School of Electrical Engineering and Computer Science
Oregon State University, 1148 Kelley Engineering Center, Corvallis, OR 97331
Phone 541-737-3617 | **Fax** 541-737-1300 | http://eecs.oregonstate.edu

November 16, 2007

Comments regarding the FTC Town Hall Meeting on "Ehavioral Advertising: Tracking, Targeting, and Technology" of November 1-2, 2007

By

Carlos Jensen, Ph.D.
Assistant Professor, School of EECS
Oregon State University

First I would like to thank the Federal Trade Commission for arranging and hosting this discussion. I would also like to thank the commission for the opportunity to participate in the town hall and for allowing me to file a written comment.

As an academic with a background in Computer Science and Usability, my focus is on determining whether current consent and information practices are usable, accessible, and appropriate. In my research I both monitor the data collection practices and trends online, and conduct usability experiments and analysis to see how these meet the information and usability needs of the US population.

While I had the opportunity to share some of my findings and concerns with the commission and those attending the town hall meeting, I wish to summarize what I see as the most important obstacles to a fair and level playing field in terms of market forces and to supporting user involvement and consent online.

Surveys indicate that consumers are deeply concerned about their privacy and online data-practices they are exposed to. We have also seen indications that consumers are willing to vote with their feet and wallets in order to protect their privacy. This has been recognized by the major online players, who recognize the critical role that earning and keeping public trust plays in the growth and wellbeing of the online marketplace. As such, we have seen serious actors adopt and embrace voluntary online privacy policies and self-regulation programs such as the BBB and TRUSTe seal/certification programs.

While these are significant advances from where we were 7 years ago when the FTC last organized a workshop on this issue, several factors conspire to undermine public trust in the online industry. The argument that consumers who value privacy and object to any given privacy practice will vote with their feet and wallets is currently being undermined by a lack of transparency and standardization.

The primary mechanism for consumers to determine the privacy practice of a website is through its privacy policies. Privacy policies are voluntarily posted, and not universally adopted. Privacy policies are often long, complex, and written as legal disclaimers rather than to address the concerns and information needs of consumers. Because there are no standards or legal requirements, consumers are often left comparing apples to oranges. Furthermore, the lack of transparency and oversight is very much in the consumers mind and helps undermine trust.

In order for the market to function, there needs to be a level playing field, and information and policies need to be readily available, relevant, directly comparable, binding, and enforceable. Voluntary and unregulated efforts serve to punish serious actors by putting them at a competitive disadvantage to unscrupulous actors who either choose not to adopt voluntary efforts, or do so in a misleading fashion. Serious industry actors should therefore embrace and support efforts which help them advertise and promote their policy choices and values.

This is an issue which has been addressed in other industries with success, and without hurting the market or consumer choice. One example is regulation of the insurance and banking industries, setting standards for the issues which have to be disclosed, the language used to disclose these, and the enforcement and redress mechanisms available to consumers.

Requiring policies to address certain minimum sets of information would make them much more meaningful, and make consumers more likely to consult them. Such an online "nutrition label" should include, in a standard layout and language clear and unambiguous information on opt-in/opt-out options and mechanisms, what information sites collect, how it is collected, how it is processed and combined, how it is used, shared, or sold, and to whom. Terms such as "trusted partners" and "general statistics" should be strongly discouraged. If it too much of a burden for companies to list and explain every use they have for data and every partner they share with, how is it any more reasonable to ask every consumer visiting said site to attempt to gather that information themselves?

Requiring industry to file quarterly or yearly statements detailing their privacy and security practices could also be productive ways to building public trust. While companies may have legitimate competitive reasons for not disclosing every partner or use they make of data, there should be accountability somewhere.

Certain disclosure practices should also be strongly discouraged or banned, especially those putting undue burden on the consumer. One example is an update policy requiring the consumer to constantly check the policy document to see if the policy has been updated since their last visit, or since the start of their interaction with the site. Simple mechanisms such as requiring sites to display prominent notice of a policy change if the site tracks when the user last visited the site are simple solutions which would have significant impact.

We will seek to do our part by providing industry, government, and consumers with more detailed and relevant information on what online actors are doing, and what risks consumers face. We hope industry and the government will identify common needs and concerns to move us forward. While many industry representatives at the town hall spoke out against government regulation, an equal, and overlapping set of industry representatives spoke up for the need for stronger and more formal standards and best practices. Whether initiated and proposed by industry or government, it is in the vital interest of consumers that such standards be agreed and implemented immediately.

Thank you for your time,

Carlos Jensen

Tracking Website Data-Collection and Privacy Practices with the iWatch Web Crawler

Carlos Jensen
School of EECS
Oregon State University
Corvallis, OR 97331, USA
+1-541-737-2555
cjensen@eecs.osrt.edu

Chandan Sarkar
School of EECS
Oregon State University
Corvallis, OR 97331, USA
sarkar@eecs.orst.edu

Christian Jensen
Department of Economics
Southern Methodist Uni.
Dallas, TX 75275, USA
christia@mail.smu.edu

Colin Potts
College of Computing
Georgia Institute of Tech.
Atlanta, GA 30332, USA
potts@cc.gatech.edu

ABSTRACT

In this paper we introduce the iWatch web crawler, a tool designed to catalogue and analyze online data practices and the use of privacy related indicators and technologies. Our goal in developing iWatch was to make possible a new type of analysis of trends, the impact of legislation on practices, and geographic and social differences online. In this paper we present preliminary findings from two sets of data collected 15 months apart and analyzed with this tool. Our combined samples included more than 240,000 pages from over 24,000 domains and 47 different countries. In addition to providing useful and needed data on the state of online data practices, we show that iWatch is a promising approach to the study of the web ecosystem.

Categories and Subject Descriptors

H.5.4 **[Hypertext/Hypermedia]**, Architecture, User issues
K.4.1 **[Public Policy issues]**. Privacy, Transborder data flow.

General Terms

Management, Measurement, Documentation, Human Factors, Standardization, Legal Aspects, Verification.

Keywords

Privacy, Demographics, Data-collection practices, Web-crawling, Cookies, Webbugs, P3P, Legislative impact.

1. INTRODUCTION

The Web is a complex place in terms of technologies and practices, especially when considering how these affect privacy and security. These are important concerns for consumers who have to decide who to trust with their data, for legislators who have to develop meaningful and effective regulation, as well as for system administrators and developers, who stand to loose significant time and money on flawed models and designs, or potentially face a user backlash and/or fines.

