Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices

Carlos Jensen, Colin Potts
GVU Center, College of Computing
The Georgia Institute of Technology
Atlanta, GA 30332, USA
{carlosj, potts} @cc.gatech.edu
+1-404-894-5551

ABSTRACT
Studies have repeatedly shown that users are increasingly concerned about their privacy when they go online. In response to both public interest and regulatory pressures, privacy policies have become almost ubiquitous. An estimated 77% of websites now post a privacy policy. These policies differ greatly from site to site, and often address issues that are different from those that users care about. They are in most cases the users' only source of information.

This paper evaluates the usability of online privacy policies, as well as the practice of posting them. We analyze 64 current privacy policies, their accessibility, writing, content and evolution over time. We examine how well these policies meet user needs and how they can be improved. We determine that significant changes need to be made to current practice to meet regulatory and usability requirements.

Author Keywords
Privacy, WWW, e-commerce, Usability, Consent, Readability

ACM Classification Keywords
H5.2 [Information Interfaces and Presentation]: User Interfaces – Evaluation, Usability; H5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia – User Issues

INTRODUCTION

Studies have repeatedly shown that users are increasingly concerned about their privacy when they go online. In a 2001 survey, 70% of respondents said they worried about their online privacy [9]. In a separate study, 69% said that they were "*concerned about* [online] *privacy invasions and try to take action to prevent them from happening to* [them]" [5]. This concern may not be unfounded. According to a recent study (91%) of U.S. Web sites collect personal information and 90% collect personally identifying information [1].

In response to public interest and regulatory pressures, privacy policies have become almost ubiquitous. The Progress and Freedom Foundation recently surveyed a sample of highly visited websites and found that 77% of those websites posted a privacy policy [1]. Website privacy policies are meant to inform consumers about business and privacy practices and serve as a basis for decision making for consumers. Not only are privacy policies important for decision making, they are often the only source of information. Policies therefore present an important challenge in terms of HCI; how to convey a lot of complicated but critical information without overwhelming users.

We know there are several common problems with policies today, including a frequent mismatch between the issues companies wish to address in their policies, and what users want to know about business practices. Part of the reason for this, and why privacy policies differ greatly from site to site is a lack regulation or industry standards. This applies both in terms of the language used in the policies and the issues they address. This lack of standardization makes it difficult to compare and contrast policies, thereby decreasing their value to users.

This issue of standards and regulations is slowly changing as different industries have become more tightly regulated in terms of privacy (e.g. Healthcare through the Health Insurance Portability and Accountability Act of 1996 (HIPAA) [15], finance through the Gramm-Leach-Bliley Act of 1999 (GLBA) [14], and the Children's Online Privacy Protection Act of 1998 (COPPA) [13] for children).

Industry standards have also emerged in the form of privacy certification services, also known as "privacy seals." These are run either by independent companies or

by industry groups. By setting requirements for what policies must address to obtain certification, these services may foster better privacy policies by encouraging consistency. However, seals often say nothing about the practices specified in policies, only that a minimum amount of information has been provided and that the company does abide by their policy. A privacy seal therefore usually says nothing about whether a company's practices are in the best interest of users, but studies show users are prone to making that assumption [3].

Given that privacy policies are everywhere and are often the only source of information about a company's privacy practices, it is important to examine whether they meet the needs of users. In this paper we present a thorough analysis of the different aspects of policies which may affect their value to users. We present a survey of representative policies; analyze how they are posted, their content and other aspects. We compare these policies with a sample from a regulated industry (healthcare in the U.S.). When HIPAA came into effect in April, 2003 with much fanfare and controversy, one concern frequently voiced was that new requirements for privacy policies would make them more like legal contracts and less understandable to average consumers. Our sample includes policies from before HIPAA came into effect and after, allowing us to see if legislation has had an effect on the quality of policies.

We compare our findings to those of user surveys and other studies to draw guidelines for how to improve current practices. Privacy policies have been around for quite some time, and therefore have been studied before. Some studies have examined the readability of policies [8] while others have focused on the content of these policies [2]. While some of this work is usability related, little has been done on evaluating the "complete" privacy policy. Closer to home, there is a body of HCI literature on designing for privacy [12], mostly focusing on the problems associated with groupware and ubiquitous computing [4, 6, 10].

