LORRIE FAITH CRANOR

'I DIDN'T BUY IT FOR MYSELF'

Privacy and Ecommerce Personalization[*]

1. INTRODUCTION

Ecommerce web sites are increasingly introducing personalized features in order to build and retain relationships with customers and increase the number of purchases made by each customer. While survey data, (Personalization Consortium, 2000; Personalization Consortium, 2001), user studies (Karat, *et al*, 2003), and experience (Manber, 2000) indicate that many individuals appreciate personalization and find it useful, personalization also raises a number of privacy concerns ranging from user discomfort with a computer inferring information about them based on their purchases to concerns about co-workers, identity thieves, or the government gaining access to personalization profiles. In some cases users will provide personal data to web sites in order to receive a personalized service, despite their privacy concerns; in other cases users may turn away from a site because of privacy concerns (Ackerman, *et al.,* 1999; Culnan and Milne, 2001; Cyber Dialogue, 2001, Berk, 2003). As recent studies have suggested that the benefits to a web site of offering personalized services often do not outweigh the costs (Berk, 2003), it is important to consider ways of designing personalization systems that will maximize the return on the investment. Improving the privacy associated with these systems so that web site visitors are more willing to trust and use them is a step in that direction.

This chapter begins with a discussion of the privacy risks associated with personalization. It then provides an overview of the fair information practice principles and discusses how they may be applied to the design of personalization systems, and introduces privacy laws and self-regulatory guidelines relevant to personalization. Finally, the chapter describes a number of approaches to personalization system design that can reduce privacy risks.

2. PRIVACY RISKS

Ecommerce personalization poses a variety of risks to user privacy. Most of these risks stem from the fact that personalization systems often require that more

[*] This chapter is a revised version of a paper presented at the 2003 ACM Workshop on Privacy in the Electronic Society (WPES).

This is an unedited draft of a paper that will appear in: Clare-Marie Karat, Jan O. Blom, and John Karat (ed.), Designing Personalized User Experiences in eCommerce. Kluwer Academic Publishers. 2004.

personal data be collected, processed, and stored than would otherwise be necessary. This section provides an overview of several privacy risks that may be caused or exacerbated by ecommerce personalization systems. These risks are summarized in Table 1.

Risk	**Examples of possible consequences**	**Examples of parties to whom personal information might be exposed**
Unsolicited marketing	Unwanted email, postal mail, and telephone calls; time wasted deleting email, throwing away mail, answering calls	Employees of personalized web site; employees of companies to whom marketing lists are sold; employees of companies that perform marketing services
Computer "figuring things out" about me	Individuals feel uncomfortable or embarrassed; characteristics inferred by computer become available to people who would otherwise not know this information; inaccurate information inferred by computer becomes available to people who believe it to be accurate	Employees of personalized web site; any other parties that gain access to profile
Price discrimination	Individuals are treated differently based on profile; higher prices	Employees of personalized web site
Information revealed to other users of same computer	Other users of computer may learn confidential information; other users of computer may be able to gain access to accounts	Other users of computer such as family members or co-workers
Unauthorized access to accounts	Identity theft, fraud, stalking	People that run personalized web site, someone who steals password
Subpoena	Information used against individual in court case	Law enforcement officers or participants in legal dispute; public (if information obtained becomes part of public record)
Government surveillance	Individual could be detained by law enforcement for questioning or arrested	Law enforcement officers

Table *1. Privacy risks from ecommerce personalization*

One of the first privacy risks that Internet users mention is unsolicited marketing (Cranor*, et al.,* 2003). Arguably, the consequences of unsolicited marketing are less severe than the potential consequences of some of the other privacy risks discussed here. Nonetheless, this risk is of great concern to users, and a strong desire not to receive unwanted marketing communications may be a factor in some users' decisions not to engage in ecommerce (Culnan and Milne, 2001; Cyber Dialogue, 2001). Users have concerns that information they provide for use in personalized ecommerce may be used to send them targeted advertising, or may be sold to other companies that may advertise to them. They often fear that the more a company knows about them, the greater the interest that company will have in marketing to them.

Many users are also concerned about a computer "figuring things out" about them. They are not comfortable with the idea that a computer might be able to make predictions about their habits and interests. In some cases, individuals are frustrated because the computer's predictions appear to be off base and they are afraid that someone might find out and draw incorrect conclusions as a result. In other cases, individuals are concerned because the computer's predictions are uncannily accurate, and perhaps reveal information that they thought other people didn't know about them. Some users of the TiVo digital video recorder have been surprised at the television selections their TiVo makes for them based on their TV viewing history, and some even believe their TiVo has made inferences about such personal characteristics as their sexual preference (Zaslow, 2002). Regardless of the accuracy of a computer's inferences and prediction, many individuals are simply uncomfortable with the idea that their activities are being "watched." Additional concerns arise when there is a mismatch between users' perceptions about privacy and the types of data collection and use that actually occur (Adams, 1999).

