Supporting Privacy Management via Community Experience and Expertise

Jeremy Goecks, Elizabeth D. Mynatt
Georgia Institute of Technology, USA
{jeremy, mynatt}@cc.gatech.edu

Abstract. We propose a novel approach for supporting privacy management that leverages community experience and expertise via the process of social navigation. Social navigation simplifies the often complex task of managing privacy settings, and systems that employ social navigation can advantageously complement user privacy management processes. We implemented our approach to privacy management in the Acumen system; Acumen uses social navigation to enable individuals to manage their Internet cookies both manually and automatically based on the behavior of others in the community. We present the Acumen system in detail and discuss data obtained from a six-week, preliminary deployment of Acumen. Lastly, we discuss challenges that systems implementing our approach must address if they are to be successful.

Introduction

Privacy has come to the forefront of societal concerns about technology in recent years (Ackerman et al., 1999, FTC, 2000, Harris, 2003, Turow, 2003). The rise of the Internet and ubiquitous computing technologies has made it possible to easily collect, store, aggregate, and share personal information. Such technologies offer many benefits; however, there is significant concern that applications and entities using these technologies (e.g. companies, governments) are marginalizing people's desire for personal privacy. Much can be gained by addressing people's privacy concerns. These technologies and applications are more likely to be fully accepted by and integrated into society if privacy is addressed; in addition,

scholars argue that privacy is beneficial to both individuals and society (Westin, 1967, Lessig, 1999).

For the purposes of this paper, we define 'privacy' to be an individual's ability to control when and how he shares his personal information with other people and third parties; when an individual manages his privacy, then, he manages when and how he shares his personal information.

The HCI community has substantially advanced its understanding of privacy management over the past decade. We now understand that privacy management cannot be addressed solely or even largely by a static set of preferences that determine how a user's information can be shared. Rather, privacy management is a fluid, organic process in which users are constantly refining their choices based on any number of contextual facets (Bellotti, 1996, Bellotti and Sellen, 1993, Palen and Dourish, 2003). This is true both for privacy management among peers and also for privacy management between individuals and "third parties," such as corporations and government agencies.

Supporting privacy management, then, is a challenging task for a computational system. In this paper, we propose a novel approach for supporting privacy management that leverages community experience and expertise via social navigation. Social navigation is the process of using other people's behavior to inform one's own behavior (Dieberger et al., 2000). Systems that employ social navigation leverage users' data in aggregate as a form of information or advice to help individuals make decisions. We can consider the users of a social navigation system to be a community; thus, when a community member engages in social navigation, the system provides the member with community data, the aggregate activity data of all community members. Social navigation systems have proven to be successful in many domains (Hill et al., 1992, Resnick et al., 1994, Svensson et al., 2001, Wexelblat and Maes, 1999).

Our approach employs social navigation to support individual privacy management decisions and enable novel privacy management techniques. Social navigation is a promising approach to privacy management for several reasons. Social navigation simplifies the often complex task of managing privacy settings by leveraging people's tacit ability to infer information from others' decisions and use that information as form of advice. Studies have shown that, if advice is available when making a decision, users very often use the advice and make a better decision as a result of using the advice (Harvey and Fisher, 1997, Yaniv, 2004). In addition, social navigation systems offer a technological complement to user privacy management activities; both evolve as user behavior changes, and both are frequently collaborative and situated in other, principal activities.

We have implemented our approach to privacy management in a system that enables users to manage an important facet of their Internet privacy: Internet cookies. Websites use cookies most often to identify users, store preferences, and record users' browsing behavior. While users derive benefits from cookies, the

ability of websites' to use cookies to identify and monitor users' activities means that cookies also present a threat to users' privacy.

Internet users are becoming increasingly concerned about their privacy online and about cookies in particular (Ackerman et al., 1999, Harris, 2003, Turow, 2003). Managing cookies on an individual basis is impractical, and existing solutions for managing cookies, such as P3P user agents (Cranor, 2002) and web browsers' tools, are insufficient at times. These tools are not well understood by users, offer little awareness of ongoing cookie activity, and provide inflexible settings that do not adapt to changes in users' needs and attitudes (Friedman et al., 2002, Millett et al., 2001).

The Acumen system employs social navigation to help users manage their cookies. Acumen addresses many problems of existing cookie management solutions and also provides novel methods for managing cookies. Acumen's interface is an Internet Explorer toolbar (Figure 1) that enables users to manage their cookies by leveraging community data. To this end, Acumen collects information about how users are managing their cookies and employs this data for three purposes.

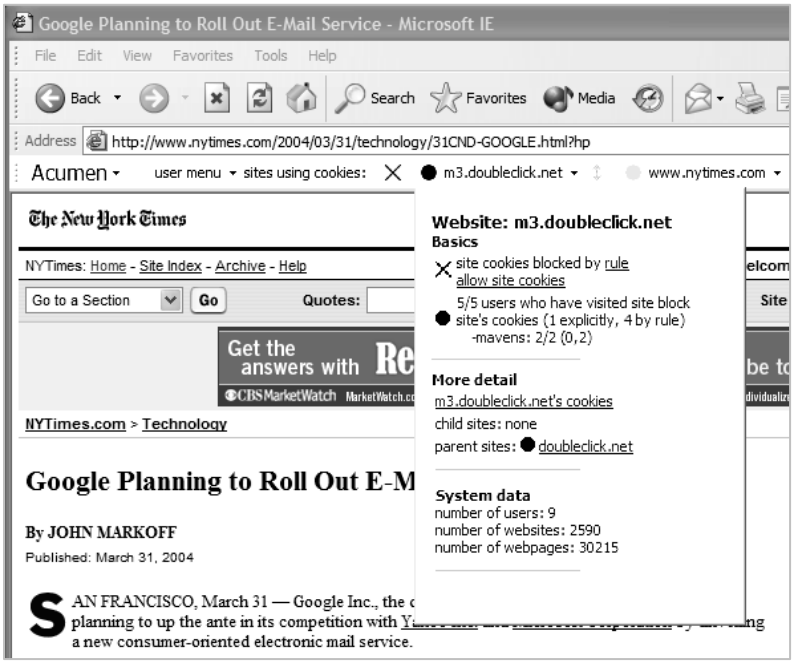


Figure 1. Acumen toolbar for a page on *The New York Times* website.

