User Agents in E-commerce Environments: Industry vs. Consumer Perspectives on Data Exchange

Sarah Spiekermann[1], Ian Dickinson[2], Oliver Günther[1], and Dave Reynolds[2]

[1]Institut für Wirtschaftsinformatik
Humboldt-Universität zu Berlin
Spandauer Str. 1
10178 Berlin, Germany
{sspiek, guenther}@wiwi.hu-berlin.de
[2]Hewlett-Packard Laboratories
Filton Road
Stoke Gifford
Bristol BS34 8QZ
United Kingdom
{ian.dickinson, dave.reynolds}@hp.com

Abstract. This paper focuses on the protection of user privacy in business-to-consumer (B2C) settings. In the first part of the paper we discuss today's commercially driven customer relationship management (CRM) practices and report on the results of an interview study we conducted with nine significant Internet industry players. We analyse their current practices and expectations on service and product differentiation, price discrimination, as well as data and advertisement sales. We discuss these data usage practices critically from a user as well as privacy rights perspective. In the second part of the paper we then use those insights and propose a combination of currently researched privacy technologies into one overall approach which we call "the user model". Here, we report on how a compromise could be achieved between industry's desires for one-to-one marketing and peoples' wish to maintain control over their privacy while profiting from personalization. We discuss the role of client-side profiling, identity management, and privacy metadata and propose development principles for a user-friendly interface solution.

1 Introduction

As use of the World Wide Web has grown, more and more information about individuals – their tastes, preferences, purchases and demographic details – has been codified in electronic form. Personal information has become an economic good. However, the rules of trade for such a good are still being determined. One way to describe the current situation is as follows: two opposing players, corporations (which collect personal data) and privacy rights advocates (who seek to curtail the abuse of personal data) negotiate over the terms and conditions of the personal information exchange. The third party, the consumer, generally has to abide by whatever the two parties agree upon. Corporations, in their struggle to survive in a competitive market with increasingly disloyal customers, regard customer data as a strategic asset. It

J. Eder and M. Missikoff (Eds.): CAiSE 2003, LNCS 2681, pp. 696–710, 2003.

promises them the chance to realize the vision of true one-to-one marketing of their products and services. Privacy rights advocates, on the other hand, fear "database nations" and the manipulative or discriminatory power of customer knowledge in the hands of profit-seeking corporations [Garf2000]. Consumers mostly don't know how much and what kind of data their product suppliers hold about them and what they are doing with it. There is evidence to suggest that, even knowing about such practices, individuals do not fully appreciate the consequences of misuse of their personal data. Consequently, it is hard for individuals to express their preferences for the trade-off between disclosure and access to individualized services and bonuses.

Drawing on the results of a case study of nine companies with significant Internet businesses, this paper investigates a change to the basic model of collecting and storing personal data as a basis for assisting users to gain better control of their privacy. We propose that it is essential to understand corporations, their business models, and the role of consumer data in marketing, in order to develop privacy technologies and frameworks that are acceptable to both companies and individuals. To this end, we look at the business models of two classes of online company, marketers and mediaries, and the role of customer data in them. Our analysis leads us to the conclusion that companies will not willingly give up the opportunity to identify those with whom they do business. We also recognize companies' desire to segment their customer base in order to do personalized or relationship-based marketing.

In contrast to the commercial trend towards personalization, we observe online users' stated desire to maintain their privacy. Reconciling these different views, we examine a one framework for achieving practical privacy based on client-side profiling. Specifically, we look at client-side profiling based on software agents, and show how, in principle, agent technology could provide a means for effective privacy protection. The paper is organised as follows: section 2 gives an overview of companies' data collection and usage practices and the benefits they derive from personalized marketing. It also contains a critical discussion of the benefits that users can gain from personalization practices, and why this concerns privacy advocates. Section 3 then analyses where compromise could be achieved between companies' customer relationship marketing (CRM) aspirations and privacy conscious individuals, based on client-side profiling. Section 4 concludes with a summary of findings.

2 Data Collection and Usage Practices

To evaluate companies' data collection and usage, practices, and some of their expectations for the future, we base our arguments on a) business studies literature on direct marketing, and b) on a case study based on nine extensive interviews that we developed in summer 2001. We interviewed experts in marketing and personalization at some of the most influential and well-known Internet portals, retailers and services providers in Germany and the USA. The interview study was jointly designed by research teams at Hewlett-Packard Laboratories and Humboldt University Berlin. In the following sections we primarily refer to online data collection practices and usage. However; as online and offline channels often work in parallel, and are intermingled

from a company perspective, our report on data collection and usage practices is not exclusively restricted to Internet-centric practices and business models.

