making and behavior [11].

In many situational contexts, decisions are based on heuristics rather than on rational consideration of all factors for or against all possible courses of action (for an overview, see [4]). In our case, the shopping context may in particular have drawn attention to the potential *gains* from disclosing personal information: product recommendations and the chance to obtain a discount. In addition, the wealth of choices in this store interface may have led to a certain decision aversion and the accompanying wish to collect all possible information. In contrast, both the questionnaire items and the direct evaluation of bot questions in [3] may have framed information disclosure in terms of a *loss* of privacy, and the possible lack of legitimacy and importance. Moreover, the first impression of Luci may have engendered a

positive mood, which makes positive memories of productive (and harmless) interactions with shop assistants more *available* and leads to the expectation that the current interaction will be like this too.

Heuristics that specifically simplify *communication* are likely to have played a role as well. Since the first anecdotal evidence of interactions with ELIZA, Joseph Weizenbaum's 1960s computer "psychotherapist," it has repeatedly been observed that people tend to treat interactive software like a trustworthy human communication partner. Human communication is usually characterized by adherence to the "Gricean maxims of cooperativity," which implies in particular saying things that are true and relevant to the conversation. This generally holds for agent communication as well. (The Gricean maxims are in fact a popular guideline for agent communication design.) At least as pervasive as the *actual* adherence to these maxims, however, is the—often implicit—*expectation* of adherence. Even in surveys and experiments, people assume that their dialog partners ask questions that are relevant [9]. This expectation often leads people to re-frame their perception of something said that, at first sight (or viewed in isolation [3]), seems to violate the cooperativity maxims.

These results and their interpretation need to be substantiated and investigated further. Our sample was a self-selected sample that was relatively well-educated, young, and with considerable online experience: 92.7% were students; 44.2% were females; and 98.5% (91.7%) were (regular) Internet users. A more diverse population should be investigated. It was not possible to control for the different roles that the financial incentives and the specific shop interface may have played in the sample. Also, it could not be ruled out that in spite of our efforts to the contrary, the university setting may have made participants more trustful. Nonetheless, the findings appear significant enough to warrant better measures of protection.

How Can Privacy Preferences be Protected?

An important result of this study is that while many users have strong opinions on privacy and do state privacy preferences, they are not able to act accordingly. Once they are in an online interaction, they often do not monitor and control their actions sufficiently; and *seen* privacy statements have no impact on behavior. Users rely on legal protection, even though it is widely known that laws and regulations have difficulties to respond to the fast changes in Internet communications. Given this discrepancy, software appears to be a better basis for effective privacy protection.

Currently, P3P is regarded as one of the most promising tools, as it can give automatic warnings if a website's privacy policy does not correspond to one's personal privacy preferences. Yet, beyond these warnings, the tool does not protect a user once a website is entered. Privacy preferences cannot be expressed on a per-service level; they are static across the Web. Thus, P3P may not be effective enough.

We therefore advocate the development of alternatives in a new form of client-sided privacy enhancing technology (PET) which combines most of today's research efforts.

To support the rich and service-dependent interaction users desire,

1. PET should monitor third-party services as P3P does today and bring potential problems to the user's attention. Yet in addition, it should *learn* users' privacy preferences by observation [7], change settings dynamically and on a per-service level.
2. PET should record all interactions with all Web services, creating information-rich *client-side profiles* [10]. At the user's discretion, parts of that profile can be made available to marketers or peer networks.

To empower users and protect them against the described context effects,

3. PET should have easy-to-use interfaces and privacy-friendly default settings.
4. PET should provide *identity management* [6], allowing users to adopt new pseudonyms whenever they (re-)enter a site and sheltering client-side profiles.
5. PET should *decontextualize*. Recognition and blocking of dangerous interactions could be done automatically if PET were able to understand all interactions. However, since automatic language understanding is anything but perfect, PET must employ its user interfaces to support users' thinking *about* their actions *while* they are acting, e.g., with windows that disturb the flow of interaction popping up upon unclear information requests. Based on its learning capabilities, the agent should issue warnings selectively to avoid ineffectiveness. However, learning from (ineffectual) behavior is not enough. Rather, PET could, for example, cluster interactions and periodically submit them to a critical review by the user. Alternatively, "good" interaction histories could be pooled in a peer network, and used as a basis for individual PET agents' learning.

An additional desirable feature is the use of knowledge about Web services and their privacy practices. PET could, for example, periodically compare users' privacy preferences to changes in a site's privacy policy, changes in ownership of an organisation, or recorded abuses of private data. When detecting a suspected violation, the agent could change its settings and recommend actions to take. This is becoming possible with the rapidly increasing availability of *privacy-related metadata* from independent agencies and public review boards.

By combining these techniques, this PET would represent a more timely privacy protection and trust tool for modern Web applications. Even though people can potentially still reveal everything about themselves, this PET would go far beyond P3P, ensuring identity protection and serving as a learning and intelligent watchdog at the user's service.

References

1. Ackerman, M.S., Cranor, L.F., and Reagle, J. Privacy in E-commerce: Examining user scenarios and privacy preferences. In *Proceedings of the ACM Conference on Electronic Commerce EC'99* (Denver, CL, Nov.). 1999, 1-8.
2. Adams, Anne. *Users' Perceptions of Privacy in Multimedia Communications.* PhD Thesis, University College London. 2001. http://www.cs.mdx.ac.uk/RIDL/aadams/thesis.PDF. Access date: 06 Oct. 2003.
3. Annacker, D., Spiekermann, S., and Strobel, M. E-privacy: A new search cost dimension in online environments. In *Proceedings of the 14th Bled Conference on Electronic Commerce*. 2001.
4. Bettman, R., Luce, M.F., and Payne, J.W. Constructive consumer choice processes. *Journal of Consumer Research 25* (1998), 187-217.
5. IFAK GmbH & Co. *Nur begrenztes Vertrauen der Verbraucher beim Einkaufen im Internet.* Research study. Taunusstein. Jan. 18, 2002. http://www.ifak.de/about/presse_8.php. Access date: 06 Oct. 2003.
6. Jendricke, U. and Gerd tom Markotten, D. Usability meets security – The Identity Manager as your personal security assistant for the Internet. In *Proceedings of the 16th Annual Computer Security Applications Conference* (New Orleans, LA, Dec.). 2000.
7. Lieberman, H. and Maulsby, D. Instructible agents: software that just keeps getting better. *IBM Systems Journal 35*, 3-4 (1996), 539-556.
8. Pew Internet & American Life Project. *Trust and privacy online: Why Americans Want to Rewrite the Rules.* 2000. http://www.pewinternet.org/reports/toc.asp?Report=19. Access date: 06 Oct. 2003.
9. Schwarz, N. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1996.
10. Shearin, S. and Liebermann, H. Intelligent profiling by example. In *Proceedings of the ACM Conference on Intelligent User Interfaces* (Santa Fe, NM, Jan.). 2001, 145-151.
11. Simonson, I., Carmon, Z., Dhar, R., Drolet, A., and Nowlis, S.M. Consumer Research : In search of identity. *Annual Review of Psychology 52* (2001), 249-275.
12. Spiekermann, S., Grossklags, J., and Berendt, B. E-privacy in 2nd generation E-Commerce: Privacy preferences versus actual behavior. In *Proceedings of the ACM Conference on Electronic Commerce (EC'01)* (Tampa, FL, Oct.). 2001, 38-47.