First, Acumen uses its data to raise users' awareness of cookies, especially those that others have blocked. This information is conveyed via color-coded icons in the toolbar; icons are colored using biased thresholds that accentuate user activity. Second, Acumen makes its data available to help individuals make more informed decisions about whether to block or allow a website's cookies. Third, Acumen enables users to automate cookie management by using simple rules which automatically block cookies that others have blocked.

To address herd behavior (Banerjee, 1992), a common problem in social navigation, Acumen employs data from all users and also from a subset of expert users called *mavens* (Gladwell, 2000). Mavens' data can help deter individuals from blindly following the decisions of others.

Overview

We offer three research contributions in this paper. First, we introduce a novel approach to privacy management that employs social navigation. Second, we demonstrate an application of this approach in Acumen, a system that enables users to manage their Internet cookies. Third, we evaluate Acumen using data from a six week, preliminary deployment of the system. We evaluate how well Acumen helps users manage their cookies and also discuss five challenges that we identified during our design and implementation of Acumen. These challenges are: (1) raising awareness of cookies; (2) understanding decision support; (3) obtaining data coverage; (4) mitigating herd behavior; and (5) addressing a privacy paradox.

Building on Related Work

Previously, we described how privacy management has come to be understood as a dynamic process. Our approach for supporting privacy management builds both on this work and on work in two other areas: (a) social navigation research and (b) research investigating management of Internet privacy and Internet cookies.

Social Navigation

The adage "a crowd draws a crowd" describes one instance of social navigation. When an individual sees a crowd outside an unfamiliar restaurant, she can infer that many people enjoy the food at the restaurant and that the restaurant generally serves good food. Hence, she is more likely to dine at the restaurant than she otherwise would be because the crowd provides information about how much other people like the restaurant. The crowd's behavior, then, is a simple form of data or advice that a bystander can use to make a decision.

In general, social navigation is the activity of using other people's behavior to inform one's own behavior. Social navigation is quite common in everyday life; it has also been demonstrated to be a powerful concept in digital systems (Dieberger et al., 2000). Researchers have built systems that enable users to perform social navigation in numerous domains; these domains include editing and reading documents (Hill et al., 1992), reading newsgroup messages (Resnick et al., 1994), exploring an online food and recipes store (Svensson et al., 2001), and browsing the Internet (Wexelblat and Maes, 1999).

We believe that our approach is the first attempt to explore social navigation as a privacy management solution. Social navigation is a particularly promising approach to privacy management for several reasons. Social navigation is a tacit and natural facet of the decision-making process because people are social beings. People routinely make inferences based on others' behavior and use this information as a form of advice; this advice simplifies and informs what can otherwise be complex decisions (e.g. how fast should I be driving on an unfamiliar road?). Finally, there is substantial evidence that people very often use advice and make better decisions as a result (Harvey and Fisher, 1997, Yaniv, 2004). Social navigation, then, can simplify and improve the often complex decisions that people must make when managing their privacy.

Social navigation systems also offer a technological complement to user privacy management processes. Users' privacy settings evolve to reflect changing needs; changes in how personal information is collected and used or in community norms surrounding use of personal information prompt changes in users' privacy settings (Palen and Dourish, 2003). Social navigation systems evolve as well because user's activities shape the system; thus, the system evolves as users' activities change.

Privacy management is frequently a collaborative process; conventions regarding privacy management develop within communities, and an individual's privacy management decisions are made in the context of these conventions (Bellotti, 1996). Social navigation systems support a similar process. Community conventions are made visible by aggregating community members' data, and an individual's decisions are made in the context of, and often directly using, this aggregated data.

Internet Privacy and Cookies

Internet users have numerous privacy concerns. One of their main concerns is the collection of personal data by third parties; users want the ability to control when, how, and what information they share with third parties. Internet cookies (RFC, 2004) are particularly troublesome in this respect because websites can use cookies to collect and store information about users; sites often use cookies to monitor users' browsing activities. In fact, at least thirty-five percent of websites use cookies to collect such information (FTC, 2000). Managing cookies on an individual or per-request basis is confusing, tedious, and overly invasive for most users. Hence, there is a need for tools that enable users to better manage cookies.

Some online privacy policies describe how a website uses cookies and what data they collect using them. However, online privacy policies are often difficult to locate and understand (Jensen and Potts, 2004). The Platform for Privacy Preferences (P3P) specification enables websites to encode a privacy policy in a

machine-readable format; software agents can then interpret and utilize P3P policies (Cranor, 2002).

Much work has been done in an attempt to help users manage their cookies. Both of today's major browsers, Internet Explorer and Mozilla, provide users with the ability to filter out cookies with particular characteristics. Users can block cookies without acceptable privacy policies or cookies from particular websites. However, there are problems and inadequacies with both browsers' cookie management tools. For example, cookie settings are often nested deep inside many menu levels, making them hard to find and modify; also, browsers provide little on-going awareness of cookies and do not enable users to adapt their settings to changes in their needs. Finally, many users do not understand the terminology that is used by browsers (Friedman et al., 2002, Millett et al., 2001). Neither browser uses social navigation to support cookie management.