2.1 Internet Business and Data Collection Models

Data-mining and CRM are currently important subject on corporate agendas:

> "We need to know our customers better. That's the name of the game. Anything and everything is pretty much useful" [Net2000]

In order to understand why Internet companies regard user data as such a strategic asset, we must understand what role user and customer data plays in those companies' business models. For the purposes of this paper, we classify businesses on today's Internet into two types: marketers and mediaries.

- *Internet marketers* are organisations that derive their profit from selling goods and services to end-consumers through the on-line channel. A typical example for this type of organisation is Amazon.com. Marketers also include traditional offline retailers (e.g. WalMart) and direct marketers (e.g. Otto) that also offer their products to offline purchasers. Marketers derive their revenue from the sale of goods.
- *Internet mediaries* are organisations that offer mediating or supporting services to online users. This includes many different services, including e-mail, newsletters, information portals and referral services. These organisations, which include, for example, Yahoo! and AOL, derive their profit from selling banner advertisements, from collecting monthly user fees, and from transfer provisions.

Today, both, Internet marketers and mediaries collect and hold customer information themselves, which we term the *host model* (see Figure 1 a and b).

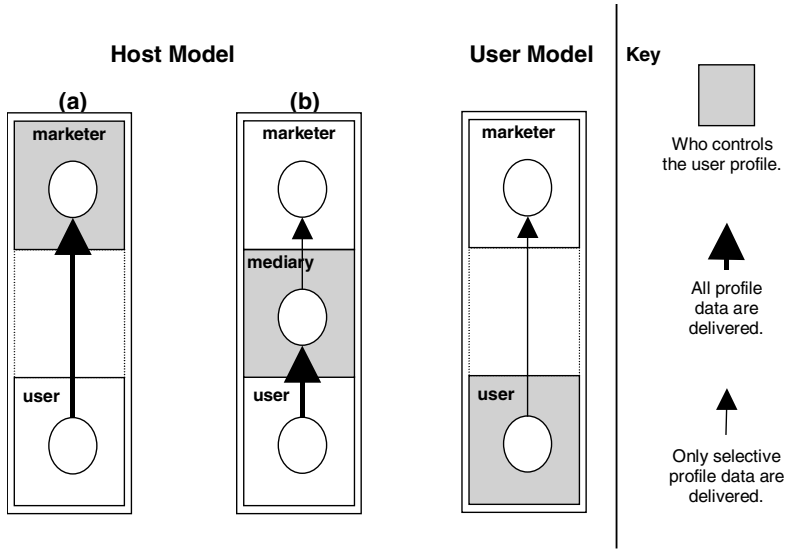

Fig. 1. The host and the user model for profile control

The host model implies that the user has no direct control over what information is stored by a given web site. As technologies such as P3P [P3P2001] achieve more widespread adoption, users will potentially gain a clearer understanding of the precise conditions under which their personal data is collected and used. However, even given this, their options are essentially limited to accepting without reservation the privacy practices of a given site, or using an alternative service. In other words Internet users in the current host model are regularly confronted with a "take it or leave it" decision: data for service (or discounts or access), or no service at all. Whether user data is then collected by a web marketer, or a mediary, once the data leaves the user's purview they cease to have control over it.Good faith or trust in the respective host is what remains as the basis for transactions. We compare this situation to an alternative in which the customer retains control of their data, and relies less on trust (see the "user model" in figure 1).

Whether this faith and trust is always justified is another question. Both marketers and mediaries use user data in many ways to their own profit and not always to the very best of all customers. The following sections will summarise how and why marketers and mediaries use customer data. For this purpose, we look at: service and product differentiation, price discrimination, targeted advertising, and data sales.

2.2 Data Usage Practices, Benefits, and Challenges

We distinguish *internal* and *external* uses of customer data. Internal data use implies that the company uses the data they collect only in order to adjust products and services to their own customers. Customer data thus serves to increase internal efficiency and/or profitability. External data use means that a company uses its user or customer data in order to derive revenue from outside the company. These revenue streams may be the result of targeted advertising or direct data sales. Figure 2 gives an overview.

2.2.1 Internal Data Usage

Our interview study revealed that four out of the five marketers we spoke to currently interact differently with different customers, depending what knowledge they have of that customer. *Service differentiation* means that companies group their customers into different value segments (e.g. A, B, C, and D customers), and then provide differing levels of service to each segment. Value in this context corresponds to the revenue a company has from, or expects to have from a customer (i.e. the number and volume of transactions), combined with the cost he or she creates. The primary reason why companies pursue service differentiation is customer retention.

Our study revealed that four out of the five marketers we spoke to segment their customers on the basis of their current value. Increasingly, however, interviewees plan to factor in the costs, and migrate to a profitability-based segmentation. Retailers in particular want to develop better models of potential future customer revenue.

From a user perspective, service differentiation can be very positively perceived. Consider the popularity of frequent flyer airline programmes. However, from a privacy perspective, we can criticize the systematic classification of people according

to their financial means or spending, as it can lead to direct discrimination against those who cannot or are unwilling to spend.