Part of what makes this such a challenging problem is that technology and business practices are constantly evolving. Keeping up with changes and trends can sometime seem like a full-time job. Another challenge is that the web is a global system, crossing and blurring many of the traditional lines of jurisdictions. A company can be registered in one country, be hosted in a number of other countries, and do business with consumers from anywhere in the world. This picture can get even more complicated when we start talking about multi-national companies, and potential business-to-business (b2b) partners. This issue of jurisdiction has been, and will continue to be for the foreseeable future, a serious challenge to e-commerce and e-business. Determining compliance should therefore be a major concern for designers, developers, and administrators of such systems.

For legislators and policy makers it is therefore important to understand the impact of policy decisions in order to craft rules and legislation which will be effective and meaningful, and enforce such rules once adopted. Given that legislation often lags behind technological adoption and development, it is important to monitor when safeguards are needed, and when they are no longer meaningful or necessary. It is equally important to monitor developments following the introduction of new legislation as well, to ensure that these are having the intended and desired effects, something which is not always the case [22].

For consumers it is important to understand the risks out there - including the prevalence of undesirable or dubious security and privacy practices - in order to make better decisions about whom to trust. This is especially important as a mechanism for ensuring market forces take effect. If consumers are unaware of companies using undesirable practices, they cannot express their preferences by taking their business elsewhere. Such knowledge can help spur the adoption of effective and necessary safeguards and detection mechanisms, and can help end-users press legislators for regulation of practices.

For researchers, it is important to know what problems, technologies and practices are worth addressing, or which remedies are having effect. When designing monitoring, notification, blocking, or any other type of technologies, it is important to know where best to invest time and effort, especially given the limited resources in many academic settings. Such an overview could help researchers make the necessary justifications for their decisions.

In order to meet the information needs of such diverse stakeholders we need access to a reliable set of data about current data practices and technology use. Because this data may influence public policy, consumer perception, as well as business practices, it is essential that the data be publicly available, and collected in a transparent and unbiased fashion. A technique for doing this is to instrument a web crawler, specifically designed to

Symposium On Usable Privacy and Security (SOUPS) 2007, July 18-20, 2007, Pittsburgh, PA, USA.

go out and index web-pages based on publicly visible and machine identifiable data-collection practices and policies. This data could then be made available to the public, and/or scrutinized, and used as a common benchmark or reference set. This basic approach has been used in the past [10], though not on the scale of what we demonstrate in this paper.

Our proposal for filling this function is a web crawler named iWatch. The name is derived from the famous question "Q*uis custodiet ipsos custodes?*" or "*Who watches the watchers/guards?*" originally posed by Plato in *The Republic* [31] and popularized in Latin in Juvenal's *Satires* [24]. In this case, iWatch monitors those who normally monitor us; websites.

iWatch is meant to serve as a source of basic statistics on the state of privacy, security, and data-collection practices on the web. Because we have no access to information on what websites are doing behind the scenes we have to limit our analysis to the information and technologies which are publicly visible, and what we can automatically detect and analyze. Though this naturally limits the accuracy and scope of our analysis, it still allows us to examine and detect some fairly interesting practices and situations.

In this paper we set out to demonstrate the feasibility and value of this approach to analyzing real-world data-practices from the perspective of the outside observer (no knowledge of internal website workings). We will look at several interesting practices, and ways of examining the data. This paper is also meant to serve as a point for reflection and discussion about which practices to observe, and how the raw data from a system such as iWatch, which is still a work in progress, can and should be evolved and made available to a wider audience.

The structure of the rest of this paper will be as follows: We will first discuss a selection of related work, followed by a description of the terminology, conventions and definitions used in this paper. We then discuss the workings and implementation decisions made in our web-crawler, and present two sets of data, from 2005, and 2006, and explore the changes which have taken place in this period, as well as the impact of geography and regulation. We wrap up with a discussion of these results and future plans.

2. RELATED WORK

Privacy and security have long been recognized as important areas of concern, both offline and online. As such, this is one of the areas where online activity already has a long history of legislation. These laws have taken different forms across the globe. In Europe, comprehensive or omnibus laws for data protection have been enacted, while the US has largely implemented sector specific laws. These two approaches are fundamentally different, both approaches having advantages and disadvantages, which are often hotly debated [26, 33].

Regardless of approach, the goal of these privacy laws is to protect the Personally Identifiable Information (PII) of the individual, as well as regulate how information may be collected, for what purpose, and how it must be protected. Examples of such laws include the US Health Insurance Portability and Accountability Act of 1996 (HIPAA) [36], the US Children's Online Privacy Protection Act of 1998 (COPPA) [34], the US Gramm-Leach-Bliley Financial Services Modernization Act of 1999 (GLBA) [35], and the European Union Directive on the protection of personal data (95/46/EC) [17].

Given that studies have shown that users fail to read sites' privacy policies [22, 27], the kinds of minimum protections these laws put in place are particularly important. Previous research has shown that legislation can have mixed effects on policies, especially their readability and usability [3, 22].

Despite legislative efforts, privacy concerns have been shown to be major obstacles to the adoption and success of e-commerce [1, 23]. Numerous surveys indicate that people consider privacy to be important [6, 7, 11, 15]. Privacy concerns are the most cited reasons for avoiding the use of e-commerce systems, an aversion that industry groups estimate costs e-commerce companies USD 25 billion per year in lost revenue opportunities [23]. Surveys have also found that people are more concerned about their privacy online than offline [21], even though most cases of identity theft occur offline [20]. It is not surprising that industry groups invest significant resources to build consumer confidence and engage in voluntary efforts such as publishing privacy policies and seeking different forms of certification.

Such self-regulation attempts through programs such as seal programs such as TRUSTe (http://www.truste.org), BBBOnLine (Better Business Bureaus Online Seal, http://www.bbbonline.org), MultiCheck and WebTrust (offered by American Institute of CPAs http://www.cpawebtrust.org) allow licensees who abide by posted privacy policies and/or allow compliance monitoring to display the granting organization's seal of approval on their web site. Such programs have been found to significantly increase consumer trust [21, 28, 29], though some questions remain over whether what they imply matches user expectations, and questions remain about the ease with which sites may misrepresent their certification status [29]. In other words, there is some indication that users are being misled, intentionally or unintentionally, by some of these efforts [27].

We also know from surveys that though users think it is important for sites to present privacy policies, they are less than impressed with their quality and accuracy [12]. Surveys show that users find privacy policies to be boring, hard to read and understand, hard to find, and that they don't answer the kinds of questions they are interested in. The same survey also found that most people do not believe the claims and guarantees made in privacy policies [12, 20]. While most surveys report that a sizable portion of users claim to read such policies or notices regularly [12], there is evidence to suggest these reports are greatly exaggerated [21].