There have been efforts to develop tools that extend browsers' cookie management capabilities. Cookie Watcher is an awareness tool that displays cookie activity in real-time but does not support cookie management (Friedman et al., 2002). The Privoxy web proxy blocks cookies from websites known to track users' activities via cookies for targeted advertising purposes (Privoxy, 2004). Privoxy provides little awareness to users about cookies; moreover, Privoxy's list of blocked websites is static, and thus it does not support dynamic or flexible cookie management.

Finally, two systems that employ explicit user voting to help users manage other facets of Internet privacy are Cloudmark's SpamNet (Cloudmark, 2004) and the Social Contract Core (SCC) (Kaufman et al., 2002). SpamNet utilizes users' votes to identify and filter spam, and the SCC uses votes to help companies improve their privacy policy or develop separate policies for different groups. Our approach builds on ideas in these systems. While they indirectly support privacy management through user voting, our approach directly leverages community activity data to help individuals manage their privacy.

Acumen

The ACUMEN system (Figure 2) employs a community's activity data to help individuals manage their Internet cookies. Acumen's community is the individuals who use Acumen and thus contribute data to the system via their cookie management activities. Individuals manage cookies at the website level, allowing or blocking cookies from websites. Acumen allows all cookies by default. Acumen's community data consists of the number of individuals who have "visited" a website (i.e. requested a file from the site), the number of such individuals who allow the site's cookies, and the number of individuals who block the site's cookies.

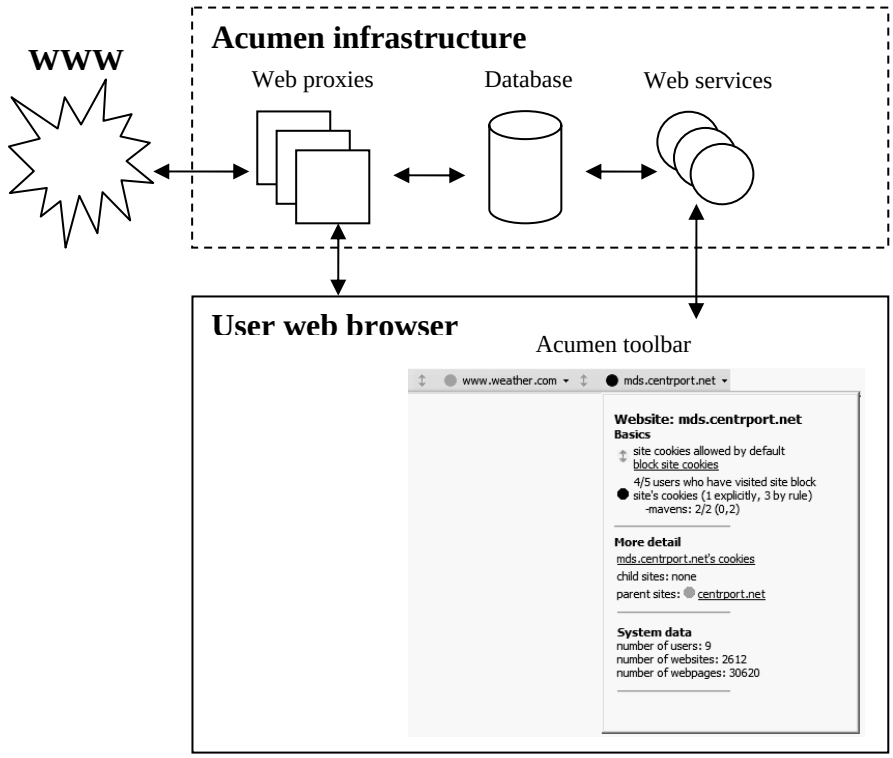


Figure 2. Acumen architecture.

Acumen enables community members to manage their cookies via indirect collaboration through their actions; each individual benefits by leveraging the community's collective knowledge and experiences via social navigation. Acumen enables individuals to leverage its community data in three ways. During their normal web browsing activities, individuals can maintain awareness of the websites using cookies on the webpages that they are visiting and whether other community members generally allow or block cookies from these sites. When making the decision to allow or block a website's cookies, individuals can view the community data for the site in detail and use this information to inform their decision.

Individuals can also employ simple rules that leverage community data to automatically block cookies. Individuals can create rules of the form 'If X% of users have blocked cookies from a website, then automatically block the site's cookies.' Individuals choose a rule's threshold percentage when they create it. The percentage of individuals that block a site's cookies includes both individuals that block cookies explicitly and those that block cookies by rule. Thus, when one individual's rule blocks a site's cookies, it can cause another individual's rule with a higher threshold to block the site's cookies as well, and so on. This chain of rule applications acts like a social epidemic (Gladwell, 2000), and it can propagate the blocking of a site's cookies quite quickly among Acumen's community.

Acumen utilizes community data from all individuals and also from a select subset of individuals called *mavens*. Gladwell defines a maven as a domain expert, someone who has both a deep understanding of a domain and also an intrinsic desire to learn as much as they can about the domain; mavens have been identified in many areas (Gladwell, 2000). Internet privacy mavens almost certainly exist as well; people that read and contribute to the Electronic Privacy Information Center (EPIC) website (www.epic.org) and those that use free cookie management software such as Privoxy are likely mavens. We believe that Acumen is the first computational system that attempts to utilize the concept of mavens.

Acumen leverages mavens' expertise by anonymously identifying and providing data from them. To identify mavens, Acumen computes a 'maven rating' for each individual; an individual's rating is the sum of the square roots of a individual's actions across all the websites that he visited:

$$R_m = \sum_{websites} \sqrt{num\_user\_actions_n}$$

Each time an individual explicitly blocks or allows a site's cookies, Acumen increments the individual's action count for that site.