We will start by explaining our methodology, including sampling methods and evaluation methods. We then examine the accessibility aspects of privacy policies and the sites that post them. An examination of policy presentation and content follows. We then summarize and analyze the results of this study, indicating how we can improve the current practice.

METHODOLOGY

For this paper we studied two sets of websites, a set of high-traffic websites and a set of health-care websites. The first set was collected to give a sample relevant to a large number of users, which they are likely to encounter frequently. The second sample was chosen to examine the effect regulatory efforts have had on policies.

For the high-traffic sites we used the "comScore Media Metrix Top 50 U.S. Internet Property Ranking" for August 2003[1]. Of these 50 websites, three were conglomerate sites with no common policy, and were therefore excluded. For the healthcare related sites we chose to use the sample studied in an earlier study of the industry [2]. This allowed us to examine how the policies had evolved over the last two years (from July 2001 to September 2003), which spanned the period when HIPAA came into effect.

Twenty-two of the original, pre-HIPAA policies were available for analysis. It was not possible to obtain the current versions of all these sites. As of September 2003 two of the healthcare websites were no longer offering publicly available privacy policies, one was no longer online, and two companies in the sample had merged. In total, 64 current policies were studied (47 from the high-traffic sample, and 18 from the health-related study, with one policy, that of *iVillage*, appearing in both samples). The sites studied are listed in Table 2 (The high-traffic sample) and Table 3 (the health-care sample). Where appropriate, the high-traffic and healthcare samples were combined for analysis.

Some sites split their privacy policies into multiple pages. In these cases all pages were analyzed as one continuous page, with the number of pages noted. Some sites offered software with privacy policies of their own. In these cases, only the site policy was analyzed to keep the sites in the sample comparable.

To set readability benchmarks for the policies, we had to make demographic assumptions about the Internet user population. Data on education levels and Internet use were collected from the National Telecommunications and Information Administration's (NTIA) report of 2002 [11] on Internet use in the USA. Given that all the sampled policies were in English, and largely from U.S. companies we chose to exclude international users from our analysis. We recognize that the Internet is a global system with a large international user base, but privacy issues must be studied against a background of national or regional cultures and jurisdictional boundaries. It is important to keep in mind that a large percentage of sites are American. Therefore their privacy practices have a large global impact.

We also restricted our analysis to adult users over the age of 25. We excluded children because in the U.S. children are afforded special protection under the law. COPPA severely restricts companies from collecting information from children. We excluded adults younger than 25 because many of them are still enrolled in educational programs, and therefore present a moving target in terms of the analysis.

[1] http://www.comscore.com/press/release.asp?id=348

Finally, we will not be analyzing the content of the policies in-depth, but rather looking at certain key policy elements. For a more in-depth analysis of the content we refer the reader to [2].

POLICY EVALUATION

Policy Accessibility

Accessibility is key to usability. Unless policies are easily found and readily available to end users the quality of the policy doesn't really matter. When we talk about the accessibility of privacy policies we are really interested in two things: First, how easy is it for users to find the policy? This is a function of where the link to the policy is placed, and how visible it is to users. Second, how easy is it to get a complete picture of the policy? This is a function of how long and how many pages the policy is spread across.

We examined the combined samples to determine how easy it is to find the policies. Of the 64 sites offering a privacy policy, we found that 55 (86%) offer a link to it from the bottom of their homepage. Three sites (5%) offered it as a link in a left-hand menu, while two (3%) offered it as a link at the top of the page. Sixty of the sites (94%), including all the health-oriented sites, offered a direct link to their privacy policy using such mechanisms; the other four sites (6%) required users to go through an intermediate page to get to the privacy policy, typically an "about us" or "help" page.

Five of the 60 sites (8%) with a direct link to the policy obscured the link through formatting. This always involved removing the typical link-underlining, and was sometimes compounded by changing the font color so it would more easily blend in with the background. Sometimes sites also placed the obscured link in the middle of a natural language sentence. Sixteen of the 60 sites (27%) with direct links offered the link in a reduced font size compared to the rest of the text on the page.