To overcome some of the problems associated with privacy policies and reduce the burden on users, machine-readable policy specification languages, such as P3P [8, 9] and EPAL [5], have been proposed. These policies can be read by automated agents (such as Privacybird [9], Privacy Fox [4], or the Microsoft IE 6 and 7, or Netscape 7 browsers themselves), only alerting users if the policy is likely to cause concern. The theory is that by filtering out the noise and drawing users' attention to only those policy elements which require attention, users are more likely to be engaged.

The most popular and widely used of these technologies without question is P3P. The Platform for Privacy Preferences (P3P) was created by the W3C to make it easier for web site visitors to obtain information about sites' privacy policies [8, 9]. P3P specifies a standard XML format for machine-readable privacy policies that can be parsed by a user-agent program. These tools have shown some indications of success [16], though there is little data on their effects during long-term or large-scale use. P3P policies have also been used as data to direct users' web-searches

[10] in a system sharing many methodological similarities to our iWatch.

A number of other tools independent of P3P have also been developed over the years, including filtering and privacy protecting proxy servers, popup-blockers, cookie blockers and analysis tools, anti-phishing tools, etc. Given that many of these functions have subsequently been absorbed by the latest generation of web-browsers, their numbers and user base is unknown today.

Regardless of the underlying technology, HCI researchers have been examining the issue of how to improve the usability and usefulness of such systems, an early shortcoming of many. Classic papers and studies include [37, 38]. This research showed that a secure system would fail unless these security measures were made usable. In recent years we have seen excellent papers on why phishing attacks work [14, 13], and how our tools and warning tend to go unheeded, regardless of the information presented [39]. While excellent results, it is obvious more work still needs to be done in this area as there are far more studies of why things fail than how to succeed.

Our approach of harvesting and examining large amounts of data via the use of a web-crawler has been employed by other security and privacy researchers. Recently, this approach has produced interesting results in the identification of malware and spyware disseminating websites [30, 32]. In these studies, researchers were able to scan and classify a large enough sample to convincingly argue about the state of the Internet as a whole.

3. Definitions

Before diving into the meat of our study, it is important to define certain terms in order to avoid misunderstandings or ambiguity. Our definitions should most often match generally accepted definitions, but may in some cases have a rather more narrow definition, chosen for practical considerations.

In this paper, domain, web server, and website are terms which are used interchangeably. While in the real-world, a given domain can host many distinct sites, we differentiate between sites based solely on domain-names. A distinct domain-name in our study identifies a distinct domain. Our classification of domains was very simplistic. We did not attempt to identify synonymous domain names (www.theregister.co.uk is not recognized as a synonym for www.theregister.com), or sub-domains (news.bbc.co.uk is not identified as a sub-domain of www.bbc.co.uk). The first is a hard problem and requires either a set of records from domain registrars, or a lot of hand-tuning. The second, though technically simple to implement, would cause problems with hosting services and smaller or related web-sites, which may lack unique second-level domain names.

We will also use the terms 1st party and 3rd party frequently. In this context a 1st party typically refers to the domain or website which served the page, and a 3rd party is any other domain/website which either receives information about the transaction, or supplies information or resources used by the requested page. Examples are 3rd party cookies, webbugs, and banner ads.

In this paper we will talk about technologies such as P3P policies, webbugs, cookies, popups, and banners. P3P stands for the Platform for Privacy Preferences, and is a standard for specifying privacy policies in a machine-readable XML format [8]. There are two types of P3P policies, the compact policy (CP) and the full policy. The P3P compact policy is a keyword abbreviated P3P policy, offering less detail and nuance, but often used by browsers to filter cookies. P3P and P3P policy will be terms that are used interchangeably in this paper.

The P3P protocol specifies 3 ways of publishing a P3P policy; in the HTTP header (can either be a compact policy, or a link to a full policy), in the HTML document as a link tag, or in a well-known location on the server. Because of some quirks of the way web servers implement the serving of P3P policies (see discussion in methodology), our current version of iWatch only finds policies posted in the HTTP header or the body of the document, it does not search the known locations. In order to fetch these remaining policies without bringing the crawler to a halt we delegate this task to a standalone program.

Privacy Seals are, in this paper, a combination of different certificates or trustmarks issued by TRUSTe and BBBOnline (BBBPrivacy and BBBReliability seals). These seals certify that the site discloses or follows a minimum set of privacy protection and security practices. While different seals or certificates are enforced by different agencies, have different meanings, and offer different enforcement mechanisms and guarantees, they are all meant to calm potential users concerns. Given the relatively low usage numbers, the different seal programs are grouped together for most of our analysis.

Webbugs, also known as web-beacons or pixel tag, are a collection of techniques aimed to tag and collect information from web and email users without their knowledge. In a web page, webbugs are typically used to track users navigating a given site, and have become quite ubiquitous. Webbugs technically can be implemented through a number of different techniques, but are most commonly associated with a 1x1 pixel transparent gif, invisible to the user. Webbugs are often used to augment the tracking available with cookies, and are most troubling when set by third parties, usually without user knowledge or consent. In iWatch we group a number of tracking techniques under the label of webbugs, but only when these are set and used by 3rd parties. We do not classify banner ads or 3rd party cookies as webbugs, but rather track these separately.

Much has been written about cookies, and so a discussion of how they work and their potential threats to user privacy is omitted here. We will just mention that in this work we do track the three main categories of cookies separately, session cookies, defined as cookies set by the first party and expiring with the browsing session, 1st party cookies, set by the 1st party and set to persist, and 3rd party cookies, which are set for any domain other than the 1st party.

Unsolicited popups, or just popups for short, refers to the much hated technique of opening new browser windows, typically for the purpose of advertising. Affiliated techniques include the pop-under (popups which try to hide themselves). They present a potential danger to end-users as they often serve up content for third parties, enabling these to track users much like webbugs. Popups have stopped being as big a focus in recent years as blocking tools and techniques have become ubiquitous and effective.

Web banners, or banners for short, do not present a privacy risk in and of themselves, unless served by a third party. In this case, they serve much the same function as a webbug, though at least remaining visible to the user. Banners in our study are identified by their size (these are the standardized sizes set by the Internet

Advertising Bureau (http://www.iab.net/standards/adunits.asp)), and the fact they are served by a 3rd party.

Some practices and technologies are ambiguous or difficult to detect reliably. This is especially true for automatic pop-ups, which at times are difficult to disambiguate from user-activated pop-ups, or webbugs from images or tricks used to layout web-pages. While we have done our best to unambiguously define and detect interesting practices, there is still room for improvement. Webbugs and unsolicited popups are still difficult to detect unambiguously, and some amounts of false-positives are still detected.