This function has two interesting features. First, taking the square root of the number of actions decreases the influence of each additional action on a individual's maven rating; for example, the first 4 actions an individual performs on a site will increase his rating by 2, but the subsequent 4 actions that he takes on the site will increase his rating by only 0.82 This feature reflects the fact that people often learn more in early experiences than they do in latter experiences (Fridland et al., 2003); thus, inexperienced individuals' ratings increase more quickly with additional actions than do experienced individuals' ratings.

The second interesting feature of the function concerns the placement of the square root operator. By taking the square root of actions performed for each site rather than the square root of actions performed across all sites, the function balances breadth and depth of individual actions, though breadth is slightly favored.

Acumen labels the individuals with the top 20% of ratings as mavens. It is not clear what percentage of individuals should be labeled as mavens; we are not aware of any estimates about how many mavens are present in a typical domain.

Finally, it is worthwhile to note that we designed Acumen to support privacy management among a small community. Acumen's data is most useful when individuals can effectively infer information from the cookie management activities of Acumen's community, and it is easiest to make effective inferences when Acumen's community members shares norms and practices. Small communities, such as extended workgroups and organizations, often do share norms and practices; thus, members of these communities are very likely to be able to use Acumen's data effectively.

Implementation

Four component types comprise the Acumen system: (1) remote web proxies; (2) a central database; (3) Acumen's toolbar; and (4) web services that act as data intermediaries between the database and the toolbar. The components communicate securely using Secure Sockets Layer (SSL) channels.

All users' web traffic goes through one of Acumen's web proxies. A proxy performs two actions: (1) records the websites that a user has visited and the cookies used by those sites; and (2) blocks cookies from webpage requests and response if a user has explicitly or by rule blocked the site's cookies.

Acumen's database acts as a central repository for all data used by the system. For each website that a user has visited, the database maintains a user-website history; the history contains the cookies that the site uses for the user, whether a user blocks or allows the site's cookies, and, if cookies are blocked, how so.

Acumen's web service acts as the intermediary between its database and its toolbar. When a user visits a webpage, the toolbar obtains data about the page from the service. The service also handles user actions (e.g. blocking a cookie, changing a rule) and updates the database accordingly.

Acumen's web proxies and web services are the performance bottlenecks in the architecture. In order to make these components as responsive as possible, Acumen supports dynamic replication and deployment of the components and uses caches in both components.

Acumen attempts to mitigate privacy concerns by ensuring that user data is anonymous both at the user interface level and at the system level. Acumen's interface enables users to view only aggregated data; users are never able to view another individual's data. Acumen does not record any identifying information about its users beyond a persistent identifier, and it records a user's browsing activities only if she is logged into the system. Finally, Acumen provides a simple interface for users to see what data the system has collected about them; this interface enables a user to exclude some or all of her data from Acumen's community data. Acumen's architecture ensures that the system's complexity is hidden from users. To use Acumen, a user needs only install Acumen's toolbar, set her browser to use Acumen's proxy, and create a pseudonym for persistent identification by Acumen.

Internet Explorer Toolbar

Acumen's toolbar uses a just-in-time approach to provide cookie information. The toolbar lists the websites that are using cookies on the page that a user is currently visiting. Next to each website using cookies are two icons. The icon to the far left of the site name denotes whether the user allows or blocks the site's cookies; a green double arrow indicates that they are allowed, and a red X indicates they are blocked.

The circle icon to the immediate left of the website name denotes Acumen's community data for the site. Recall that Acumen's community data is the number of users who have blocked/allowed a website's cookies. The icon itself has two regions: an inner region and an outer region. The inner region's color denotes data for mavens; the outer region denotes data for the entire community (Figure 3). Regions are colored using a stoplight motif. Green indicates that a great majority of users (90% or more) allow the site's cookies, yellow indicates that most users (75% to 90%) allow the site's cookies, and red indicates that only some users (less than 75%) allow the site's cookies.

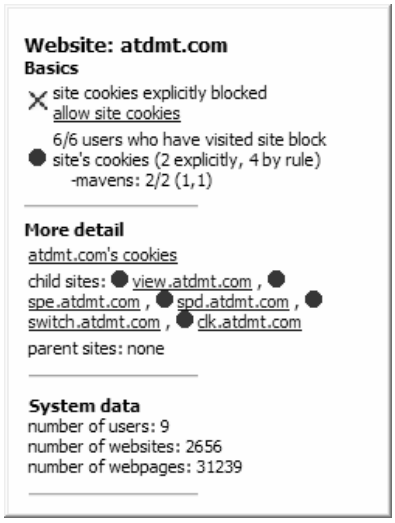
The icons serve to alert users to potentially problematic cookies. A user can quickly glance at the toolbar and determine whether there are cookies on a page that others in the community have blocked. A user can also glance at the two icons next to a website name and determine whether her decision about the site's cookies matches others' decisions. A user's decision matches that of others if the icons are the same color; if not, the icons have different colors (Figure 3).

We use non-linear, biased thresholds for the color categories to reflect the often sensitive and conservative nature of privacy management. Even if only a few community members have blocked a site's cookies, this information is reflected in an icon's color and thus communicated to the user. Also, users rarely change default settings (Mackay, 1990); biased thresholds accentuate any deviation in the community data from Acumen's default of allowing cookies.

Clicking on a website's name in the toolbar opens the site's menu (Figure 4). The menu's top section elaborates on the icons next to the site in the toolbar. If cookies are blocked, the menu indicates why; a link is provided to block/allow the site's cookies. Community data is provided numerically, and the number of users who block the site's cookies explicitly and by rule are indicated.

The menu's middle section provides a link to view the cookies that a site uses

Figure 3. User allows cookies; mavens do not.



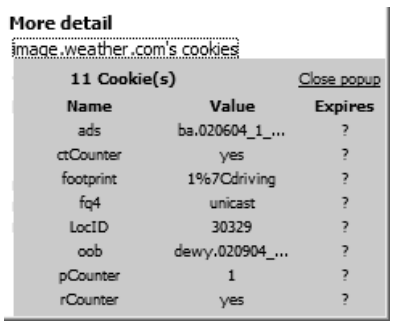


Figure 5. A website's cookies.