When it comes to the organization of policies and how many pages they are spread across, we found that thirteen sites (22%) split their privacy policies over more than one page. Most of these sites split the policy into two or three pages, although two sites (3%) split their policies into eight pages. Multi-page policies always had a uniform structure: one main policy page, with links to pages containing additional details or definitions. The sites with eight-page policies used three levels; the intermediate second-level pages were used to obscure significant privacy vulnerabilities (disclosure of and opt-out of web-bugs and spy-ware being one example from the sample).

Policy Readability

The Internet is no longer the exclusive domain of researchers and universities; it is used by people from all walks of life. According to a recent survey, 53.9% of the U.S. population is now online, and 65.6% has access to a computer [11]. As more people go online, the populations' diversity increases to reflect that of the real world. For this reason we need to make sure that we are not creating a "digital literacy divide," which would allow vulnerable populations to be exploited.

This notion of defending vulnerable populations from exploitation through confusing or intimidating language has strong legal backing, since legally binding agreements require the informed consent of all parties. In many jurisdictions, contracts and policies used in the insurance and banking industries for example, must meet certain readability criteria so that parties to these agreements can be assumed to have given their informed consent. The GLBA is one such piece of legislation, which also extends into the online sphere. It requires than any U.S. financial organizations' "privacy notice […] be a clear, conspicuous, and accurate statement of the company's privacy practices" [14].

Legal requirement for readability such as the GLBA are frequently undercut by a lack of formal definitions as to what constitutes a clear statement, or how policies should be evaluated. Given the lack of a strong formal definition, we must make some assumptions as to what can reasonably be called a clear statement, and how policies are best evaluated on this point. The remainder of this section will define the target population for these policies, and what can reasonably be expected from them in terms of reading comprehension. We will then discuss how readability may be measured, and how these readability metrics can be compared to the populations reading skills.

Reading Comprehension & Education

What constitutes a clear notice depends on whether it is reasonable to expect the target audience to understand it. This depends on the reading and comprehension skills of the target audience. Reading and comprehension skills in turn are closely linked to educational attainment. We know from the 2000 U.S. Census that 15.5% of the population over the age of 25 has less than a high school education, and only 26.9% of the population has a bachelor's degree or higher [11].

Literacy and education are closely linked to income and, as computers and Internet access are still above the means of some, we can expect the online population to have a higher than average education and literacy rate. The average education[2] of the U.S. Internet population is 14.4 years of education (approximately the equivalent of an Associate degree or two years in college), whereas the figure for the U.S. population as a whole is 13.5 years. To reflect the user population, we have used the education-level statistics for U.S. Internet users rather than that of the general population (see Table 1).

[2] Average assumes following years: Less than high school: 11, high school: 12, some college: 14, college: 16, postgraduate: 17.

One should remember that while this is sound usability practice, it overestimates the readability of privacy policies. A legally sound assessment of informed consent to privacy policies would probably refer to the adult population as a whole. Even though adult U.S. Internet users are more educated than the average American, 28.3% of them have the equivalent of a high school education or less. As more Americans go online, the percentage of users with lower educational attainment, the most underrepresented group, will inevitable grow.

Table 1: Education Levels, U.S. Adult Population

Educational Level	General Population			Internet Population	
	# People (millions)	% of Total Population	% Online	# People (millions)	% of Online Population
Less Than High School	27.5	15.5	12.8	3.5	3.8
High School /GED	57.4	32.4	39.8	22.8	24.5
Some College/Associates	45.4	25.6	62.4	28.3	30.5
Bachelors Degree	30.6	17.7	80.8	24.7	26.6
Beyond Bachelors	16.3	9.2	83.7	13.6	14.6

Source: 2002 National Telecommunications and Information Administration report [11]

Measuring Readability

With some definitions and numbers for literacy levels we can examine whether privacy notices are clear and accessible. The most commonly used method for determining the complexity of a text is to use a standardized, statistical readability metric. This allows for an objective evaluation and simple comparison between notices.

The Flesch Reading Ease Score (FRES) [7] is a popular metric, suited for evaluating more complex texts and is used extensively to evaluate school texts and legal documents. The FRES rates texts on a 100-point scale, where higher scores signify simpler texts. This score is computed by looking at the average number of syllables per word, as well as the average sentence length (Figure 1). Longer words and sentences are more difficult to read, and therefore produce a lower FRES.