4. METHODOLOGY

iWatch is a web-crawler, or spider [19], implemented in Java, and built from the ground up to search for and index data-handling practices. Similar to most crawlers, which search for and index key words, or all words within the body of a document, iWatch is designed to look for certain HTTP tokens, or HTML constructs and patterns, which may identify certain data-handling or collection techniques of interest.

Like any web-crawler, iWatch starts with a seed-list, or given set of URL's which to visit initially. iWatch downloads these pages in parallel using multiple threads, and searches the resulting download stream for web-links and a set of filters. This process is partially done using Java's built in classes and their data-handling functions (such as finding links in a HTML document), and a set of full-text searches using regular expressions.

Links found are added to a database of pages to potentially crawl. Given that most websites are complex in structure, iWatch seeks to analyze a number of pages within each domain in order to get a more complete picture of the site. At the same time, iWatch seeks to minimize the impact on the servers studied by limiting the number of pages requested from any domain. This also ensures that iWatch does not get stuck analyzing big sites, ensuring we get a minimum breadth of coverage. When a thread is idle, or is done analyzing its current page, it consults the database of links found, selecting the next eligible link and repeating the process.

Because the initial seed-list used has a tremendous effect on the overall crawling pattern it is important to choose carefully. Given the limited resources of a university/research setting, the crawler will only be able to visit a very limited number of pages and domains when compared to dedicated operations such as Google and MSN. The seed-list must therefore be selected so that the sample taken is a) as representative as possible, b) as relevant as possible, and c) leads down a path of diversity of sites.

These criteria are not always achievable. A fully representative sample would require a random sampling, which is not possible with a web-crawler, which by its nature investigates clusters of websites by following the links between these. Instead, we have chosen to construct our seed-list based on the data's potential value or impact. In other words, we ensure that the most popular sites, the sites most likely to impact the privacy of the most users, are at the heart of the crawl. In addition, to avoid an overwhelming US and English language bias, the sample must be balanced to include different countries and classes of websites. For our experiments, the crawler was seeded with a combination of the top 50 websites for that month (as determined by the Comscore MediaMetrix (http://www.comscore.com/metrix)), and a hand-picked set of popular European and Asian sites. This is far from a perfect selection of sites, but gives us an interesting and relevant sample to study.

Given a functioning web-crawler, one then needs a set of search criteria to index the pages. Table 1 gives an abbreviated list of the main bits of information we currently collect using iWatch. Many of these are composed by multiple regular expressions of mechanisms. For instance, cookies are identified by one of three filters, depending on whether they are session cookies, 1st party cookies, or 3rd party cookies. For each of these, different information is collected. iWatch collects information on 21 data-practices plus assorted site-characteristics such as geographic location based on IP address matching.

Our indices were derived from the filters used in the privacy-protecting proxy server called Privoxy (http://www.privoxy.org). Privoxy is an open-source proxy server designed to act as a filter between a browser and the web. In order to do this, Privoxy filters incoming and outgoing HTTP communication using a set of regular expressions identifying potentially dangerous or undesirable practices from an end-user perspective. These filters were manually tuned to remove some false-positives (especially in the area of webbugs) and give us more information to process.

Table 1: Main iWatch index terms

Index Terms	Description
Cookies	Identifies the use of different types of cookies (session, 1st party and 3rd party), and their characteristics
Unsolicited popups	Identifies the use of unsolicited popup windows
Webbugs	Identifies the use of third part resources potentially used to track users from site to site
Banners	Identifes the use of different types of banners and ads, potentially used to track users from site to site
P3P policies	Identifies the use of both full and compact P3P privacy policies in HTTP header
Privacy Seals	Identifies the use of Privacy seals (TRUSTe, BBBOnline, and WebTrust) in a domain's pages (link and graphic)
Data-sharing networks	A collection of the techniques used to track users across sites (3rd party cookies, webbugs, banners), and who the data is shared with
Link structure	Basic information on page's link structure and relationships between sites
Geographic information	Maps a domain/server's IP address to a country using the GeoLite database created by MaxMind (http://www.maxmind.com/)

Based on early experiments, we learned that in order to correctly identify P3P and privacy seal use, we needed to adopt a strategy other than filters. While filters effectively identify the use of compact and embedded P3P policy references, finding and downloading full P3P policies requires additional steps, which are prone to errors. As pointed out in [10], some servers will at times refuse to serve some full P3P policies from the default location (http://server/w3c/p3p.xml), skewing results. In order to ensure more correct results, we wrote a custom application that revisited each of the domains in our samples 3 times trying to get a full p3p policy. These repeated queries made a significant difference in our results, giving us an additional 117 policies for our 2006 sample, and 211 additional policies in the 2005 sample when compared with a single visit strategy. Responses were analyzed to check that what was returned was an xml document and not a html document, and that redirects were followed correctly. In the current version of the crawler, the P3P policies are not analyzed.

Our early attempts at determining seal usage directly from the pages we crawled also proved to be an ineffective strategy. Seals

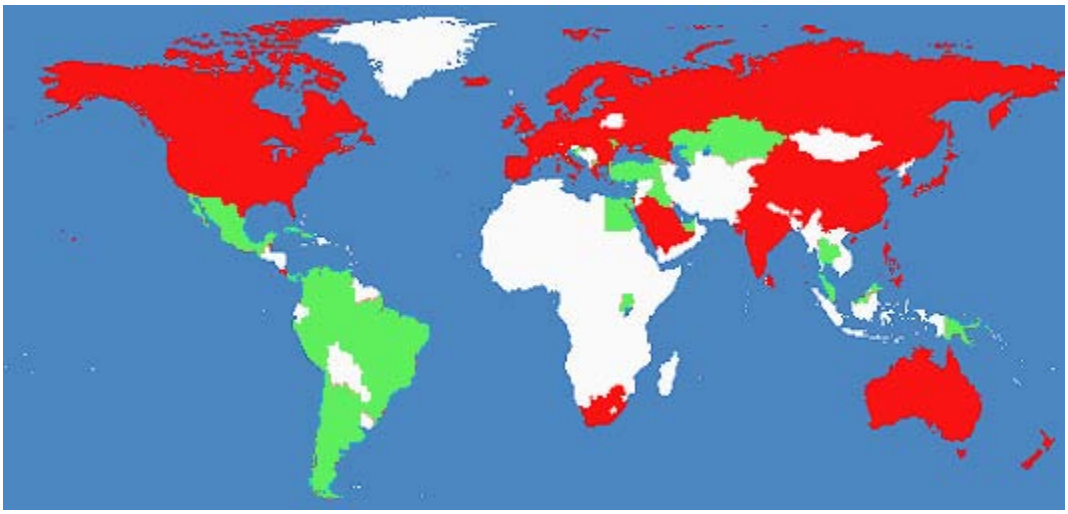


Figure 1: Geographic distribution of sample

Countries marked in red are included in the study. Countries marked in green were reached, but excluded from the study due to small sample size. Map courtesy of world66.com

are typically confined to a disclaimer or privacy policy page, therefore our ability to detect seal use through filters depends on a) the crawler having reached a policy page for the site, and b) that the seal is presented using a standard format. Of these, the first hurdle proved to be the most significant and eventually insurmountable obstacle to this strategy. To overcome these limitations we gained access to lists of certified sites directly from the certifying agencies (in this case TRUSTe and BBBOnline). These lists (http://www.truste.org/about/member_list.php, and http://www.bbbonline.org/consumer/pribrowse.asp) were then cross-referenced with our sample sites. We were unable to obtain lists for other seal providers, though this is something which we will seek to work on in the future.