Figure 4. Website menu.

Figures 3-5 (ordered clockwise). Screenshots of Acumen's toolbar interface.

(Figure 5) and provides links to menus for child and parent sites. There is an icon next to each child and parent site that denotes the community data for that site. These links enable users to explore Acumen's data via relationships between parent and child sites. For example, a user can easily move to a top-level website (e.g. atdmt.com), view community data for the site, and block all cookies from that site; blocking cookies from a parent site blocks cookies from all its child sites as well.

The menu's lower section provides system information; the number of users, number of websites, and number of documents in Acumen's database are displayed. This information is intended to serve as an incentive for users to utilize Acumen; users are more likely to use Acumen if they know that there is data in the system.

The toolbar's user menu enables users to manage their rules for automatically managing cookies and indicates whether the user is a maven.

Deployment and Evaluation

We deployed Acumen to 9 users for 6 weeks so that we could begin to understand how users would employ Acumen. At the end of the six weeks, Acumen's database contained data for over 2650 websites; users had blocked cookies from 85 websites using Acumen.

All users in our deployment utilize the Internet heavily and as part of their job. Seven users were graduate students; two users were information workers outside the technology industry. Only two users managed their cookies before using Acumen; both were graduate students. The seven other users were familiar with cookies and expressed some concern about them but did not manage them or know their browser's cookie settings. Users employed Acumen on a voluntary basis. We asked users to employ Acumen in the context of their normal browsing activities; we did not ask them to be more proactive in managing cookies than they otherwise would be, though the presence of Acumen's toolbar likely encouraged them to manage cookies more than they would have otherwise.

While the number of users that participated in this deployment is small, it is still constructive to evaluate data obtained from the deployment. Acumen is a novel privacy management system, and this deployment provides needed data that offers insight into how users did and did not use Acumen's features. Evaluation data from this deployment will also inform future iterations of Acumen and stimulate research questions that can be explored in follow up work.

We obtained data about this deployment from informal interviews with users, logging data, and data from Acumen's database. Overall, the data is promising. Users employed many of Acumen's features to manage their cookies, and the data suggests they managed cookies actively and effectively.

We present data from our deployment in the next two sections. In the following section, we address an important but challenging question: are there cookies that users generally agree are "good" and other cookies that users consider to be "bad?" Addressing this question enables us to begin evaluating whether Acumen helped users made good decisions. Then, using data from our deployment, we discuss challenges that we encountered when designing Acumen.

Allowing the "Good," Blocking the "Bad"

One important evaluation criterion for Acumen is the degree to which its community data helps users make high-quality decisions. The question, then, is whether Acumen helps users allow desirable ("good") cookies and block undesirable ("bad") cookies. Before we can evaluate Acumen using this criterion, though, a method or model is needed to label cookies as good or bad; to the best of our knowledge, there is no such model. We first introduce our model for labeling cookies and then apply it to evaluate user decisions in Acumen.

A Model for Labelling Cookies

While there are individual differences in privacy preferences (Ackerman et al., 1999), there is likely to be some agreement among people about which cookies are good and which are bad. The primary obstacle that prevents people from labeling cookies is that they have difficultly operationalizing their privacy preferences. In other words, it is difficult for people to understand how cookies and features of cookies impact their privacy, and thus it is difficult for them to manage their cookies. For example, people may not understand how persistent identification changes when a website uses a long-lived cookie instead of a session cookie or how "third-party" cookies impact their privacy differently than "first-party" cookies.

Compact P3P policies (Cranor, 2002), which are sent by websites in conjunction with cookies, have begun to address these difficulties; however, many technological issues still impede people's ability to determine how cookies impact their privacy. Operationalizing privacy preferences in cookie management is also difficult because feedback is often lacking during management; many websites do not enable people to see how allowing or blocking cookies affects the information that websites collect about them.

We have taken these difficulties into account in our model. We use a simplified objective model, based on one observable and meaningful cookie attribute, to label cookies as good or bad. This attribute is the website from which a cookie originated; a cookie's originating website is available in all cookie management interfaces, including Acumen. By using this attribute to label

cookies, we ensure that people can operationalize their privacy preferences within our model and make an educated decision about how to label a cookie.

Our model fuses results from multiple Internet privacy studies (Ackerman et al., 1999, FTC, 2000, Harris, 2003, Turow, 2003). A cookie's host website is remarkably useful information for people. Knowing the host website enables an individual to (a) associate some degree of trust/mistrust in the cookie and (b) infer a rudimentary benefit/cost ratio for using the website's cookies.

Interpreting these two facets as continuously-valued attributes of cookies yields a two-dimensional, four-quadrant model for labeling cookies (Figure 6). The dimensions are (1) the degree of trust in a website/cookie and (2) the benefit/cost ratio of using a website/cookie.

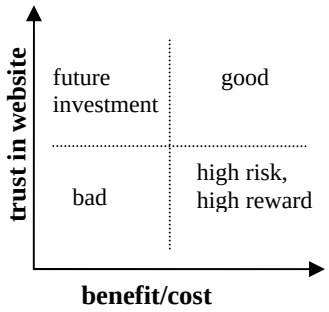


Figure 6. Cookie label model.

The model's dimensions aggregate multiple factors that influence people's privacy preferences. The dimension 'trust in a website' includes a site's reputation, business practices, privacy policies, and data sharing (and selling) policies. The dimension 'benefit/cost ratio' is simpler; people compare the benefits of using a site to its privacy costs. Privacy costs include data necessarily collected by the site so that it can provide benefits, data that is unnecessary but still collected by the site, and uses of collected data (e.g. unwanted email). While people are rarely aware of all data that a website has about them, they often have some knowledge of the data that they have given to or has been collected by a site.