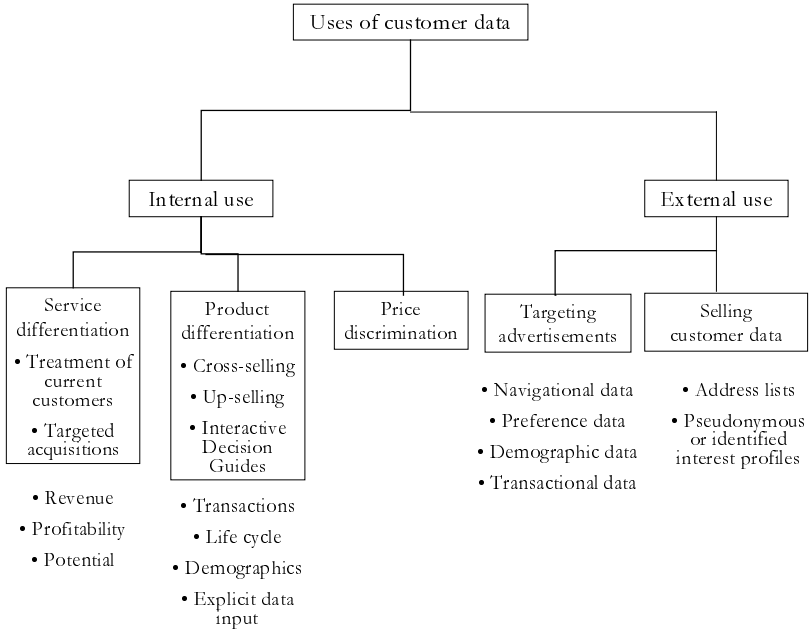

Fig. 2. How Internet companies use customer data

Product differentiation means that, based on the knowledge a company holds about its customers, it recommends products or services that it believes will suit them better. Using a customer's purchase history, demographics and clickstream (web navigation) data, companies try to derive preference and interest profiles. These profiles are then used to recommend products or services that have more functionality (at higher cost), or which enhance a product that the customer has already acquired, or which they have selected for purchase. This practice is known as *up-selling*. Alternatively, the company may recommend additional or complementary products, that might be of interest to the user according to his personal profile. This is known as *cross-selling*.

In our study, four out of five marketers we interviewed reported that they make simple personalized product offerings in the form of cross-selling or up-selling. Instead of sending bulk mails, where all customers receive the same type of offer, companies use targeted mailings or personalized websites to convince a selected group of customers of products that they believe this group to be interested in. One of our interviewees, with several million clients, claimed to offer special products and services to segments of as small as 10,000 recipients with similar profiles. As a result, the company profits from a return on marketing investment around three times higher than if it had contacted all its clients indiscriminately. The interviews confirmed that the challenge in cross-selling and up-selling is that demographic data, purchase

histories, or derived preferences are not always a reliable indicator for future preferences or budget.

From a customer perspective, personalized offers based on up-selling or cross-selling practices can also be positively perceived. It allows them to save time searching for relevant information on products that are potentially of interest to them. They are made aware of offers that they would otherwise perhaps not have seen. Privacy advocates, however, fear the potentially negative effects of current product differentiation practices. Initially, systematic product differentiation has the potential of depriving people of the richness and diversity of offers. In the language of diffusion of innovation, this is called homophilous diffusion [Rog95]. Homophilous diffusion allows rapid diffusion of innovations within one socio-economic group. But diffusion throughout society requires heterophilous diffusion, where individuals seek recommendations from more advanced peers who are unlike them. This type of heterophilous diffusion can be impeded by current recommendation cross-selling and up-selling techniques.

Price discrimination refers to a seller charging buyers different prices for the same commodity. In economic literature, first and second degree price discrimination are distinguished [Ulph2000]. *First-degree* price discrimination arises when each unit of a good can be sold at a different price, while *second-degree* price discrimination occurs when different brands of the same product are sold at different prices. Price-varying markets, such as auctions, are common in online and traditional commerce. The important distinction for price discrimination in the current context is that a single seller may adjust prices according to some characteristic of the buyer while the buyer is not aware of this.

First-degree price discrimination practices for Internet trade were publicly discussed during the summer of 2000, when Amazon.com experimented with demanding different prices for the same DVD from different customers. As the subsequent uproar showed, blatantly variable pricing can cause great image and PR damage. Our interview study confirmed this. Despite its direct impact on profit, half of the participating interviewees did not believe that price discrimination would have a great influence on their profitability in the future, as image considerations would impede its systematic use as a marketing tool. Early bookings, or being the first in a queue, can reward customers as a consequence of 'open rules of the game', where everyone initially has the same chance. The problem with the type of variable prices tested by Amazon.com is of different nature: it arises when discrimination takes place based on one's personal profile, and individuals are unwitting participants with no control.