Individuals are also concerned that companies will profile them in order to facilitate price discrimination. While economists point out that price discrimination can often benefit both businesses and consumers, consumer reaction to price discrimination is usually quite negative. In addition, effective price discrimination often leads to increases in the amount of personal information associated with a transaction (Odlyzko, 2003). Individuals may be concerned not only about the possibility of being charged higher prices because of information in their profile, but also about the fact that they are being treated differently than other people (Turow, 2003).

Another privacy risk associated with personalization is that users may inadvertently reveal personal information to other users of their computer. When cookies are used for authentication or access to a user's profile, anyone who uses a particular computer may have access to the information in a user's profile. This leads to concerns such as family members learning about gifts that may have been ordered for them and co-workers learning about an individual's health or personal issues. In addition, when profiles contain passwords or "secret" information that is used for authentication at other sites, someone who gains access to a user's profile on one site may be able to subsequently gain unauthorized access to a user's other accounts, both online and offline.

The possibility that someone who does not share the user's computer may gain unauthorized access to a user's account on a personalized web site (by guessing or stealing a password, or because they work for an ecommerce company, for example) raises similar concerns. However, while family members and co-workers may gain access inadvertently or due to curiosity, other people may have motives that are far more sinister. Stalkers and identity thieves, for example, may find profile information immensely useful. Ramakrishnan *et al.* (2001) have also suggested ways that users may be able to probe recommender systems to learn profile information associated with other users.

A risk that most people don't consider is that the information in their profile may be subpoenaed in a criminal case or in civil litigation. For example, increasingly Internet records are subpoenaed in patent disputes, child custody cases, and a wide variety of lawsuits. Information about what someone has purchased, eaten, read, or posted is proving important to many cases. In addition, other types of profile information that may reveal interests, habits, or personal preferences may be important, especially in cases where the character of the plaintiff or defendant is important. Much of this information may be logged by ecommerce systems that store transaction records, even if they offer no personalization. However, a personalized system will typically store information that goes beyond transaction records, and may potentially store the information for a longer period of time than would be necessary if it were used only to support a transaction.

Finally, as the United States and other governments have been initiating increasing numbers of surveillance programs in the name of fighting terrorism, the possibility that information stored for use in ecommerce personalization may find its way into a government surveillance application is becoming increasingly real. This places users of these services at increased risk of being subject to government investigation, even if they have done nothing wrong.

As new personalization applications are developed that take advantage of a wider range of information (such as information in a user's calendar or address book), or are designed to run on mobile devices and take advantage of information about a user's precise physical location (Gandon and Sadeh, 2003; Warrior, *et al.,* 2003), additional privacy concerns are likely to emerge. The privacy risks discussed here are all likely to become magnified in these new environments.

3. APPLYING FAIR INFORMATION PRACTICE PRINCIPLES

Several sets of principles have been developed over the past three decades for protecting privacy when using personal information. The Organization for Economic Co-operation and Development (OECD) Guidelines on the Protection of Privacy and Transborder Data Flows of Personal Data (1980) are one of the best-known sets of fair information practice principles. Many other sets of guidelines and some privacy laws are based on these principles.

The eight OECD principles provide a useful framework for analyzing privacy issues related to ecommerce personalization. The principles are paraphrased here and discussed in the context of ecommerce personalization. In these principles, the term *data subject* refers to the person about whom data has been collected, and the term *data controller* refers to the entity that controls the collection, storage, and use of personal data.

Collection Limitation. Data collection and usage should be limited. In the context of ecommerce personalization, this suggests that personalization systems should collect only the data that they need, and not every possible piece of data that they might find a need for in the future. The approaches described in Sections 5.2 and 5.3 can also serve to limit data collection.

Data Quality. Data should be used only for purposes for which it is relevant, and it should be accurate, complete, and kept up-to-date. In the context of ecommerce personalization, this suggests both that care be taken to make sure data is used for relevant purposes (that is, don't use data to make inferences that are irrelevant to the data), and personalization systems should provide the ability for individuals to update and correct the information in their profiles.

Purpose Specification. Data controllers should specify up front how they are going to use data, and then they should use that data only for the specified purposes. In the context of ecommerce personalization, this suggests that users be notified up front when a system is collecting data to be used for personalization (or any other purpose). Privacy policies are often used to explain how web sites will use the data they collect. However, by also providing notice about data use at the time the data is collected, sites can more effectively bring this information to the attention of users at the time when it is most relevant. Software tools such as P3P-enabled web browsers may also assist in conveying meaningful privacy notices to users (Cranor, 2002).

Use Limitation. Data should not be used or disclosed for purposes other than those disclosed under the purpose specification principle, except with the consent of the data subject or as required by law. In the context of ecommerce personalization, this suggests that data collected by personalization systems should not be used for other purposes without user consent. This also suggests that sites that want to make other uses of this data develop interfaces for requesting user consent.