This model is not inclusive or complete, but it is a first attempt to address how people make cookie management decisions given limited knowledge of how cookies and cookies features impact their privacy.

Based on this model, we label cookies that offer a high benefit/cost ratio and are from high trust websites as good; bad cookies are those with a low benefit/cost ratio and are from low trust websites. It is more difficult to label cookies in the other two quadrants. We term cookies that provide a high benefit/cost ratio but are from low trust websites as 'high risk, high reward.' We designate cookies that offer a low benefit/cost ratio but are from trusted websites as 'future investments'; such cookies require that users trust the site to provide

benefits in the future given that they are using cookies to collect personal information now.

Evaluating User Decisions

Using this model, we categorized two sets of website cookies: (1) the 85 websites whose cookies were blocked by at least one user; and (2) the 85 most popular websites whose cookies were not blocked by any users. We categorized the websites into one of the four quadrants using information from privacy advocate groups such as Privoxy, EPIC, and the Center for Democracy and Technology (www.cdt.org). Table 1 summarizes these categorizations.

	Allowed cookies	**Blocked cookies**
Good	34% (29)	3% (3)
Bad	13% (11)	85% (72)
Future investment	28% (24)	2% (2)
High risk, reward	25% (21)	10% (8)

Table 1. Website categorization data.

The data is encouraging. Among websites whose cookies were blocked, most are categorized as bad and only a handful fall into other categories. In contrast, the distribution among websites whose cookies were allowed is broad. Together, good and bad cookies constitute nearly half of all allowed cookies, while future investments and high risk, high reward cookies comprise the other half.

Recall that our goal is to evaluate whether Acumen helps users block bad cookies and allow good cookies. Taken together, this data suggests that it does. Acumen's true positive and true negative rates are an encouraging 91% (29/32) and 87% (72/83), respectively. Acumen false positive rate is only 3%. An interesting result is Acumen's 13% false negative rate; this rate is the number of bad cookies that are allowed by all users. There is no immediate explanation for this rate. It may reflect the fact that users acted conservatively, only blocking cookies they could confidently conclude were bad; alternatively, users may not always be able to recognize bad cookies.

Finally, it is important to note that these results are in aggregate, and Acumen certainly served some users better than others. One prominent Internet privacy study has shown that there are at least nine major factors that influence Internet privacy management (Ackerman et al., 1999). Moreover, the study found that there are three principal classes of users: privacy fundamentalists, privacy pragmatists, and the marginally concerned; each class has demonstrably different privacy needs. Pragmatists, the largest group of users (55%), often make sophisticated privacy judgments that cannot be easily reduced to rules.

We found anecdotal evidence that some users did in fact disagree with our model. One user blocked a "good" website's cookies because she felt that there was no reason for the site to be using cookies. Another user experimented with

blocking a “future investment” site’s cookies to evaluate how it affected the site’s performance; when he found that blocking cookies did not notably degrade the site’s performance, he permanent blocked its cookies.

These anecdotes suggest that simply automating cookie management by implementing our model would be insufficient for some users. In particular, Acumen’s approach of helping users make informed decisions rather than automating decision making is likely especially useful for pragmatists. Further studies are needed to evaluate how Acumen does and does not support particular types of individual cookie management practices.

Challenges

In this section, we discuss five challenges that we encountered while designing Acumen. The first challenge is universal for privacy management systems, and nearly all privacy management systems must address it. The latter four challenges are unique to our approach, and systems that employ social navigation to support privacy management must address them.

Raising Awareness for an Unknown Problem

Users will not use a privacy management system if they do not perceive a threat to their privacy. Users are largely unaware that cookies can be used to identify them and record their browsing activity, and many Internet cookie management tools go unused as a result (Turow, 2003). Making users aware of the pervasiveness of cookies and the risks that they pose is important if Acumen is to be used.

Acumen raises awareness of cookies by “pushing” information about cookies to the user. Acumen’s interface is persistent; it displays all websites using cookies on a webpage and information that details how other community members are managing cookies from these sites. Acumen attempts to draw users’ attention to sites that others have blocked cookies from; thus, Acumen attempts to highlight sites that others have deemed a risk to their privacy.

Our data indicates Acumen does raise users’ awareness of cookies. Many users commented that they were surprised and somewhat concerned about particular websites that were using cookies. One user was quite surprised when a cookie from the website ‘escapefromatlanta.com’ appeared on a webpage he was visiting for the first time. This cookie concerned him as it indicated that the website “knew” that he lived in Atlanta even though he had provided no data to the site; thus, another site had likely shared his data with escapefromatlanta.com. After this incident, he became more proactive about managing his cookies.

We were surprised to find that simply presenting community data motivated some users to start managing their cookies. One user commented that since other users were managing their cookies, it was “probably something I should do as

well." It did not matter which cookies others were blocking; simply knowing that they were managing their cookies was motivation enough for this user.

Users did take advantage of Acumen's color coding to identify websites from which other users have blocked cookies. Users found Acumen's color scheme useful and not overly distracting. Some users wanted the ability to sort websites based on community data so that sites from which many users have blocked cookies appear first in the toolbar; Acumen currently displays sites in roughly the order that their cookies are found on the page.

Understanding Decision Support

It has been shown that, for simple decisions, people use advice to improve their decision making (Harvey and Fischer, 1997, Yaniv, 2004). Acumen, however, supports privacy management decisions, which are often complex. Understanding when and how Acumen supports individual cookie management is perhaps the most important evaluation criterion for Acumen. Using this criterion, we can ask three questions about Acumen: (1) Did users employ Acumen's community data? (2) If so, how did they employ it when making cookie management decisions? (3) Did users employ Acumen's rules to automate cookie management?