2.2.2 External Data Usage

Revenue streams from *targeted advertisement*, in particular personalized banner ads, have been the basis of many online business models. There are three main principles on which banner advertising revenues depend: the number of customers who visit a site, the conversion rate, and a website host's ability to segment users. The conversion rate is a measure of your ability to persuade your prospects to take an action. In relation to banner ads it means your ability to persuade X clients to follow a banner link out of a total number of Y clients who must have seen the ad. Whether a client

has really seen a banner ad is measured with the help of webbugs in addition to cookies.

The ability to segment users strongly impacts the revenue stream a website can derive from advertising. There are two driving factors for this: first, better-targeted adverts increase the conversion rate. Secondly, offering potential advertisers a clear and well-defined market segmentation presents a better proposition, and is more likely to attract the advertisers' business, and at higher fees. The better a site knows a user, based on the segmentation, the better can it display adverts attracting the user's interest. It should also be noted that online advertising rates have fallen dramatically since the burst of the "Internet bubble", and consequently many online businesses are having great difficulty in remaining viable with fees from non-personalised advertisements. It is unclear whether even highly personalised advertisements will be sufficient to change the outlook for businesses based solely on advertising revenue.

Our interviews showed that micro-segmentation activities for advertising purposes strongly vary among companies. Thus, while one company claimed to work with around thirty segments, based on demographic data, the leading-edge company interviewed in this context claimed that they use around 700 user segments. These segments are generated by integrating historical user data, demographic data and current clickstream data into the segmentation process.

A targeting practice that benefits both companies and customers alike is facilitating users to self-customize a website. Examples of this type of service include MyCDNow, MyYahoo!, or Amazon.com's wish and recommendation lists. Typically, a user can specify his or her information preferences, and based on these, receives customized content and advertisements. However, even though this personalization service seems to offer a compelling benefit to users, the companies we interviewed reported that only about 10% of their users register to personalise their experience of the site. If the personal space on a website is enhanced with order or account tracking services, this figure rose to around 20% of customers.

Just as is the case with personalized offers, targeted advertising can certainly be seen as a benefit to consumers, as they are made aware of suitable products that they might otherwise miss. However, the problem with homophilous diffusion also arises here. Moreover, using cookies to track users' interests for advertising purposes is becoming a serious privacy problem. This is, because companies serving online advertisements, such as Doubleclick, can track users across multiple web domains. Over a period of time, doing so allows them create more comprehensive interest profiles of online users than any single service marketer or mediary can. This is done without the consent of users, and mostly unnoticed by them.

Another method of external data usage is based on *renting or selling one's data.* Privacy rights advocates fear that personal profiles are becoming a tradable good, over which the owner of the information no longer has any influence. The assumption underlying this fear is that there is a business case for corporations sharing customer data. The Electronic Frontiers Foundation states on its website:

> "Most [people] don't realize the vast information sharing chain that exists once a company or governmental agency obtains your personal information. In some cases, personal information about you that will be shared might contain only a name and an email address. Oftentimes though, personal information can include

name, address, email address, social security numbers, URLs for web sites you've visited, as well as other information that may have been built up about you in a profile." [EFF2002]

Table 1. Summary results from interviews. (Note that the meaning of > and < varies between issues considered.)

		>	no pref	<
(a) data exchange vs. profile completeness **>**: no exchange, but category pooled data **<**: exchange with others, but market basket profiles	**Aggregate**	**75%**	**0%**	**25%**
	Marketer	60%	0%	40%
	Mediary	0%	0%	100%
(b) Contact preferences **>**: online contact (email), **<:** offline contact (postal)	**Aggregate**	**56%**	**0%**	**44%**
	Marketer	40%	0%	60%
	Mediary	75%	0%	25%
(c) High-cost identified profile vs. low cost pseudonymous **>:** high cost, ID **<:** low cost, pseudonymous	**Aggregate**	**75%**	**13%**	**13%**
	Marketer	80%	20%	0%
	Mediary	67%	0%	33%
(d) Identity of user vs. quality/reliability of profile **>:** user identity **<:** profile quality (truthful and rich)	**Aggregate**	**33%**	**44%**	**22%**
	Marketer	40%	40%	20%
	Mediary	25%	50%	25%
(e) reliability/quality vs. control over profile **>:** quality (truthful and rich) **<:** control	**Aggregate**	**78%**	**0%**	**22%**
	Marketer	80%	0%	20%
	Mediary	75%	0%	25%
(f) Identified profile with 30% false data vs. pseudonymous profile, 100% accuracy **>**: identity, 30% error **<:** pseudonym, no error	**Aggregate**	**25%**	**13%**	**63%**
	Marketer	40%	20%	40%
	Mediary	0%	0%	100%

The companies we interviewed, however, all of which have a direct end-user customer base, do not have a strong interest in selling or sharing rich data on their customers. There are two major reasons for this. First, selling knowledge about customers would mean a company giving up some competitive business advantage. As described above, knowing a customer well (and, by implication, better than a

competitor) has a direct impact on profitability. Secondly, selling knowledge about customers entails a strong risk to goodwill or image if customer harm, perceived or real, arises from such sales. Conversely, buying data is exposed to the risks arising from the reliability of the source. As poor promotion is regarded as detrimental, companies fear putting customers into wrong segments due to low quality data purchased externally. Consequently, we found that companies are hesitant to purchase more than address lists and socio-economic classifications from external sources. Thus, good marketing practices prohibit, at least to some extent, the sharing of enriched customer data.