Figure 1: Flesch Formulas

Flesch Reading Ease Score (FRES):
206.835 - 84.6 * (syllables/words)
- 1.015 * (words/ sentences)

Flesch Grade Level (FGL):
(0.39 * words/sentences)
+ (11.8 * syllables/words) - 15.59

Domain specific terminology and jargon normally will make a text more difficult to understand to an outsider than what the FRES will indicate, but these factors tend to equal out over a random population sample. Though no metric is universally liked, the Flesch metrics have been in use for decades. Today the FRES is used extensively to, among other things; regulate the complexity of insurance policies in more than 16 states.

Table 2: Popular Sample

Site Name	Words	Flesch Score	Flesch Grade	Seal	Site Name	Words	Flesch Score	Flesch Grade	Seal	Site Name	Words	Flesch Score	Flesch Grade	Seal
AOL Time Warner	1101	34.2	14.87	Y	Classmates.com	3542	33.9	14.57	Y	Wal-Mart	2098	45.2	12.07	
MSN-Microsoft	6222	41.5	13.18	Y	Weather Channel	2510	32.5	14.84	Y	United Online, Inc	4403	29.7	14.04	
Yahoo! Sites	3651	37.9	12.49	Y	Overture	1641	31.0	14.20		News Corp. Online	2098	15.6	17.96	
EBay	5216	36.5	13.66	Y	eUniverse Network	1099	22.2	17.14		Travelocity	403	26.3	14.53	
Google Sites	657	45.7	11.68		Vivendi-Universal	1729	26.9	16.02		Gannett Sites	No common policy			
Terra Lycos	5522	34.7	13.96	Y	Verizon	2090	34.0	12.79	Y	Dell	2274	45.4	11.87	Y
About/Primedia	2173	35.0	13.94		EA Online	2984	31.4	14.84	Y	American Greetings	3693	40.0	12.85	Y
Amazon Sites	2427	37.8	14.67		Expedia Travel	4362	28.7	14.60	Y	Earthlink	1788	28.5	15.17	
Gator Network	1786	31.1	15.01		SBC	4693	35.2	12.97	Y	Hewlett Packard	3301	34.5	13.44	Y
Symantec	2215	38.6	12.99	Y	AT&T Properties	1946	28.7	15.54	Y	New York Times	3472	46.2	12.23	
Excite Network	3298	31.2	15.39	Y	Sony Online	3984	30.0	16.88		ORBITZ.COM	3308	40.2	13.34	
Viacom Online	No common policy				Monster Property	2752	34.6	14.82		McAfee.com Sites	2160	33.9	13.03	
InfoSpace Network	2033	34.2	13.76		iVillage.com:	3681	26.2	16.21		Adobe Sites	2417	30.8	15.17	
Walt Disney	3170	44.5	11.70	Y	Ask Jeeves	1256	34.6	14.25		Trip Network Inc.	No common policy			
CNET Networks	1723	36.0	13.26		Weatherbug.com	3461	29.4	15.20		Buy.com Sites	5773	39.6	13.38	
Real.com Network	4306	36.4	13.60	Y	Dealtime	868	43.7	12.68		NFL Internet Group	2708	33.7	14.27	
					Cox Enterprises	1755	22.7	17.40		Comcast	1158	35.9	15.48	
										Average	2806.3	34.2	14.21	40.4%
										Standard Dev	1345.4	6.5	1.50	

Sites listed in order of popularity according to the "comScore Media Metrix Top 50 U.S. Internet Property Ranking" for August 2003.