Consider the two initial questions. In our deployment, users, on average, explicitly blocked cookies from 10 websites and automatically blocked cookies from 7 websites. It is not surprising to note that the ratio of cookies explicitly blocked to cookies blocked via rules is higher for mavens than for other users.

Anecdotes obtained from interviews indicate that users employed Acumen's data for different purposes when manually managing cookies. One user stated that, when blocking a website's cookie, he felt like "it was a pat on the back" if others had blocked the site's cookies as well. Thus, this user found a measure of validation for his decision by looking at others' data. Another user said that when she considered making a decision that others disagreed with (e.g. blocking a site's cookies that others had not), she thought more carefully about the decision than she otherwise would have. This user, then, used the data to help her decide which decisions to consider more carefully. There were also users who engaged in herd behavior and blocked a site's cookies because others had.

In addition to these purposes, we expect that there are many more purposes for which users may employ Acumen's data, and more research is needed to identify and understand them. These anecdotes, taken together with our earlier analysis that indicates that users share some agreement of good and bad cookies, suggest that Acumen helped users during the process of decision making and helped them make effective decisions.

There is evidence that users found Acumen's rules for automating cookie management useful. Users created few new rules, yet all users knew about

Acumen's rules and understood how they worked. Acumen automatically created two rules when a user signed up for an account: (1) if 20% or more of all users have blocked cookies from a website, automatically block the site's cookies; and (2) if 10% or more of mavens have blocked cookies from a website, automatically block its cookies.

In our limited deployment, these rules were sufficient for all but the most advanced users. Due to the small number of total users in our deployment, though, there were only two mavens; the small number of mavens did not allow for the latter rule to be meaningful for users. In interviews, most users said that they appreciated the rules as a mechanism for blocking cookies. Users favored using rules to block cookies because they either didn't know which cookies to block or were simply too preoccupied with other tasks to manage their cookies.

Community Data and Coverage

A system that employs community data to support privacy management is most effective when it can provide data about many of the potential decisions a user may face. If the number of potential decisions is too large or many potential decisions are not explored by users (and thus there is no data for these decisions), the system's data is unlikely to provide sufficient coverage over potential decisions and the system is rendered ineffective.

We can evaluate Acumen's data coverage in terms of websites. There are millions of websites, and Acumen will not contain community data for every site. The question, then, is which websites will Acumen likely have data for? And will users notice or be significantly harmed by the data that Acumen is missing? For the purposes of this discussion, we assume that Acumen has community data for a website if two or more uses have visited the site. Thus, any user who visits the site can observe how at least one other person manages the site's cookies.

We can posit answers to the above questions by observing that Acumen's data is tied to user's browsing activities; users will manage only cookies that they encounter while browsing. Hence, Acumen's data coverage mirrors the coverage obtained by users' browsing activities. Traffic among websites on the Internet has been shown to obey a power law distribution (Huberman, 2001). A basic power law distribution for web traffic says that the Nth most popular website receives about 1/N as much traffic as the most popular website. It follows that traffic to the most popular sites is a very, very large proportion of total traffic. Acumen, then, should contain community data for the most popular websites and be exponentially less likely to have data for less popular sites.

Data from Acumen's deployment confirms this hypothesis. The website traffic through Acumen's proxy exhibits a power law distribution (Table 2); traffic to 1.4% of websites accounts for 20% of all traffic, and traffic to 10.1% of sites accounts for 60% of all traffic. Overall, nearly one quarter of websites in Acumen

have community data. Among the 20% of websites with the most traffic, however, nearly half have community data, and among the 60% of sites with the most traffic, one-third have data. Table 3 summarizes Acumen's coverage data.

% of Websites	% of Total Traffic
1.4%	20%
2.5%	40%
10.1%	60%
11.2%	80%

Table 2. Acumen's site traffic obeys power law.

% of Websites	% with Com. Data
20%	48%
40%	41%
60%	33%
80%	25%

Table 3. Acumen's attains substantial site coverage.

This data provides preliminary evidence that website coverage can be attained for much of the Internet and for nearly all of the Internet's popular sites. A related question is whether community data is more useful when making cookie management decisions for popular or obscure sites; we plan to explore this question in the future.

Mitigating Herd Behavior

One problem that social navigation systems sometimes experience is herd behavior (Banerjee, 1992). That is, users blindly follow the decisions or behavior of others because that data is available and assumed to be correct. This phenomenon could be especially problematic for Acumen because many users have little experience managing cookies. We attempted to address this problem in Acumen by identifying mavens and presenting community data from them. We hypothesized that mavens' data could offer a more informed–and thus more useful–set of information than could data from the general community.

It was difficult to determine whether mavens' data was useful to the users in our deployment. There were only two mavens during the deployment as the total number of users was small, and many websites did not have data from mavens. When sites did have mavens' data, it was only somewhat useful to users. Some users were skeptical that mavens were more knowledgeable than other users and thus relied on community data from all users rather than on mavens' data.

We speculate that mavens may need to provide credentials in order for users to trust that they are more knowledgeable than others. In addition, more work is needed to investigate how to effectively identify mavens.

Addressing a Privacy Paradox

A paradox arises when employing community data to support privacy management: it is possible to solve one privacy problem while creating another. By collecting and making users' data visible, as Acumen does, it is possible that users' privacy could be compromised.

Recall that Acumen ensures users are anonymous at both the interface level and at the system level. Even with these precautions, many users were mindful that Acumen was recording their activities. Rather than just log out, users sometimes chose to both log out and stop using Acumen's proxy so that they were confident Acumen was not recording their activities. It is unlikely that we can alleviate all users' privacy concerns as long as Acumen employs a central database; one solution is to develop a distributed architecture in which users' data is kept on their personal machine and shared anonymously.