To demonstrate the effectiveness and value of this approach to the study of online privacy, online regulation, and online data collection practices, we performed two experiments, one in May of 2005, and the second in August 2006, where we collected information on web-sites' privacy and data-collection practices. Each of these crawls was performed over a period of 10-14 days, with our crawler running on a single dedicated server. In this paper we will use these two samples to examine the changes that have taken place online over the last year.

Before concluding this section we wish to say a few things about the statistical testing using our two sample databases. Given the large size of our two samples, finding statistical significance is relatively simple even for relatively small changes in behavior. The reader is therefore advised that it is important to make a distinction between statistically significant and meaningful changes when considering this data. We therefore, uncharacteristically within the field of computer science, choose to set our threshold for statistical significance at the $p<0.001$ level throughout this paper, unless otherwise noted.

5. RESULTS

5.1 Sampling Results

Across both samples, a total of 240,340 web pages were crawled, from a total of 24,990 unique domains. There was an overlap of 1,223 domains between the two samples, or 4.7% of the total sample domains, for a total of 26,213 non-unique domains across both samples. Given that these samples were taken 15 months apart, and the speed with which websites evolve, we decided to use the non-unique total in our calculations, and treat the two samples as statistically independent. This means that on average we analyzed 9.17 pages per domain, a relatively solid basis for drawing conclusions about any given domain. Table 2 summarizes the basic characteristics of the two samples.

Overall, our two samples reached 81 countries or territories, 69 in the first sample and 60 in the second, despite the crawler being primarily seeded with US web-sites (Figure 1 shows an overview of our geographic reach). Many of these countries were represented by extremely small number of domains and pages in our data-sets, which forced us to filter some of the data to avoid drawing conclusions on overly thin data. We decided to exclude from analysis any country which was not represented by more than 10 domains across both samples, unless they were part of the European Economic Area (EEA).

Table 2: Data sample summary statistics

	Sample 1	Sample 2	Total
Collection	May 2005	August 2006	
Web-pages	119,237	121,103	240,340
Domains (unique)	15,792	10,421	26,213 (24,990)
Web-Pages/Domain (unique)	7.55	11.62	9.17 (9.62)
Total Countries	69	60	81
Filtered Countries	43	43	47
Domains/Country	367.26	242.35	557.72

The EEA is composed of the 25 European Union (EU) members, plus Iceland, Liechtenstein and Norway. All domains belonging to any EEA country were included in our sample because all EEA countries are signatories to the EU privacy directive [17], and therefore have similar privacy legislation in place. For the purpose of this analysis, the EEA countries will be viewed as a block. Of the 28 EEA countries, we found 27 in our sample (Lichtenstein being absent, see table 3 for list of all countries included in study). EEA countries make up 9.66% of our total sample.

Applying the above filtering rules, we lose 56 domains and 26 countries from Sample 1, and 34 domains and 17 countries in sample 2. Overall, 34 countries were filtered from the combined data-set, leaving 47 (43 in each of the samples). On average, the

excluded countries were only represented by 2.64 domains. As could be expected, our probes primarily reached the most net-active countries in world. Though we only saw a total of 47 countries, those countries account for more than 96% of all active domains (http://www.webhosting.info/domains/country_stats). This means that though our samples only reached approximately 0.019% of all registered domains, these samples are representative of a large percentage of the net.

Table 3: Geographic distribution of sample and bias

Countries highlighted in grey to indicate EEA membership.

Based on a test of proportions, and # in the bias column together with green and tan highlight indicates significant positive or negative bias (P<0.001) respectively*

Country	Total		Samples		Bias (% of expected)
	Number of Domains	% of Domains	Number of Domains	% of Domains	
United States	46,036,912	67.56%	21,949	83.73%	* 123.94%
EEA	12,526,739	18.38%	2,531	9.66%	# 52.52%
Germany	4,039,278	5.93%	416	1.59%	# 26.77%
United Kingdom	2,947,932	4.33%	930	3.55%	# 82.01%
Canada	2,495,501	3.66%	585	2.23%	# 60.94%
China	2,099,671	3.08%	114	0.43%	# 14.11%
France	1,733,082	2.54%	197	0.75%	# 29.55%
Australia	1,393,853	2.05%	177	0.68%	# 33.01%
Spain	884,969	1.30%	210	0.80%	# 61.69%
Japan	871,196	1.28%	213	0.81%	# 63.56%
Korea	837,088	1.23%	171	0.65%	# 53.10%
Hong Kong	763,480	1.12%	27	0.10%	# 9.19%
Italy	721,992	1.06%	43	0.16%	# 15.48%
Netherlands	547,838	0.80%	157	0.60%	# 74.50%
India	342,735	0.50%	102	0.39%	77.36%
Denmark	263,789	0.39%	40	0.15%	# 39.42%
Russia	240,386	0.35%	31	0.12%	# 33.52%
Sweden	209,208	0.31%	63	0.24%	78.28%
Switzerland	186,619	0.27%	62	0.24%	86.36%
Norway	172,123	0.25%	289	1.10%	* 436.47%
Austria	163,612	0.24%	37	0.14%	58.79%
Poland	141,423	0.21%	14	0.05%	# 25.73%
Finland	123,288	0.18%	22	0.08%	# 46.39%
Belgium	122,048	0.18%	37	0.14%	78.81%
Czech Republic	91,051	0.13%	12	0.05%	# 34.26%
Israel	81,883	0.12%	39	0.15%	123.81%
Bulgaria	81,290	0.12%	2	0.01%	# 6.40%
Ireland	73,363	0.11%	21	0.08%	74.41%
Portugal	56,850	0.08%	5	0.02%	# 22.86%
New Zealand	53,517	0.08%	14	0.05%	68.00%
South Africa	48,384	0.07%	13	0.05%	69.85%
Taiwan	48,254	0.07%	34	0.13%	183.17%
Romania	35,479	0.05%	8	0.03%	58.62%
Hungary	31,249	0.05%	5	0.02%	41.59%
Saudi Arabia	29,696	0.04%	30	0.11%	262.62%
Greece	27,661	0.04%	8	0.03%	75.18%
Philippines	25,859	0.04%	17	0.06%	170.90%
Luxembourg	23,819	0.03%	5	0.02%	54.57%
Gibraltar	19,162	0.03%	2	0.01%	27.13%
Costa Rica	19,152	0.03%	16	0.06%	217.17%
Estonia	14,640	0.02%	1	0.00%	# 17.76%
Lithuania	9,988	0.01%	2	0.01%	52.05%
Slovakia	9,892	0.01%	1	0.00%	26.28%
Latvia	8,332	0.01%	1	0.00%	31.20%
Sri Lanka	5,821	0.01%	41	0.16%	* 1830.99%
Malta	5,813	0.01%	1	0.00%	44.72%
Iceland	3,047	0.00%	2	0.01%	170.63%
Sample Total	**68,142,225**	**96.34%**	**26,213**	**100%**	
Global Total	**70,733,538**				