Table 3: Health-care sites

		July 2001				September 2003					
	Site Name	Words	Flesch Score	Grade	Seal	Words	Flesch Score	Grade	Seal	Diff words	Diff grade
Health Insurance	AETNA	806	39.4	14.20		802	37.3	14.14		-4	+0.24
	AFLAC	1930	30.4	14.98		2160	26.4	15.37		+230	+0.33
	BCBS	638	40.2	15.20		716	37.2	14.98		+78	+0.77
	CIGNA	875	45.2	10.70		1115	42.2	11.50		+240	+0.87
	EHealthInsurance	1546	23.1	15.35	Yes	2113	29.9	14.03	Yes	+567	-1.32
	Kaiser Permanente	689	32.0	14.11		4678	40.5	13.45		+3989	-0.66
	OnlineHealthPlan	1390	31.9	13.83	Yes	No publicly available Policy					
Online Drugstore	CornerDrugstore	1906	37.6	12.98	Yes	No publicly available Policy					
	DestinationRX	1925	38.7	13.20	Yes	1871	36.0	13.46	Yes	-54	+0.25
	Drugstore	1499	38.7	13.75	Yes	2139	37.8	14.12	Yes	+640	+0.37
	Eckerd	1340	35.5	14.02		6404	34.0	16.24		+5064	+2.22
	HealthAllies	1025	34.5	13.81	Yes	1414	29.3	14.94	Yes	+389	+1.12
	HealthCentral	1283	41.1	13.10		675	38.5	13.31		-608	+0.66
	IVillage	3382	28.9	15.89		3681	26.2	16.21		+299	+0.33
	PrescriptionOnline	753	33.8	12.69		No longer Online					
	PrescriptionsByMail	1082	39.9	12.90	Yes	706	36.8	12.65		-376	+0.33
Pharmaceutical	Bayer	760	40.9	13.10		953	41.4	13.60		+193	+0.63
	Glaxo	448	39.5	12.60		396	37.9	13.19		-52	+0.67
	Lilly (Eli)	507	40.4	13.60		1014	35.2	14.76		+507	+1.15
	Novartis (Ciba)	1340	39.7	13.50		1366	36.5	13.68		+26	+.022
	Pfizer	393	41.1	12.10		331	35.8	12.39		+38	+0.57
	Pharmacia	957	38.7	13.08		Now part of Pfizer					
	Average	1203.4	36.5	13.45	31.8%	1807.4	35.5	14.03	22.2%	+604	+0.58
	Standard Deviation	1216.3	5.1	1.16		1613.7	4.7	1.26			

The FRES was of course developed to measure the readability of printed material. When we read on a display, the process is somewhat different because of the affordances of technology. Web pages have hyperlinks, which may help make information more accessible, or easier to find for users. When it comes to policies, and especially policies which are not regulated on form and content, it is necessary for users to read the entire policy. Hyperlinks and keyword searches are not going to be efficient simply because you don't always know what it is you are looking for. For this reason, we are forced to revert back to the normal linear paper processes.

A number of tools calculate the FRES automatically, including Microsoft Word, which was used to evaluate the policies discussed herein. MS Word also calculates the FGL, but only up to the 12th grade; for more complicated texts we calculated these scores manually using the formula above. We performed these evaluations on both sets of policies (See Table 2 and 3). The rest of this analysis will use the FGL equivalents, not the FRES.

The FRES can also be converted into a grade level score. The Flesch Grade Level (FGL) determines the U.S. grade-school equivalency level of a text, and is also based on the average number of syllables and sentence length. By using the FGL we can easily compare a population's educational attainment to the complexity of a text.

Analysis

For the popular sample, our survey found the average FGL of 14.21 (SD=1.50) (See Table 2). For the healthcare sites the average FGL was 14.03 (SD=1.26) (see Table 3). Across both samples the average FGL was 14.15 (SD=1.43). These averages are lower than the average education level of Internet users (14.4), but higher than that of the general population (13.5). The most difficult policy across both samples had a FGL of 17.96, the equivalent of a postgraduate education. The most readable policy required just under a high school education (11.50).

Of the 64 policies examined, only four (6%) were accessible to the 28.3% of the Internet population with less than or equal to a high school education. Thirty-five policies (54%) were beyond the grasp of 56.6% of the Internet population, requiring the equivalent of more than fourteen years of education. Eight policies (13%) were beyond the grasp of 85.4% of the Internet population, requiring the equivalent of a postgraduate education. Overall, a large segment of the population can only reasonably be expected to understand a small fragment of the policies posted.