However, there is a tension between the users' anonymity and the usefulness of community data. Community data becomes more useful as the level of anonymity decreases; knowing more about the data's source enables users to better evaluate and employ the data.

Future Work and Concluding Thoughts

We expect to iterate on Acumen's design and deploy Acumen for an extended period of time to an established community. We anticipate that a larger, extended study will enable us to address some of the questions posed in this paper about Acumen's usage and about the challenges discussed above. In addition, deploying Acumen to different communities would enable us to begin to understand how community attributes (e.g. size, purpose, attitudes) affect usage of Acumen.

We are interested in understanding and better supporting the large-scale dynamics that undergird Acumen. Identifying mavens is critical to the success of Acumen, and we plan to further explore methods to identify mavens. Mavens may be best identified by a combination of methods, such as through the recommendation of others, the demonstration of expertise, or via machine learning techniques (e.g. manually identify a few mavens and find others that are similar). Understanding how privacy management norms develop and the role that mavens play in the development process is also an area for future research.

Using social navigation to address privacy management in domains beyond Internet cookies is promising. Cookies are only one area of Internet privacy; other areas include voluntary submission of personal information and adware/spyware. We are interested in developing systems that use social navigation to support privacy management in these areas. We also intend to explore applications of our approach to privacy management problems in ubiquitous computing systems.

Acknowledgments

We thank the Everyday Computing Lab, The Privacy Place, and Sun Microsystems for their support of this research.

References

Ackerman, M., Cranor, L. and Reagle, J. (1999) Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. Proc. of 1999 ACM Conference on Electronic Commerce, p. 1-8.

Banerjee, A. (1992) A Simple Model of Herd Behavior. Quarterly Journal of Economics 107,3(1992), 797-818.

Bellotti, V. (1996) What You Don't Know Can Hurt You: Privacy in Collaborative Computing. Proc. of the 1996 HCI Conference on People and Computer, 241-261.

Bellotti, V. and Sellen, A. (1993) Design for Privacy in Ubiquitous Computing Environments. Proc. 1993 ECSCW, 77-92.

Cloudmark SpamNet. (2004) http://www.cloudmark.com/products/spamnet/

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

Dieberger, A., Dourish, P., Hook, K, Resnick, P, Wexelblat, A. (2000) Social Navigation: Techniques for Building more Usable Systems. Interactions 7(6), 36-45.

Federal Trade Commission Report. United States Government (2000) Privacy Online: Fair Information Practices in the Electronic Marketplace: A Federal Trade Commission Report to Congress, May 2000. http://www.ftc.gov/reports/privacy2000/privacy2000.pdf

Fridland, A.J., Reisberg, D., and Gleitman, H. Psychology. W.W. Norton & Company (2003), New York, NY.

Friedman, B., Howe, D., and Felten, E. (2002) Informed Consent in the Mozilla Browser: Implementing Value-Sensitive Design. Proc. of 35th HICSS (2002), Abstract p. 247; CD-ROM for full paper.

Gladwell, M. The Tipping Point: How Little Things Can Make a Big Difference. Back Bay Books (2000), New York, New York.

Harris Inc. Poll #17. (2003) Most People Are "Privacy Pragmatists" Who, While Concerned about Privacy, Will Sometimes Trade It Off for Other Benefits, http://www.harrisinteractive.com/harris_poll/index.asp?PID=365

Harvey, N. and Fischer, I. (1997) Taking Advice: Accepting Help, Improving Judgement, and Sharing Responsibility. Journal of Organizational Behavior and Human Decision Processes, 70(2), p. 117-133.

Hill, W., Hollan, J., Wroblewski, D., McCandless, T. (1992) Edit wear and read wear. Proc. 1992 CHI, 3-9.

Huberman, B. The Laws of the Web: Patterns in the Ecology of Information, MIT Press (2001), Cambridge, MA.

Jensen, C. and Potts, C. (2004) Privacy Policies as Decision-Making Tools: A Usability Evaluation of Online Privacy Notices. Proc. 2004 CHI, 471-478.

Kaufman, J., Edlund, S., Ford, D. and Powers, C. (2002) The Social Contract Core. Proc. of 11th World Wide Web Conference, 210-220.

Lessig, L. Code and other Laws of Cyberspace. Basic Books (1999), New York, NY.

Mackay, W.E. (1990) Users and Customizable Software: A Co-Adaptive Phenomenon. Dissertation, Sloan School of Management. Cambridge, MA, MIT (1990).

Millett, L, Friedman. B., and Felten, E. (2001) Cookies and Web Browser Design: Toward Realizing Informed Consent Online. Proc. 2001 CHI, 46-52.

Palen, L. and Dourish, P. Unpacking "Privacy" for a Networked World. Proc. 2003 CHI, 129-136.

Privoxy web proxy, 2004. http://www.privoxy.org

Request For Comments #2965: HTTP State Management Mechanism, The RFC Archive (2004) http://www.rfc-archive.org/getrfc.php?rfc=2965

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994) GroupLens: an open architecture for collaborative filtering of netnews. Proc. 1994 CSCW, 175-186.

Svensson, M., Höök, M., Laaksolahti, and J., Waern, (2001) A. Social navigation of food recipes. Proc. 2001 CHI, 341-348.

Turow, J. (2003) Americans and Online Privacy: The System is Broken. Online report; available: http://www.annenbergpublicpolicycenter.org/04_info_ society /2003_online_privacy_version_09.pdf

Westin, A, Privacy and Freedom. Atheneum Press (1967), New York, NY.

Wexelblat, A., Maes, P. (1999) Footprints: History-Rich Tools for Information Foraging. Proc. 1999 CHI, 270-277.

Yaniv, I. (2004) "Receiving other people's advice: Influence and benefit." Journal of Organizational Behavior and Human Decision Processes, 93(1), p. 1-13.