Table 3 shows the distribution of domains across countries, as well as the bias of the sample relative to the countries current (October 2006) internet footprint. As noted earlier, the sample is skewed in favor of US web-sites, and as a consequence many other countries are underrepresented (highlighted in shades of orange in Table 3), including most EEA countries (highlighted in light grey in Table 3). Some smaller countries, through quirks of the way websites link to each other, or current events at the time of data-collection, are over-represented in the sample. As an anecdote, the bulk of our Sri Lanka sample was collected during May 2005, when peace negotiations efforts were receiving widespread press.

Given the size of the sample we collected, and the fact that there was only minimal steering of the crawler through the initial seed-list, we expected there to be significant bias in our sample when compared to the real-world. Though, as Table 3 shows, the bias in our sample is statistically significant at the p<0.001 level for approximately half of the countries in our sample (predominantly among the most net-populous nations), this bias was less than we had expected. This shows that great care needs to be taken in ensuring a seed-list which is geographically proportionate, at least for the top 20 countries (each representing more than 0.50% of the overall global domain-population). Once we exit this exclusive group, quirks and bias are less important, given the small relative size of these countries. For instance, while Norway is over-represented with 223 domains (436.47% of the sample size we should have seen), this only accounts for 0.85% of the overall sample size. This is negligible when compared with the US sample, overrepresented by 4240 domains (123.94% of the expected sample size), or 16.18% of the overall sample size. The main source of bias in our sample stems from the US being heavily over-represented. Most other countries and regions are consequently underrepresented.

5.2 Data Practices and Evolution

These data-sets have the potential to facilitate the tracking of trends in data collection practices, to gauge the effect or adoption of new technologies, new legislative requirements, best practices, and help determine if we are seeing the intended or desired effects on practices on a national or global scale. Such an analysis requires historical data going back far enough to judge long-term and short-term effects, and enough detail to determine specific causes. Our current data-set only spans 1 year, and does not, to the best of our knowledge, span any immediately obvious legislative event of relevance, making it difficult for us to perform an in-depth analysis here as proof of concept. Instead, we will focus on identifying overall trends rather than testing a specific hypothesis.

Table 4: Global data-practices

Table shows % of domains adopting practices, and the geographic spread of these practices as % of all countries in our sample.

*Based on a test of proportions a * with green highlight indicates statistically significant increase from one year ago (P<0.001)*

(1) Note that the sum of cookies used is not the same as the sum of Session, 1st, and 3rd party cookies, as sites may set multiple cookies of different types.

	2005		2006	
Practice	**Domains**	**Countries**	**Domains**	**Countries**
Any P3P Use	24.84%	72.09%	25.90%	60.47%
Only Compact P3P Policy	1.37%	27.91%	* 1.83%	18.60%
Only Full P3P Policy	17.43%	72.09%	17.13%	58.14%
Compact & Full P3P Policy	6.05%	32.56%	* 6.94%	20.93%
Any Privacy Seal	1.99%	11.63%	* 2.03%	11.63%
Truste	0.73%	6.98%	0.95%	9.30%
BBBPrivacy	0.12%	2.33%	0.16%	2.33%
BBBReliability	0.46%	4.65%	0.92%	6.98%
Any Cookie (1)	24.03%	72.09%	* 29.08%	86.05%
Session Cookies	18.02%	72.09%	* 23.07%	86.05%
1st party Cookies	4.74%	53.49%	* 6.11%	51.16%
3rd party Cookies	3.53%	41.86%	* 5.76%	39.53%
Popups	23.59%	72.09%	24.61%	81.40%
Webbugs	33.85%	81.40%	34.52%	86.05%
Banners	8.73%	55.81%	* 10.31%	58.14%

Table 4 gives an overview of the most common and relevant data-practices with the potential to affect end-users' privacy (both negatively and positively). In this table we see both the prevalence of the data-practices for the two samples (as percentage of total domains exhibiting data-practice), as well as their geographic spread (as percentage of countries where at least 1 domain exhibits this data-practice).

Our first finding is that P3P is alive and well, with adoption among the sites in both our samples circling 25%. There were no statistically significant changes in adoption rates overall from 2005 to 2006, though the use of Compact Policies, with or without Full policies did increase significantly. These high adoption rates are likely in part due to the ubiquitous Microsoft IE 6 web-browsers' inclusion of P3P as a factor in blocking some types of cookies. Another area of good news is that though the use of compact policies is growing, use of the more expressive and meaningful Full policies dominates by a large factor.

Using our new and improved seal matching technique we see a small, but statistically significant increase in the use of privacy seals. We realize that our list of seal providers is simplistic and short, and that more providers need to be added in order to provide a more realistic picture of the use of seals today. As a point of contrast, others [1] have found that 11% of US websites had privacy seals in 2001. It is unlikely that seal adoption has decreased this significantly over the last 5 years.

Looking at the much maligned cookie, we see that overall use has increased markedly over the course of the year. This increase is seen both in the use of inoffensive session cookies as well as the more troubling 3rd-party cookie. We also see more sites using more than one type of cookie, though we have not computed statistics on how many cookies of the same type a site uses. The one bright note to raise here is that though the number of domains using 3rd party cookies grew, geographic distribution declined.

As expected from the improvements seen in terms of online ad revenues in the past year, we see a significant growth in the number of domains using banner ads. On the other hand, the use of unsolicited popups and webbugs is flat from a year ago, though geographic distribution is up.