Security Safeguards. Data should be protected with reasonable security safeguards. In the context of ecommerce personalization, this suggests that security safeguards be applied to stored personalization profiles and that personalization information should be transmitted through secure channels.

Openness. Data collection and usage practices should not be a secret. In the context of ecommerce personalization, this suggests, as with the Purpose Specification Principle, that users be notified up front when a system is collecting data to be used for personalization. Users should be given information about the type of data being collected, how it will be used, and who is collecting it. It is especially important that users be made aware of implicit data collection.

Individual Participation. Individuals should have the right to obtain their data from a data controller and to have incorrect data erased or amended. In the context of ecommerce personalization, this suggests, as with the Data Quality principle, that users be given access to their profiles and the ability to correct them and remove information from them.

Accountability. Data controllers are responsible for complying with these principles. In the context of ecommerce personalization this suggests that personalization system implementers and site operators should be proactive about developing policies, procedures, and software that will support compliance with these principles.

Table 2 provides a summary of the OECD principles and how they can be applied to ecommerce personalization. The lessons for ecommerce personalization derived from each principle can be expanded further in the context of a specific application. For example, Patrick and Kenny (2003) have performed a similar analysis and made detailed user interface design recommendations for an Internet job search tool.

Principle	Lessons for ecommerce personalization
Collection limitation	Collect only the data you need
Data quality	Don't use data to make inferences irrelevant to the data; provide mechanisms for individuals to update and correct information in their profiles
Purpose specification	Tell users when data is used for personalization
Use limitation	Don't use personalization data for other purposes without user consent
Security safeguards	Take reasonable security precautions with stored personalization profiles and transmit personalization information through secure channels
Openness	Tell users when data is being collected for personalization and make sure they are aware of implicit data collection
Individual participation	Provide mechanisms for individuals to update and correct information in their profiles
Accountability	Be proactive about developing policies, procedures, and software that will support compliance with these principles

Table *2. OECD privacy principles and their lessons for ecommerce personalization*

4. PRIVACY LAWS AND SELF-REGULATORY GUIDELINES

Privacy laws and self-regulatory guidelines can influence the types of personalization systems that can be deployed in practice. Here is an overview of some of the ways laws and guidelines may impact ecommerce personalization systems. It is by no means a comprehensive review of privacy laws or guidelines.

In the United States, most privacy laws are sector-specific. In many sectors, no privacy laws restrict personalization systems on ecommerce web sites. However, financial sites, children's sites, and health-related sites may need to design their personalization systems carefully to comply with legal requirements. For the most part this involves providing adequate notice about the personalization system. In some sectors, there are restrictions on third party sharing of data that may be relevant. Children's web sites are prohibited from collecting personally identifiable information from children under age 13 without consent of a parent. In addition, US sites need to be aware of any state laws that may impact them as well as the privacy laws in other countries where some of their customers may reside.

US companies that provide targeted advertising services to multiple web sites and are members of the Network Advertising Initiative (NAI) must comply with the NAI Principles (2000), which are enforceable by the US Federal Trade Commission. These principles prohibit use of sensitive data in targeted marketing and require that merger of personally identifiable information with previously collected non-personally-identifiable information occur on an opt-in basis only. They also require companies to provide adequate notice, allow individuals to access their information, and offer opt-out opportunities.

A number of other industry organizations such as the Online Privacy Alliance, the Direct Marketing Association, and the Personalization Consortium have adopted self-regulatory guidelines that may be applicable to their members' ecommerce personalization efforts.

In Europe, comprehensive privacy laws impact the design of ecommerce personalization systems across every sector. These laws, which are based on the OECD principles, require privacy notices and access provisions and restrict secondary uses and third-party data sharing. Kobsa (2002) analyzed the European Data Protection Directive and the German Teleservices Data Protection Act and found a number of restrictions that would affect ecommerce personalization on sites under the jurisdiction of German law. For example, raw data from usage logs must be deleted after each session and usage logs from different services must not be combined, except for accounting purposes. In addition, anonymous and pseudonymous services must be provided when possible, and user profiles must always be pseudonymous. These laws also restrict the ability of sites to fully automate decisions that would have significant impacts on individuals (for example, related to employment, credit, etc.).

5. REDUCING PRIVACY RISKS IN ECOMMERCE PERSONALIZATION

The previous sections have identified privacy risks and outlined privacy-related legal requirements, guidelines, and principles that are relevant to ecommerce personalization. This section discusses several approaches to system designs that reduce privacy risks and make privacy compliance easier. No single approach to ecommerce personalization will always provide the desired functionality while protecting privacy. There are tradeoffs associated with each of these approaches.