We discard the hypothesis that the health-care (HIPAA regulated) policies were more readable than those of the high-traffic sample (n=63, t=0.324, p=NS). In terms of evolution, the policies in the health-care sample did not show an improvement in readability from July 2001 to

September 2003 (n=39, t=-1.015, p=NS) despite the passing of special legislation. There was no significant difference in the length of the policies (n=39, t=-1.241, p=NS).

We also examined the relationship between the length of the policies and their complexity. Users are often put off by lengthy policies, but are these policies in fact any harder to read? There proved to be no linear correlation between the length of the policy (in words) and the FGL for the combined sample set (r=0.049).

Finally, we examined the effect privacy seals had on policies, as certifying institutions usually have a set of minimum requirements on content. In terms of readability there was no difference between the two groups in terms of FGL (n=65, t=-1.256, p=NS). The two groups did prove to be marginally different in terms of the length of the policies, with the certified group on average offering policies, which were 50% longer than the non-certified group (n=63, t=1.730, p=0.09).

Policy Content

Privacy policies contain a great deal of information, enough subject matter for a paper in its own right. We shall therefore focus on a single policy element that greatly affects the usability and validity of privacy policies, namely how policy changes are handled, and what burden this puts on the user. All privacy policies build on the assumption that visiting the site implies the user's consent to the site's policy, whether or not the user reads it. This is typified by statements such as "[Company name] *may change this statement from time to time*" and "*Your continued use of this site constitutes acceptance of these terms.*"

In the combined sample, eight of the 64 policies (13%) offered no mention of how changes to the policy would be conveyed to the user. Twelve policies (19%) offered to notify users on the policy page and through email, while 44 policies (69%) required users to check the policy page periodically.

Of the policies which required users to check for changes, sixteen (25%) posted no modification date. Four (12.5%) of the policies which did not specify a modification policy also offered no modification date. Overall, only 41 polices (64%) were dated. Thus, in many cases, the user's only way of assessing changes to a policy would be to re-read the policy regularly to see whether it had changed. Based on the dates posted, policies varied in freshness from a few days to three and a half years, with an average of thirteen months. Eight (20%) had been changed in the previous three months.

Of the sites that specified how changes to their policy would be communicated, only eleven (19%) promised to give prior notice when significant changes were made. Four of these did not specify how much advance notice would be given; six specified a 30-day warning period, while one site promised to give six months notice.

ANALYSIS

Notification

A privacy policy builds on the concepts of fair warning and implicit consent. If a company posts a policy in a public place (such as linked off the main page of its web site), it can assume that users have been warned, and that by the act of continuing to use the service they have agreed to its terms. Fair warning, a well-established legal principle, sets three requirements [16]:

- The warning should be readily available to affected parties
- Affected parties should be given a clear way to voice their concerns or questions; and
- The warning should be understandable to any reasonable person making a good faith effort.

If the three requirements are met, sites can assume consent.

In general, websites did poorly on notification for notifying users of changes to their policies. It would of course be easy to require users to read the policy before accessing a website, but this would likely have no positive effect. Users would probably find this to be an annoyance and click through without reading. Even though sites do not require users to read their policy before access, they do place the burden of monitoring changes on the user. Over two thirds of sites (69%) require users to monitor the site's privacy policy regularly for changes.

We found the average age and the enormous variability in ages of the dated polices (mean and standard deviation each being about one year) to be surprising. There are three potential explanations for the long-lived policies in the tail of this distribution. The first, taking the age of the policies and their accuracy at face value, is to assume that that the policy is indeed up to date, but the business has not altered the way it handles users' information since it was posted. Given the length and complexity of most of the policies, together with the volatility of modern marketing practices, we think this explanation is unlikely.

A second explanation is that some companies may post privacy policies as legal disclaimers. These are blanket statements authorizing the company to do whatever it wishes with the information. This is really a variation of the first explanation, but with the policy, irrespective of its complexity and length, essentially promising little and therefore seldom requiring revision. Based on a close reading of the policies, we have encountered some of these, but again they are not common.