Table 5: Effects of P3P and Privacy Seals on practices

Table shows % of domains adopting practices, the expected rates (product of the probability of the two practices), and the difference (diff) from this expected rate.

*Based on a test of proportion, cells marked by *or # with green or tan highlight in 2006 "Detect" column indicates statistically significant increase or decrease from one year ago (p<0.001, 2-tailed)*

Based on Chi-Square tests of independence, combinations marked with a ^ and highlighted blue in the "diff" columns were not statistically independent (P<0.001)

	2005			2006		
Practices	**Detect**	**Expect**	**diff**	**Detect**	**Expect**	**diff**
P3P+Webbugs	11.99%	8.41%	^ 142.6%	* 13.75%	8.94%	^ 153.8%
Seal+Webbugs	0.96%	0.44%	^ 217.0%	0.92%	0.32%	^ 289.7%
P3P+Popups	11.61%	5.90%	^ 196.9%	12.15%	6.37%	^ 190.6%
Seal+Popups	0.89%	0.31%	^ 286.9%	1.13%	0.50%	^ 226.1%
P3P+Session C	4.51%	4.48%	100.8%	* 5.70%	5.97%	95.4%
Seal+Session C	0.41%	0.24%	^ 174.2%	* 0.86%	0.47%	^ 184.0%
P3P+1st party C	1.48%	1.18%	^ 125.9%	1.66%	1.58%	104.9%
Seal+1st party C	0.24%	0.06%	^ 387.6%	0.33%	0.12%	^ 262.4%
P3P+3rd party C	1.61%	0.88%	^ 183.2%	* 3.22%	1.49%	^ 216.2%
Seal+3rd party C	0.24%	0.06%	^ 228.3%	* 0.51%	0.12%	^ 434.2%
Seal+P3P	0.60%	0.33%	^ 184.7%	# 0.33%	0.53%	^ 61.9%

Some of the most interesting findings from our study deal with the effect that the use of P3P and privacy seals has on the prevalence of other data-practices. What we are looking for here is whether the group of other practices is statistically independent from the use of privacy seals or P3P policies. In Table 5 we present the basic data, as well as the results of tests of proportions seeing whether the rate increased or decreased from one year to the other, and Chi-Square (test of independence) to determine whether the differences between observed or detected rates and expected rates differ in a statistically significant way.

As Table 5 shows, P3P and privacy seal use was not statistically independent from most of the other privacy indicators examined in this study. The presence of either of these indicators was usually associated with a positive co-occurrence rate. This may have had (and likely does have) a perfectly reasonable explanation in that sites with more complex information needs and data collection practices seek to assure and explain the use of other technologies through a P3P policy, or provide assurance of their intent through the presence of a seal. Because P3P policies were not analyzed in this study, we cannot say whether policies addressed or explained the use of the correlated technologies, though this is something which should be investigated in the future.

From 2005 to 2006 we saw a statistically significant increase in the use of P3P in conjunction with webbugs, session cookies, and 3rd party cookies, while the same was observed for privacy seals and session cookies and 3rd party cookies. This represents a mixed bag for end-users, as both desirable and undesirable practices showed an increase. On the other hand, the co-occurrence of privacy seals and p3p policies decreased significantly from 2005 to 2006, part of an observed trend in avoiding overlapping certification or explanation systems.

The prevalence of P3P use was an issue which we decided to explore in greater depth. Specifically, we wanted to explore to what extent P3P use was constrained, or influenced by the site's popularity (as defined by our seed-list selection). By partitioning the domains crawled into segments of 1000 domains we get a rough ranking of the sites (see Figure 2). This is dependent on the acceptance of a definition of popularity being the distance from the seed-list sites. While not a fully fair metric, it does fit with the way browsing patterns affect page rankings, and is probably good enough for the purposes of this investigation. As can be seen in Figure 2, popularity does indeed affect the adoption of P3P, though much more markedly today than in 2005.

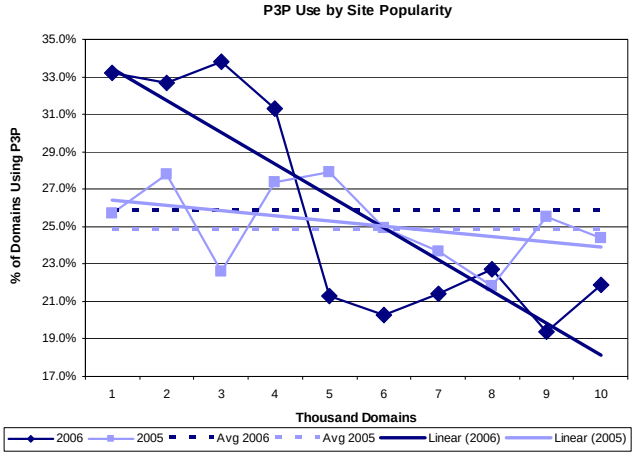


Figure 2: P3P use by site popularity

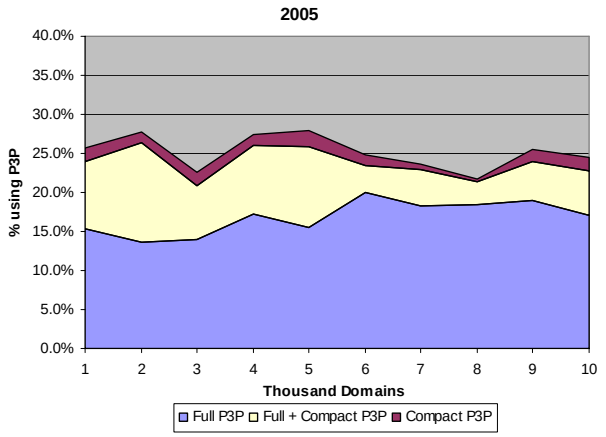


Figure 3: P3P use by site popularity and type, 2005

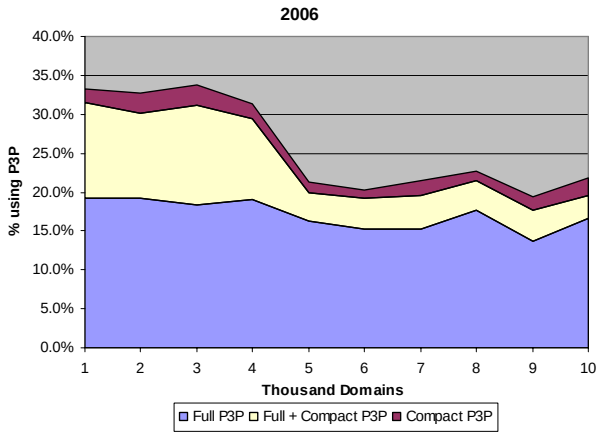


Figure 4: P3P use by site popularity and type, 2006

Figures 3 and 4 show how the use of P3P has evolved from 2005 to 2006 in terms of the types of P3P policies used, and the popularity of the sites using them. From figure 3 we can see that in 2005 as a sites' popularity decreases, fewer offer dual policies (fewer sites offer compact policies), instead offering only full policies. From figure 4 we can see that the increase in P3P use observed over the two samples is in large part due to a significant increase in the pre 4,000 sites, which are offering more dual and full policies. Beyond this, the distributions look very similar.