In fact, none of the companies interviewed generated revenue from selling rich customer data. Only 22% said that they would sell address or mailing lists. And when asked to choose between a category-pooled profile (where only the spending of customers in one product category is known) that is not shared with others, or a market-basket profile (where the entire consumption pattern of a consumer is know) but which is shared with others, both marketers and mediaries preferred the less rich profile that they would not have to share (see Table 1 (a)).

However, concerns about the trade of customer data are not completely unfounded. All companies we interviewed also agreed that market-based profiles are more valuable than category-restricted profiles (see Figure 3). Thus, pooling is seen as a valuable profile enhancement. Also, there are scenarios in which sharing of customer data occurs beyond trading. One is where a company that owns customer data is taken over, or its assets are bought by another company. A second scenario in which customer data is pooled is when 'friendly' companies from different business areas, and with equally rich profiles, share data on equal terms to enhance their profiles.

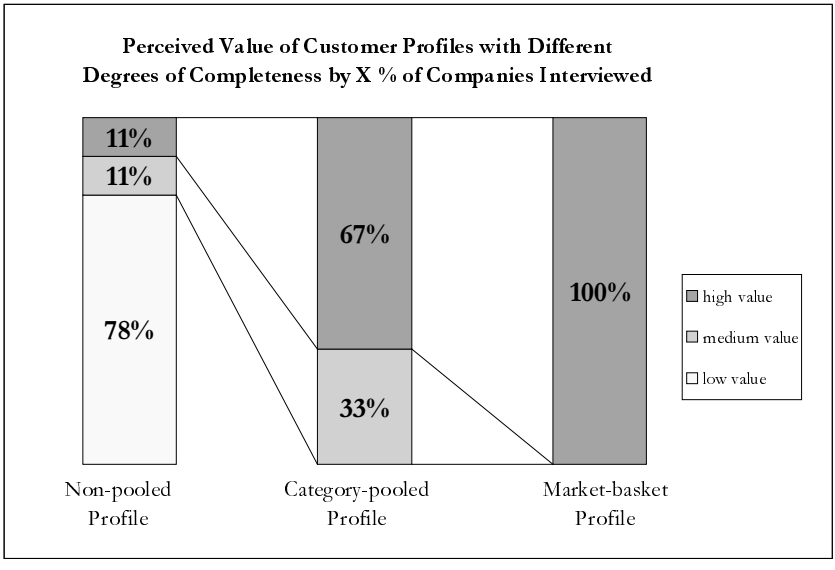

Fig. 3. Perceived value of customer profiles

3 Information System Design Implications from Data Usage Practices

3.1 Case Study Findings on Pseudonymity, Data Reliability, and Control

The above description of marketing practices and business models summarises how customer personal data is used by businesses today. Internally, marketers, particularly, use customer data to ensure higher campaign response rates and increase customer loyalty while reducing marketing expenses. Externally, customer data can be used to directly increase revenue from advertisements, or raise revenue through data rental.

Given these common business practices, developers of privacy enhancing technologies must recognize that companies view customer data as a valuable asset and will insist on having the possibility to segment customers and personalize marketing material. Moreover, as was outlined above, personalization also brings benefits for consumers that many may wish to take advantage of, or have already done so.

Against this background, it is questionable to what extent privacy enhancing technologies can be successful that lead to a default 'hiding' of users and denial of information sharing practices. Current anonymizing services such as Anonymizer.com, Freedom or JAP (Java Anon Proxy) pursue this strategy, supported by cookie management software including Junkbuster, WebWasher or CookieCooker. However, by protecting users by default from information revelation, they also deprive them of the benefits of personalization. What follows is that neither companies nor users seem to be enthused about using these software solutions; especially not to the point that they would be willing to pay for them. Given this conflict of interest between personalization and privacy, many privacy technology developers have started to propose the idea of identity management systems that protect users' privacy on the basis of transaction or relationship pseudonyms [Jend2000, Köhn2000, Libe2001]. Pseudonyms can protect a person's identity, but still allow for a personalized relationship between a customer and a company. A user can simply choose to re-use the same pseudonym to build a business relationship over time. This relationship pseudonym can then be used by the marketers or mediaries as an identifier to build up a personal profile [Köhn2000, Jend2000, Bert2000]. Marketing communications can consequently be personalized without forcing the user or customer to permanently reveal his or her true identity.