The degree of privacy risk posed by an ecommerce personalization system is often directly related to the type of personalization the system performs. Section 5.1 describes four axes of personalization and discusses where on each axes the more privacy-friendly personalization systems tend to fall. Sometimes other system requirements prohibit a design that falls on the privacy-friendly end of each of these axes, however. In this case designers may need to take steps to add privacy enhancements to a system design, using the fair information practice principles as a guide. For example, the collection limitation principle suggests system designs that minimize the amount of personally identifiable data stored by the ecommerce web site. This in turn reduces the risk that data may be misused by the company or its employees, limits exposure in the event of a security breach, and minimizes the amount of data that might be subject to subpoena. Section 5.2 and 5.3 discuss two approaches to data minimization that may be useful for designers of ecommerce web sites: pseudonymous profiles and client-side profiles. Section 5.4 discusses the importance of designing systems that put users in control, addressing the data quality and individual participation principles and supporting the ability to request consent from users in compliance with the use limitation principle. Of course, to be effective, all of these approaches need to be augmented by appropriate security safe guards and well-articulated privacy policies that are enforced throughout an enterprise.

5.1 Types of personalization systems

Several general types of personalization systems are considered here that differ on four axes, as illustrated by Figure 1. In this chapter the two extreme ends of each axis are discussed. However, many personalization systems include components representative of both ends and thus fall somewhere in the middle of the spectrum. The ends of each axis are labelled in the table according to whether they tend to be more or less privacy invasive. I have used the word "tend" here because there are exceptions. In general it is relatively easy to design a privacy-friendly personalization system if it is placed on the end of each of the four axes where systems tend to be less privacy invasive. Designing a privacy-friendly personalization system that sits on the other end of these axes is possible, but requires that mechanisms be put in place to reduce privacy risks and concerns.

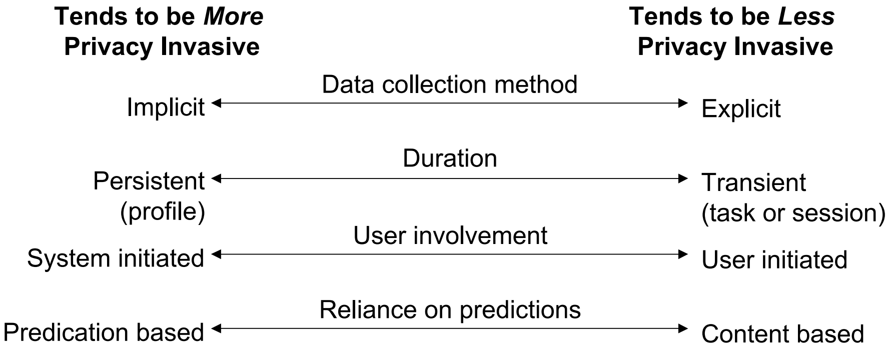

Figure *1. Four axes of personalization systems and their impacts on privacy*

5.1.1 Data collection method

- *Explicit data collection.* Personalization is based on demographics, preferences, ratings, or other information explicitly provided by a user. Typically, *recommender* personalization systems require users to rate a number of items in order to receive recommendations about other items that may interest them. Other systems allow users to create personal pages or customize their view of a site based on their personal preferences or demographics.
- *Implicit data collection.* Personalization is based on information inferred about a user. For example, a user's search queries, purchase history, or browsing history may be used to infer interests or preferences (Claypool, *et al.,* 2001).

Systems that use explicit data collection methods tend to be more privacy-friendly than systems that use implicit data collection methods because users are more aware that data collection is taking place and may be able to make a conscious decision about whether or not to provide their data. When data is collected implicitly, systems can be made more privacy-friendly through the use of easy-to-understand notices and opportunities to control what information about themselves gets collected and stored. This is discussed further in Section 5.4.

5.1.2 Duration

- *Task- or session-focused personalization.* A simplistic way of providing task-focused personalization is to place advertisements on pages where they are most obviously relevant—for example, advertising pay-per-view boxing matches in the sports section of a news site and cookware in the home and garden section. A more sophisticated way of providing task-focused personalization is to make suggestions based on actions a user has taken while performing a task (Herlocker and Konstan, 2001). For example, if a user places a pair of women's running shoes in her shopping basket, a web site might suggest that she also purchase athletic socks, running shorts, or a sports bra. Such personalization is based on information provided by or inferred from the user during the current session or while completing the current task.
- *Persistent profile-based personalization*. Many personalization systems develop profiles of users and add explicitly provided or inferred information about users each time they return to the site. Cookies may be used to recognize returning visitors automatically and retrieve their stored profiles, or users may be asked to login to the site.

A focus on task- or session-based personalization reduces privacy concerns and facilitates compliance with privacy laws because little or no user profile data need be stored in order to facilitate personalization (Herlocker and Konstan, 2001). A session cookie might be used to store some information temporarily, but that information can be deleted at the end of the user's session.

Depending on the goals of personalization, task-based personalization may be able to provide many of the benefits of profile-based personalization. It may be sufficient to know only the kind of task in which the user is currently engaged rather than information about her preferences or past activities. Focusing on a user's current task may allow for a simpler system architecture that need not facilitate the storage and retrieval of user profile data. In addition, it eliminates the problems that may occur when a system offers recommendations to a user that are consistent with her overall profile but not relevant to her current task. For example, when a user is shopping for a gift for someone else, recommendations based on her personal preferences may not be relevant. Likewise, once a user completes a particular task, she may no longer be interested in receiving recommendations related to that task. For example, while a user may be interested in advertisements from car dealers while she is shopping for a new car, once she has completed the purchase these advertisements will no longer be relevant to her.