We believe that the most plausible explanation is that many policies are posted as the product of a one-off privacy project, after which the perceived importance of user privacy dwindles within the company. This is a potentially dangerous situation, as the posted policy may quickly cease reflecting the company's practices. Not only is this damaging to users, who may be exposed to privacy violations that are apparently forbidden by the policy, it is also damaging to the companies who may face negative

publicity and legal actions. Re-examining the health-care policies in a year's time would test this hypothesis. HIPAA's going into force in April, 2003 was an exogenous stimulus that synchronized the internal privacy projects of many companies in a single industry. If many companies adopt the single-project mode of privacy management, we would expect the average age of the policies in this industry to increase.

As to the requirement of fair warning, it is general practice for sites to provide at least an email address for the webmaster. Whether this person is qualified or willing to answer questions about the privacy policy is unknown. All the HIPAA compliant sites included physical contact information as well. A more interesting question is whether providing contact information really matters, as online privacy policies are non-negotiable. The user is presented with a set of terms and conditions, and has no leverage, or voice to negotiate new terms.

Accessibility

The sites in our combined samples generally had accessible privacy policies. They tend to be found down at the bottom of the homepage, together with legal disclaimers and assorted pieces of information. While this is an unglamorous location, it is fairly consistent across sites, and users can use this consistency as a location cue. We did not do any usability testing to verify that users did or did not correctly anticipate where policies could be found, though it is a reasonable assumption that they would given the data.

Of some concern is the practice of splitting policies across multiple pages, especially when policies span more than two pages. While this practice may make policies less intimidating to users, it has the potential to confuse or obscure. This practice has great potential for hiding important facts from users, in a maze of links, as was seen in our sample.

Readability

For websites, privacy policies are a compelling practice; they require very little effort or expense. However, websites currently undermine the legal basis for this practice by posting policies that are too complicated. The fact that only 6% of policies are readable by the most vulnerable 28.3% of the population, and that 13% of policies were only readable by people with a post-graduate education goes well beyond a reasonable burden for informed consent.

DISCUSSION

We have presented an in-depth evaluation of the different usability aspects of privacy policies and the practice of posting them as public warnings or disclaimers. Overall we have to conclude that while policies seem to be pervasively available online, there are serious problems with their structure and content. Even if one assumes that companies sincerely follow practices that comply with their posted policies, the form, location and legal context of policies make them essentially unusable as decision-making aids for a user concerned about privacy.

Too much of a burden is put on the end-user by failing to provide adequate notification of changes, or presenting privacy policies in language the user can understand. Users must, if they are serious about protecting their privacy, check the privacy policy of every site they visit, and in most cases check it again every time they visit the site. Failure to do so may mean that the user has agreed to different conditions and practices not only for additional personal information that the user provides subsequently but even for information that has already been collected by the site. The longevity of most privacy policies is a disincentive to re-reading them, since it is very unlikely that the privacy policy of an average frequently-visited site will have changed from the last time the user visited it. However, failure to do so may mean that the user has agreed to different conditions and practices, not only for additional personal information that the user provides subsequently but also for information that has already been collected by the site.

Furthermore, the practice of assuming that access implies consent has serious flaws that bring the whole practice into question. In order to access and evaluate a site's privacy policy, the user must access at least two pages on the site: the home page and the page containing the privacy policy. This means that the terms of an implied-consent policy contain a "Catch-22" implication: The user must accept the policy before he or she may read it. All the policies we surveyed contained language to this effect. Most sensitive personal information web sites collect can only be disclosed by users through direct input. Such information may be even more sensitive when combined with less sensitive information, such as your surfing patterns after leaving a site. Users may think about entering information, but often don't think that they may be subsequently be tracked.

Though users are concerned about their privacy, and claim to take steps to protect themselves, it is unreasonable to assume that anyone goes to the lengths required by current practice. It is our experience that survey respondents tend to greatly over-report the frequency and likelihood with which they read privacy policies. From a small survey done in a university setting we found from log file analysis that for a standalone website requiring registration, virtually no-one read the policy. We saw a total of 55,158 sessions, out of which only 131 (0.24%) included a visit to the privacy policy. Comparable numbers are difficult to get for industry sites and may be higher, but are unlikely to differ by the two orders of magnitude that would be necessary for even a quarter of users to visit a privacy policy.