With a view to these technological advancements, we asked companies in our interviews for their views on the concept of pseudonymity in electronic business relationships. The responses revealed that, from a business perspective, there are some major challenges for the pseudonymity concept: First, companies believe that a customer profile gains in value for them, the more identified it is (Figure 4). This is because they view themselves as having an active role in commerce, which implies that they need to personally address customers and users to pro-actively market products and services. Also there is a strong belief that addressing a customer personally increases the perception of service quality [Kief2001]. And, as other

authors have pointed out [Clark], there equally is an emotionally strong scepticism towards the systematic use of pseudonyms in business relations. Short-term pseudonyms or anonymous communications preclude the possibility of offering high-quality personal service, and impede the personalized marketing practices described in section 2. Relationship pseudonyms would allow for personalized marketing, but, in its current technical realization, precludes addressing the customer by their name. Finally, the use of pseudonyms presents challenges for the management of delivery or after-sales services.

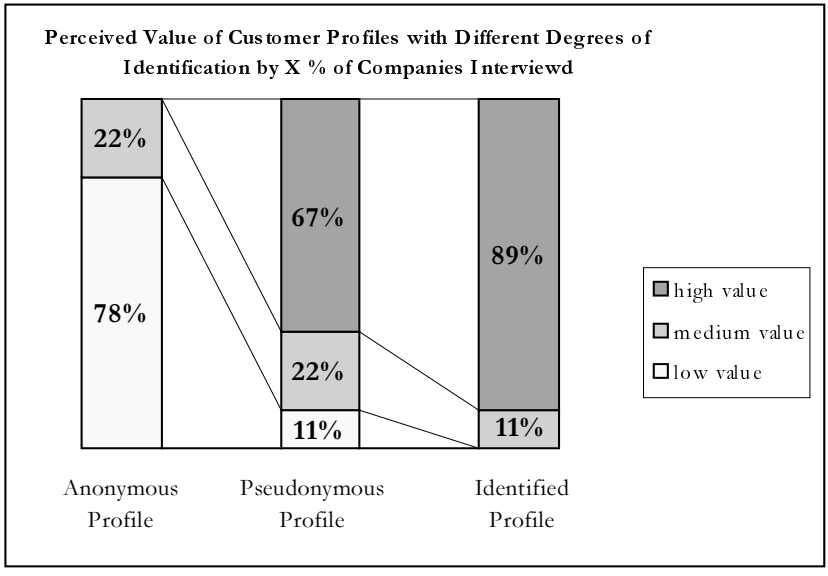

Fig. 4. Perceived value of profile identification

In addition to asking companies about the economic value of different profile types, we also asked them to express their relative preference for pairs of bipolar alternatives that would display different qualities in terms of privacy, control and reliability (Table 1, b and c). The trade-off preferences expressed by our interviewees showed that all parties are willing to invest having identified data rather than pseudonymous data (high cost identified profiles versus low cost pseudonymous profiles). Marketers in particular want to have an offline postal address from their customers, rather than just an online e-mail address.

All these arguments indicate that companies operating in the current host model will probably be opposed to implementing a privacy system on their side that supports the systematic use of pseudonyms. The profile collected would lose economic value, such at least is the current view of companies on the issue (see Figure 4).

However, as discussed above, even if companies wish to identify customers, they must still be concerned about profiles that are of low reliability. In fact, the quality and reliability of a profile, especially as far as customer preferences are concerned,

were reported to be of concern for companies fearing flawed customer segmentation and flawed personalization. This is reflected in interviewees' responses when presented with a trade-off between identifying a customer, and having a reliable profile (Table 1, (d) through (f)). The results show indecision. This is a significant finding, given the arguments above where personal identification of customers appears to be a major requirement for businesses (see also the valuation of identified profile versus pseudonymous profile in figure 4). Furthermore, all companies stated a preference for a highly reliable and rich profile over actual physical profile control. And finally, interviewees also displayed a tendency to prefer 100% correct pseudonymous contact and demographic customer data than the data being identified, but with a 30% lying rate.

On the basis of these findings, we conclude that despite the high valuation of identified and rich profiles, companies are, if they are forced to make choices, willing to make some sacrifice in order to ensure the reliability of their customer data. Companies also recognize that high quality and reliable profiles are difficult to accumulate if people are afraid of losing their privacy online. Currently, they observe a high rate of false data in their databases due to people lying on their demographics and contact information. As Sheehan et al. report, 15% of individuals falsify information more than half the time they are asked to provide it. [Shee1999]. A very recent IFAK study [IFAK2002] found an even higher rate of false data provision: around 47% of 1200 Internet users interviewed stated that they sometimes lie about their e-mail address. As people become more computer and information literate, this rate could even rise, as could eventually the use of protective software.