Personalization derived directly from a user's request rather than from predictions based on that request allows for less data to be stored and fewer privacy concerns. A system that simply reports the availability of other products in the same category of products a user has expressed interest in, for example, is unlikely to raise the kinds of concerns about a computer knowing a user too well that are often raised by recommender systems.

The simplest kind of task-based personalization—simply promoting like products together—does not require the development of a personalization “system” and may not even be considered personalization. For some applications, this approach is often more cost effective than developing a system that attempts to infer user preferences (Berk, 2003).

5.1.3 User involvement

- *User-initiated personalization.* Some sites offer users the option of selecting customizations such as stock tickers that display stocks of interest, weather forecasts for the user’s region, or news related to topics the user has selected. Users might also select their preferred page layout or the number of items they want to see displayed, or they might provide information about their display and bandwidth constraints and ask to have a site optimized accordingly.
- *System-initiated* personalization*.* Some sites attempt to personalize content for every user, even if users do not request customized features and take no explicit actions to request personalization. In some cases, sites provide a way for users to opt-out of personalization.

User-initiated personalization tends to be more privacy-friendly than system-initiated personalization because users are more aware that personalization is taking place and can make a conscious decision about whether or not to activate it. System-initiated personalization can be made more privacy friendly through the use of notices and opportunities to disable the personalization.

5.1.4 Reliance on predictions

- *Prediction-based personalization.* Some sites use user’s explicit or inferred ratings to build user profiles that can be compared with the profiles of other users. When users with similar profiles are discovered, the system predicts that they will have similar preferences and offers recommendations to one user based on the stated preferences of the others. Such systems are often referred to as *recommender* systems or *collaborative filtering* systems. Thus, for example, if Jane and Sue provide similar ratings for 10 books, a recommender system might suggest to Jane two other books that she didn’t rate at all but had been rated highly by Sue. The suggested books may not necessarily be on the same topics as any of the books Jane rated herself.
- *Content-based personalization.* Some sites use the specific requests or other actions of a user to trigger automatic personalization. For example, if a user buys a book on Internet privacy, the site may suggest other books on Internet privacy. In this case the site is not using ratings to predict other types of books the user might like to buy, but simply offering the user additional books on the same topics as the book she already bought.

Content-based personalization tends to be more privacy-friendly than prediction-based personalization because it does not require that user profiles be stored. Prediction based personalization can be made more privacy friendly through the use of techniques to improve the privacy associated with user profiles.

5.1.5 Real world examples

Examples of personalization are readily apparent at many ecommerce web sites. For example, Riedl (2001) found 23 independent applications of personalization on the Amazon.com web site. As of December 2003, the Amazon.com[1] web site appears to use all of the types of personalization mentioned in this chapter, including features that fall on both ends of each of the four axes of personalization.

– *Data collection method.* Amazon allows users to provide explicit ratings for books and other products, which it uses to recommend other items to a user. It also uses information about past purchases and what items a user has looked at as implicit data with which to make recommendations. Users are directly in control of the explicit ratings they provide. In order to reduce privacy concerns and improve the usefulness of their recommendations, Amazon allows users to specify that some of the implicit data in their profiles should not be used when making recommendations. However, users cannot have this data removed from their profiles altogether.
– *Duration.* Amazon provides task-based personalization by creating a link to a page of items recently viewed by the user with suggestions for related items that might be of interest. Amazon also provides profile-based personalization by offering recommendations to the user based on her entire purchase and recommendation history.
– *User involvement.* Most of the Amazon personalization is done by the system automatically. However, users can edit their personalization settings and turn off some types of personalization or ask that certain items not be considered as part of their profile. A user can proactively rate items in order to have them considered as part of her profile. She can also request that payment information be stored to enable more convenient ordering.
– *Reliance on predictions.* Amazon makes predictive recommendations to users based on an analysis of a user's ratings and purchases compared with other users – including a "customers who bought this book also bought" feature. Amazon also provides users with lists of items in the same category as the item they requested.

[1] Throughout this chapter Amazon.com is cited as an example because it is a well-known web site on which ecommerce personalization can be observed in a variety of forms. The author has no affiliation with Amazon.com and no knowledge of the Amazon.com personalization systems beyond what can be inferred from reading material posted on the Amazon.com web site as of June 2003.