Many of the issues we have been discussing were in the minds of the designers of P3P (the Platform for Privacy Preferences[3]). P3P is a set of practices and a way to encode

[3] http://www.w3.org/P3P/

privacy policies in XML so that interpretation and checking can be automated. P3P specifies a "safe area" for policies so they may be pre-fetched and examined by users before accessing the site itself, thus avoiding the "Catch-22" paradox noted above. It also makes it easier to implement software agents that check policies on behalf of users, screening the mundane and drawing the attention of users to the important decisions they must make. P3P is in use today along-side regular privacy policies. However, it has yet to gain significant momentum, and its current implementations restrict the enforcement of user preferences largely to acceptability of technical mechanisms such as cookies, not the full set of information-use preferences and policies made possible by the standard.

It is clear that the HCI community has a significant contribution to make in improving current privacy awareness and management techniques, a contribution that goes beyond the usability and user-interface design of web-browsing and security-enhancement tools, and is concerned also with the management of attention and awareness by users about what personal information they are voluntarily disclosing over time, what information is being leaked by the technology they use, and how this information flow interacts with business practices of the companies that own the web sites they visit. Without significant usability improvements in this broader sense, users cannot effectively take charge of their own information and protection, regardless of their motivation.

ACKNOWLEDGMENTS

This work was supported by NSF ITR Grant #0113792. The authors thank Khai Truong and Gregory Abowd for generously sharing their data. We also wish to thank Annie I. Antón, Julia B. Earp, David L. Baumer, William Stufflebeam and Qingfeng He for discussions leading to the development of these ideas.

REFERENCES

1. Adkinson, W. F., Eisenach, J. A., and Lenard T. M. "Privacy Online: A Report on the Information Practices and Policies of Commercial Web Sites" Progress and Freedom Foundation, Washington DC. March 2002
2. Antón, A. I., Earp, J. B. and Reese, A. "Analyzing Web Site Privacy Requirements Using a Privacy Goal Taxonomy", IEEE Requirements Engineering Conference (RE'02), Essen, Germany, September 2002.
3. BBBOnLine. "Third-Party Assurance Boosts Online Purchasing: BBBOnLine Privacy, Reliability Seals Increase Consumer Confidence; Privacy Remains Public's Chief Concern (survey summary)". Arlington VA, October 17, 2001.
4. Bellotti, V. and Sellen. A. "Designing for Privacy in Ubiquitous Computing Environments". European Conference on Computer-Supported Cooperative Work, ECSCW '93, Milan, Italy., ACM Press. 1993
5. Culnan, M. J. and Milne, G. R. "The Culnan-Milne Survey on Consumers & Online Privacy Notices: Summary of Responses". Washington DC: FTC, December 2001.
6. Dourish, P. and Bellotti., V. "Awareness and Coordination in Shared Work Spaces." Computer-Supported Cooperative Work, CSCW'92, Toronto, Canada, ACM Press. 1992
7. Flesch, M. "The Art of Readable Writing", Macmillan Publishing, 1949
8. Hochhauser, M. "Lost in the Fine Print: Readability of Financial Privacy Notices." Privacy Rights Clearinghouse, July 2001.
9. Jupiter Research, "Security and Privacy Data." FTC Security Workshop, May 20, 2002
10. Langheinrich, M. "Privacy by Design: Principles of Privacy-Aware Ubiquitous Systems." Proceedings of Ubicomp 2001, Springer. 2001
11. National Telecommunications and Information Administration. "A Nation Online: How Americans Are Expanding Their Use of the Internet" Washington, D.C. February 2002
12. Palen, L. and Dourish, P. "Unpacking 'Privacy' for a networked world" Conference on Human Factors in Computing Systems, CHI'03, Ft. Lauderdale, FL. 2003
13. U.S. Children's Online Privacy Protection Act of 1998, Public Law No. 105-277, October 21, 1998.
14. U.S. Gramm-Leach-Bliley Financial Modernization Act of 1999, Public Law No. 106-102, November 1, 1999.
15. U.S. Health Insurance Portability and Accountability Act of 1996, Public Law No. 104-191, August 21, 1996.
16. U.S. Regulatory Fair Warning Act of 1999. H.R. 881 One Hundred Sixth Congress, June 29, 1999