3.2 Towards a Compromise: Client-Side Profiling and Agent Technology for Privacy Support

Given the above insights into companies' data usage practices and profiling preferences, we can essentially observe the following paradox: vendors want accurate, detailed information about their customers, which consumers often are willing to supply in order to gain access to better outcomes, but privacy advocates (and more aware consumers) worry about the misuse of identifiable, detailed personal information. Vendors want data, consumers want the benefit of the vendor having the data, but neither they nor their privacy guardians want detailed disclosure.

Two extremes of this argument are frequently advocated. At one extreme, it is suggested that full disclosure is economically necessary, and that market forces and self-regulation will curb the worst excesses. At the other extreme, it is suggested that those with a financial interest in our data can never be trusted to be responsible, and hence all possible measures should be taken to prevent disclosure. Clearly, there is scope for a middle ground.

We suggest that practical privacy means allowing a reasonable degree of disclosure, under controlled conditions, and clearly in exchange for a benefit to the user. The starting point for this practical privacy may be the fact that companies seem to be ready for compromise when it comes to the reliability and quality of the data they can receive from their customers. Moreover, data that is being provided out of free will by users, driven by the desire to receive personalized services and not apt to the risk of a

backlash on privacy. For this type of highly reliable data, companies reported, marketers and mediaries alike, that they would sacrifice some control over data, meaning that the data does not have to be stored necessarily in their own databases.

From this starting point, we propose the use of a client-side software agent, profiling the user under his or her control, and managing personalized commercial relationships for the user in a privacy-friendly manner. The software agent realizes what we labelled the user model of a company-client relationship (Figure 1), because the user in this scenario would take control over data revelation and data use.

The user-model is based on a framework of mediation between users and companies (and other entities) on the WWW and includes the following key agent capabilities:

1. *Building and maintaining a rich profile under the user's control and direct sphere of influence (client-side profiling).* The central idea is that profiles are not developed exclusively by companies on the server or host side, but by software on the user's client computer. Thus the profile is built under the user's control. Parts of that personal profile can then be placed at the disposition of marketers or peer networks in order to receive appropriate and personalized recommendations.
2. *Managing multiple user identities.* Clearly it is not possible to prevent another party from observing, and hence recording, utterances (here: in the very general sense of units of interaction, such as clicking on a URL within a web site) in an interaction, since otherwise no discourse could occur. The only way in which people can prevent a decrease in their privacy despite progressive disclosure of their profiles is to manage their identities. A number of identity management systems are now being proposed by researchers [Köhn2000, Jend2000, Bert2000, Libe2001]. These tools will be able to assist online users in controlling their virtual identities, and ensure that customers revisit sites under the same virtual identity if they wish to (situational pseudonyms) but not necessarily under their true physical one.
3. *Providing an innovative user interface to assist the user to better manage their privacy settings.* Privacy, like security, is asymmetric in the sense that having it seems to have a low value, but losing it has a high cost. This leads users typically to devote little time and effort to maintaining privacy settings, or attending to privacy risks, even though the potential negative consequences are ones they wish to avoid. We therefore suggest that a key role for a privacy *agent*, in addition to being a repository for the client-side profile and identity manager, is to engage the user in an appropriate conversation around privacy preferences, and redistribute the burden of maintaining the relevant safeguards (e.g. by managing the settings of web browsers or learn by observation of the user to fine-tune his or her privacy preferences).
4. *Managing user relationships using privacy metadata.* The decision to put private information at the disposition of a third party is a choice that should be made on the basis of the sound knowledge of the reputation of the other party. This suggests an important additional privacy measure: the pro-active use of knowledge about web services and their privacy practicesas part of reputation index that indicates whether a party is trustworthy to receive ones' personal data. Such data about a service or organisation is known as *metadata*, particularly if it is machine processable. An example of such metadata is a web site's privacy

policy – a statement about that site's use of personal information. In particular, the *Platform for Privacy Preferences Project (P3P*) standard has been defined by the W3C to allow web sites to make their privacy policies machine-readable [P3P2001]. If metadata about a company's privacy (or, more particularly, lack of privacy) practices is available, a range of new services are enabled that can assist both user and vendor.

4 Conclusion

This paper has surveyed a number of aspects of protecting users' privacy in business-to-consumer (B2C) settings. In the first part of the paper we focused on the results of an interview study with nine significant Internet industry companies. We analyzed their current practices for collecting and using data on their customers, and critically discussed them from a user as well as privacy rights perspective. Based on these interview-based insights we conclude that companies fear a move to a (perceived) more passive role in commerce, a scenario in which users can choose whether or not to identify themselves. On the other hand, we also noted companies' wish for more reliable, complete and timely profile data. The wish for this type of higher quality user data may be an incentive for companies to accept and work with profiles generated on the client side. This would be a compromise between user privacy concerns and personalization benefits for both users and vendors.