5.2 Pseudonymous Profiles

Often an individual's name and other personally identifiable information are not needed in order to provide personalized services. For example, recommender systems typically don't require any personal information in order to make recommendations. If personal information is not needed, personalization systems can be designed so that users are identified by pseudonyms rather than their real names. This reduces the chance that someone who gains unauthorized access to a user's profile will be able to link that profile with a particular individual, although it does not eliminate this risk. Someone who gains access to a user's account by using her computer or by learning her user name and password may be able to gain access to a pseudonymous profile. Furthermore, some combination of non-identifiable information contained in a pseudonymous profile may prove identifiable in practice, especially when combined with information stored in web usage logs (Malin, *et al.,* 2003). Nonetheless, pseudonymous profiles are a good way to address some privacy concerns. In addition, companies may find it significantly easier to comply with some privacy laws when they store only pseudonymous profiles rather than personally identifiable information.

For increased privacy protection, sites that employ pseudonymous profiles should make sure that this profile information is stored separately from web usage logs that contain IP addresses and any transaction records that might contain personally identifiable information. Web usage logs should be scrubbed so that they do not contain information that would allow pseudonymous profiles to be linked with other data.

Arlein *et al.* (2000) propose an architecture for pseudonymous personalization using information collected by multiple web sites. This system allows users to specify multiple personae that are stored on persona servers residing in the network. Users can grant web sites privileges to read or write to a specific persona. In addition, web sites can further restrict access to data they have written to a persona.

Kobsa and Schreck (2003) propose a more complex architecture for personalization services that use pseudonymous profiles. They envision the existence of user modeling servers that can communicate with users and personalization services via anonymous channels. While this architecture may prove too heavy for adoption by a single ecommerce web site, it is an interesting model that might be considered by a group of sites or as part of a single-sign-on/electronic wallet protocol.

5.3 Client-Side Profiles

Another option for reducing the privacy concerns associated with user profiles and satisfying some legal requirements is to store these profiles on the user's client (computer) rather than on a web server. This will ensure that the profiles are accessible only by the user and those who have access to her computer.

Client-side profiles may be stored in cookies that are replayed to a web site that uses them to provide a personalized service and immediately discards them. The information stored in these profiles should be encoded or encrypted so that it is not revealed in transit and it is inaccessible to other people who have access to a user's computer or to viruses or other malicious programs that may look for personal information stored in cookies.

A personalization interface that uses client-side scripting may be able to provide personal services by examining user profile information on the client without ever having to transmit it to the web site.

Canny (2002) proposes an architecture for a recommendation system in which participants compute a public "aggregate" of their data to share with members of their community. Individuals can then compute their own personal recommendations without revealing their individual data. He suggests that such an approach might be particularly useful in a ubiquitous computing setting where users may desire recommendations about everyday activities but are concerned that detailed information about their own activities not be revealed.

5.4 Putting Users in Control

Regardless of the approach taken to personalization, implementers who want to be sensitive about privacy concerns and comply with the fair information practice principles need to develop systems that give users the ability to control the collection and use of their information. Users should be able to control what information is stored in their profile, the purposes for which it will be used, and the conditions (if any) under which it might be disclosed. They should also be able to control when and if personalization takes place. In some cases, such controls may be required by law.

Developing a user interface that allows users to control the information in their profiles is a complicated problem, especially if the interface provides controls that go beyond a very course level of granularity. Lau *et al.* (1999) explored interfaces for a software tool that allows a user to create privacy rules for sharing web browsing histories. They found interfaces that require users to set privacy rules individually for every object in the system were too tedious for users, and they recommended that interfaces be developed that allow users to specify privacy policies that apply automatically to objects as they are encountered. However, formulating a privacy rule is a complicated task, which may require a deeper understanding of privacy issues than many users have as well as the ability to anticipate future activities that hold particular privacy concerns for a user. Some of the lessons learned by developers of Platform for Privacy Preferences (P3P) user agents may prove useful in developing privacy interfaces for personalized ecommerce services (Cranor, 2002; Cranor *et al.,* 2003).

A number of ecommerce web sites give users access to their profiles; however, it is not clear that many users are aware of this, and reports from operators of some personalization systems indicate that users rarely take actions to proactively customize their online experiences (Manber and Robison, 2000). To update personalization profile information on Amazon.com, for example, requires users to proactively go to their personalized "Your Account Page" and select from several items in a “Recommendations” section near the bottom of the page. Here users can edit previous explicit ratings they have given, as well as review their transaction and rating history and request that certain items be excluded from consideration when Amazon makes recommendations to them in the future. This interface essentially requires users to make individual privacy decisions for every object in the system, which can be quite time consuming. In addition, as users make new purchases, they have to remember to update their settings.

An interface might be developed that could allow Amazon shoppers to specify general privacy policies that would apply automatically. Such policies might allow users to specify, for example, that certain categories of purchases never be used to make recommendations, or that purchases be excluded from their profiles after six months. Or perhaps a user might want to specify that purchases made using her business credit card should be considered in her recommendations but purchases made using her personal credit card should be excluded. These types of rules would be useful to a user who can anticipate in advance the types of purchases that she would not want to have influence her recommendations. However, it might prove difficult for most users to formulate these kinds of rules.

An alternative approach that would require less foresight on the part of users would be to allow them to specify privacy preferences as part of the transaction process. Thus, when a user enters her credit card number and shipping address, she would also be prompted to decide whether this transaction should be excluded from her profile. She might establish a default setting that would apply to all her purchases unless she indicated otherwise, or even specify general policies like the ones described above that could be overridden easily for a specific purchase. A system-wide default might be that items that users have indicated have been purchased as gifts are excluded from a user’s recommendation profile (indeed, Amazon appears to exclude gift purchases from recommendation profiles already). A user interface might even include a box that allows a user to claim a purchase is a gift (“I didn’t buy it for myself”) as a way of disassociating herself from that particular purchase—similar to the habit some people have of requesting advice “for a friend” in an attempt to protect their own privacy.

A more sophisticated approach might allow users to establish multiple personae that would each have their own personalization profile. Thus, a user could have a separate profile for personal and business purchases, and could have a profile for each individual for whom she buys gifts. Besides addressing some privacy concerns, such an approach would likely lead to better personalization because it could offer recommendations within the appropriate context. Of course, designers of such a system should consider potential privacy concerns of gift recipients.

The Amazon interface allows users to exclude purchases from their recommendations, but not to remove them from their profile altogether. Excluding purchases from recommendations addresses some privacy concerns, but leaves others unaddressed. While legal and liability issues may require that Amazon retain transaction histories for some amount of time, there should be some retention period after which these histories need not be retained if a user prefers. Furthermore, even within the retention period, Amazon might allow users to request that all or part of their transaction histories not be made available through the web.

When user interfaces are developed that allow users to control the use of their information, it is also essential that back end systems be put in place that can properly carry out each user's instructions. This is easiest when personalization profiles are used only for web site personalization; however, some companies also make use of this data for postal mail marketing or other purposes. When these companies have databases spread across many different computer systems, as users change their personal settings these changes must be propagated across multiple systems that may store data in different formats. Furthermore, policies and procedures need to be put in place to limit access to these databases and ensure that those who have access to this data respect each user's privacy settings.

6. CONCLUSIONS

This chapter has reviewed several privacy risks related to ecommerce personalization and discussed privacy principles, laws, and guidelines that may impact the design of personalization systems. While no simple universal formula exists for designing a privacy-protective ecommerce personalization system, there are a number of approaches that may be helpful depending on the functionality and design requirements of a particular system. In general personalization systems tend to be most privacy friendly when they are explicitly activated by users, make use of data explicitly provided by users, use data obtained only during the current session, and perform personalization based directly on the content of information the user provides. When data must be stored in profiles and retained beyond a single session, pseudonymous profiles and client-side profiles may enhance the privacy of a system. Pseudonymous profiles are a good approach when personalization information need not be tied to personally identifiable information. Client-side profiles are useful when personalization services can be performed on the client. Interfaces that put users in control of the collection and use of their data as well as the types of personalization provided can make most personalization systems more privacy friendly, although further work is needed to develop privacy interfaces that are both usable and provide flexible control.

I see two major challenges for the human computer interaction community in the area of ecommerce personalization and privacy. The first challenge is to develop usable interfaces that allow users to control the use of their personal data and make privacy-related choices in a meaningful way that does not interfere with their use of a web site. The second challenge is to design interfaces that allow users to find what they are looking for with minimal need for user profiles, especially profiles tied to personally identifiable information.

Web site designers should also keep in mind that in some cases the goals of ecommerce personalization can be achieved at a lower cost and with fewer privacy risks through good web site design that makes it easy for users to find what they want. Many personalization interfaces build user profiles so that they can try to infer or anticipate user needs and offer relevant suggestions; however, users often already know what they are looking for, and good navigation design can help them find it (Berk, 2003). Where personalization adds value to a site, careful attention to design can reduce the amount of personally identifiable information necessary to make personalization successful, minimizing privacy risks and increasing user acceptance and trust.

7. REFERENCES

Ackerman M.S., Cranor, L.F., and Reagle, J. (1999) *Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences.* In *Proceedings of EC'99* (Denver, CO), ACM Press, 1-8. http://doi.acm.org/10.1145/336992.336995

Adams, A. (1999) *The Implications of Users' Multimedia Privacy Perceptions on Communication and Information Privacy Policies*. In *Proceedings of Telecommunications Policy Research Conference*, (Washington, DC). http://www.cs.mdx.ac.uk/RIDL/aadams/TPRC%20final.PDF

Arlein, R.M., Jai, B., Jakobsson, M., Monrose, F., and Reiter, M.K. (2000) *Privacy-Preserving Global Customization.* In *Proceedings of EC'00*, (Minneapolis, MN, October 17-20), ACM Press, 176-184. http://doi.acm.org/10.1145/352871.352891

Berk, M. (2003) *Beyond the Personalization Myth: Cost-effective Alternatives to Influence Intent.* Jupiter Research Site Technologies and Operations, Volume 2.