Pursuing a client-side profiling strategy however, also bears some risks. Once a user reveals comprehensive data about him or herself, the circumstances under which he does so must be clearly defined. His choice must be taken on the basis of knowledge on the opposite party's reputation. Furthermore, he must have the option to shelter his identity when revealing comprehensive information. This again calls for a sophisticated identity management system operating under the user's control. Finally, all measures for privacy protection are in vain if people are not capable of using them. This means that they need not only default protection settings in their privacy tools, but also an easy to use interface supporting them in their daily privacy management tasks.

References

[Bert2000] Oliver Berthold, Hannes Federrath: Identitätsmanagement; in: Helmut Bäumler (Hg.): E-Privacy; Tagungsband zur Sommerakademie des Unabhängigen Landeszentrums für Datenschutz Schleswig-Holstein, 28. August 2000 in Kiel; Vieweg, Wiesbaden 2000; 189–204

[Bick2000] Bicknell, C., "Online Prices Not Created Equal", in: Wired News, 7th September 2000. Available from: http://www.wired.com/news/ business/0,1367,38622,00.html

[Duff2001] Duffy, D., "Get ready for the privacy backlash", Darwin B2B Network, August 2001. Available from: http://www.darwinmag.com/read/ 080101/backlash.html

[EFF2002] EFF Topics: Privacy: Marketing and Commercial. Electronic Frontier Foundation. 25th April, 2002. Available from: http://www.eff.org/ Privacy/Marketing/

[Garf2000] Garfinkel, Simon, "Database Nation – The Death of Privacy in the 21st Century", O'Reilly, 2000

[Gart2001] DeLotto, R., „Client Self-Profiling Protects Consumer Privacy“, Gartner Research Note, Strategic Planning SAP-13-6930, September 2001

[Gold1997] Goldberg, Ian; Wagner, David and Brewer, Eric “Privacy-enhancing technologies for the Internet”. 1997. Available from: http://www.cs.berkeley.edu/~daw/papers/privacy-compcon97-www/privacy-html.html

[Hage1997] Hagel and Rayport, “The coming battle for Customer Information”, McKinsey Quarterly, no 3, pp. 64–77, 1997

[IFAK2002] IFAK study on online communication, January 2000, available at: http://www.ifak.de/about/index.php3

[Jend2000] Uwe Jendricke, Daniela Gerd tom Markotten: Usability meets Security – The Identity-Manager as your Personal Security Assistant for the Internet (PDF); in: Proceedings of the 16th Annual Computer Security Applications Conference (ACSAC 2000); New Orleans, USA; December 11–15, 2000

[Kief2001], Kiefel, Nicola, “Perspektiven von Personalisierungs-Marktmodellen und Verwendung von Kundendaten durch Unternehmen“, unpublished master thesis, Berlin, September 2001

[Köhn2000] Köhntopp, M., Pfitzmann, A., “Datenschutz Next Generation“, in: E-Privacy, ed. by Helmut Bäumler, Wiesbaden, 2000, pp. 316–322.

[Kotl1994] Kotler, P., Marketing Management – Analysis, Planning, Implementation, And Control, 8th edition, New Jersey, 1994

[Libe2001] The Liberty Alliance Project [Web Page]. 2001. Available from: http://www.projectliberty.org

[Lieb1996] Lieberman, Henry and Maulsby, David. Instructible agents: software that just keeps getting better. IBM Systems Journal. 1996; 35(3 - 4):539–556

[Micr2001a] Microsoft. Introducing .NET My Services [Web Page]. 2001 Sep. Available from: http://msdn.microsoft.com/library/default.asp?url=/library/en-us/Dndotnet/html/Myservintro.asp?frame=true

[Micr2001b] Microsoft Internet Explorer version 6 features. August 27, 2001. See http://www.microsoft.com/windows/ie/evaluation/features/default.asp

[Net2000] Net Genesis, “E-Metrics – Business Metrics For The New Economy”, Cambride 2000

[P3P2001] World Wide Web Consortium (W3C). The Platform for Privacy Preferences Project (P3P) 1.0 Specification [Web Page]. 2001. Available from: http://www.w3.org/TR/p3p

[Pew2000] Pew Internet & American Life Project, 2000. www.pewinternet.org

[Rog95] Rogers, Everett M., Diffusions of Innovations, Fourth Edition, The Free Press, 1995

[Shea2001] Shearin, S., Liebermann, H., “Intelligent Profiling by Example”, Proceedings of the Conference on Intelligent User Interfaces, IUI’01, Santa Fe, 2001

[Shee1999] Sheehan, K., Hoy, M., “Flaming, complaining, abstaining: How online users respond to privacy concerns”, Journal of Advertising, Fall 1999