Canny, J. (2002) *Collaborative Filtering with Privacy.* In *IEEE Symposium on Security and Privacy*, (Oakland, CA), 45-57. http://citeseer.nj.nec.com/canny02collaborative.html

Claypool, M., Brown, D., Le, P., and Waseda, M. (2001) *Inferring User Interest*. IEEE Internet Computing, (November/December): 32-39.

Cranor, L. (2002) *Web Privacy with P3P.* O'Reilly & Associates.

Cranor, L., Guduru, P., and Arjula, M. (2003) *User Interfaces for Privacy Agents.* Under review.

Culnan, M. J. and Milne, G. R. (2001) *The Culnan-Milne Survey on Consumers & Online Privacy Notices: Summary of Responses.* http://www.ftc.gov/bcp/workshops/glb/supporting/culnan-milne.pdf.

Cyber Dialogue. (2001) *Cyber Dialogue Survey Data Reveals Lost Revenue for Retailers Due to Widespread Consumer Privacy Concerns.* http://www.cyberdialogue.com/news/releases/2001/11-07-uco-retail.html.

Gandon, F.L. and Sadeh, N.M. (2003) *A Semantic e-Wallet to Reconcile Privacy and Context-awareness.* In *Proceedings of the Second International Semantic Web Conference (ISWC03).*

Herlocker, J.L. and Konstan, J.A. (2001) *Content-Independent Task-Focused Recommendation.* IEEE Internet Computing, (November/December): 40-47.

Karat, C. M., Brodie, C., Karat, J., Vergo, J., and Alpert, S.R. (2003) *Personalizing the user experience on ibm.com.* IBM Systems Journal 42(4): 686-701.

Kobsa, A. (2002) *Personalized Hypermedia and International Privacy.* Communications of the ACM, 45(5): 64-67.

Kobsa, A. and Schreck, J. (2003) *Privacy Through Pseudonymity in User-Adaptive Systems.* ACM Transactions on Internet Technology, 3(2):149-183.

Lau, T., Etzioni, O., and Weld, D. S. (1999) *Privacy Interfaces for Information Management.* Communications of the ACM, 42(10): 89-94.

Malin, B., Sweeney, L., and Newton, E. (2003) *Trail Re-identification: Learning Who You are From Where You Have Been.* Carnegie Mellong University, School of Computer Science, Data Privacy Laboratory Technical Report, LIDAP-WP12. Under review. http://privacy.cs.cmu.edu/people/sweeney/trails1.html

Manber, U., Patel, A., and Robison, J. (2000) *Experience with Personalization on Yahoo!* Communications of the ACM, 43(8): 35-39.

Network Advertising Initiative Principles. (2000) http://www.networkadvertising.org/aboutnai_principles.asp

Odlyzko, A. (2003) *Privacy, Economics, and Price Discrimination on the Internet.* In *Proceedings of the Fifth International Conference on Electronic Commerce* (ICEC'03), Pittsburgh, PA.

Organization for Economic Co-operation and Development. (1980) *Recommendation Of The Council Concerning Guide-Lines Governing The Protection Of Privacy And Transborder Flows Of Personal Data.* Adopted by the Council 23 September 1980. http://www.datenschutz-berlin.de/gesetze/internat/ben.htm

Patrick, A.S. and Kenny, S. (2003) *From Privacy Legislation to Interface Design: Implementing Information Privacy in Human-Computer Interaction.* In *Proceedings of Privacy Enhancing Technologies Workshop (PET2003),* Dresden, Germany. http://132.246.128.219/legint/pet-workshop-patrick-kenny.pdf

Personalization Consortium. (2000) *Survey Finds Few Consumers Unwilling to Provide Personal Information to Web Marketers in Exchange for Better Services.* http://www.personalization.org/surveypress.html

Personalization Consortium. (2001) *New Survey Shows Consumers Are More Likely to Purchase At Web Sites That Offer Personalization.* http://www.personalization.org/pr050901.html

Ramakrishnan, N., Keller, B.J., Grama, A.Y., and Karypis, G. (2001) *Privacy Risks in Recommender Systems.* IEEE Internet Computing, (November/December): 54-62.

Riedl, J. (2001) *Personalization and Privacy.* IEEE Internet Computing (November/December): 29-31.

Turow, J. (2003) *Americans & Online Privacy: The system is Broken.* A Report from the Annenberg Public Policy Center of the University of Pennsylvania. http://www.asc.upenn.edu/usr/jturow/internet-privacy-report/36-page-turow-version-9.pdf

Warrior, J., McHenry, E., and McGee, K. (2003) *They Know Where You Are.* IEEE Spectrum, (July) 20-25.

Zaslow, J. (2003) *If TiVo Thinks You Are Gay, Here's How to Set It Straight: What You Buy Affects Recommendations On Amazon.com, Too; Why the Cartoons?* The Wallstreet Journal, (26 November). http://online.wsj.com/article_email/0,,SB1038261936872356908,00.html