5.3 Legislation and Data Practices

As previously mentioned, one of the intended uses of these data-sets is to examine the effects that legislation and regulation have on data-practices. Given that no major new US privacy legislation took effect between our two samples, we instead use our samples to examine the privacy practices, and evolution of these between the US, Canada, the UK, and the EEA, all countries or regions with different levels of legislation regulating data-practices and the collection and use of PII. Table 6 gives an overview of the geographic clustering of data.

The most interesting elements for this analysis is the EEA and US columns, as they represent two ends of the spectrum in terms of privacy regulation and enforcement activity. The UK and Canadian samples are interesting because they serve as interesting points along this continuum. Both the UK and Canadian privacy regulations are stricter than those seen in the US, yet both are influenced by similar culture, language, technology adoption, etc. If legislation and user activism have an effect on the adoption of technologies and practices, we should see some systematic differences in this data, especially between the US and EEA.

Table 6: Geographic clustering of domains

Table shows number of countries and the % of all domains in each group and sample. In the total column we give the actual number of domains.

** UK appears both on its own and as part of the EEA sample*

*Based on a test of proportion, cells marked by *or # with green or tan highlight in 2006 Detected column indicates statistically significant increase or decrease from one year ago (p<0.01)*

	2005		2006		Total	
Geographic Area	Country Count	Domains	Country Count	Domains	Country Count	Domains (unique)
EEA	24	9.75%	25	9.52%	27	2,531 (2,483)
Canada	1	2.41%	1	# 1.96%	1	585 (576)
United Kingdom*	1	3.18%	1	* 4.11%	1	930 (899)
United States	1	83.28%	1	* 84.43%	1	21,949 (20,815)
Other	17	4.57%	16	4.10%	17	1,148 (1,117)

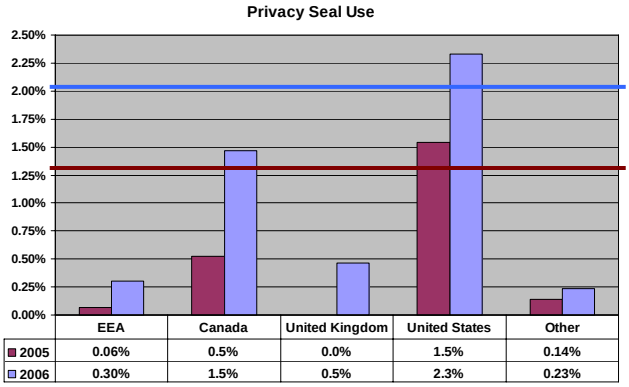


Figure 5: Privacy seals by geographic area

Horizontal bars showing global average for the two samples (by color). All changes from 2005 to 2006 except 'Other' category are statistically significant (p<0.005)

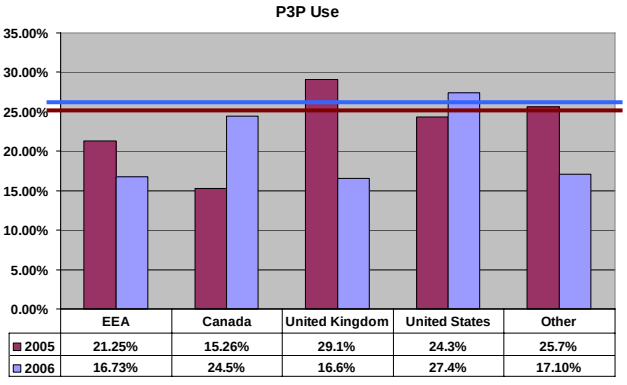
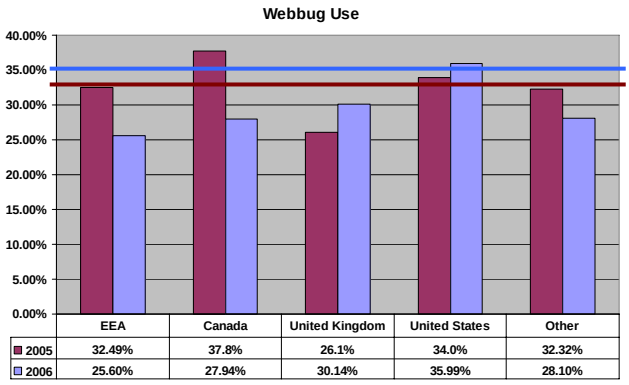


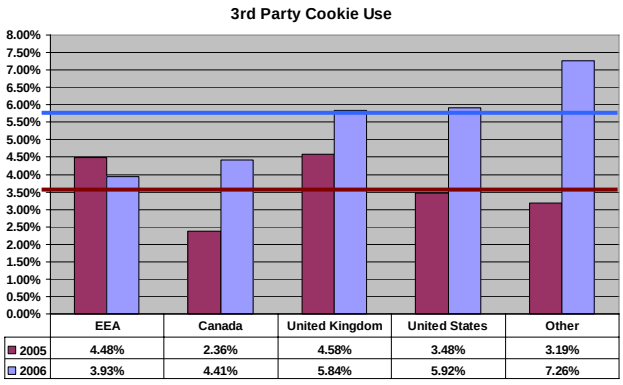
Figure 6: P3P adoption by geographic area

Horizontal bars showing global average for the two samples (by color). All changes from 2005 to 2006 are statistically significant (p<0.005)



	EEA	Canada	United Kingdom	United States	Other
2005	32.49%	37.8%	26.1%	34.0%	32.32%
2006	25.60%	27.94%	30.14%	35.99%	28.10%

Figure 7: Webbug use by geographic area
Horizontal bars showing global average for the two samples (by color). All changes from 2005 to 2006 are statistically significant ($p<0.005$)



	EEA	Canada	United Kingdom	United States	Other
2005	4.48%	2.36%	4.58%	3.48%	3.19%
2006	3.93%	4.41%	5.84%	5.92%	7.26%

Figure 8: 3rd party cookie use by geographic area
Horizontal bars showing global average for the two samples (by color). All changes from 2005 to 2006 are statistically significant ($p<0.005$) except for EEA group ($p<0.05$).

Some of the interesting observations are that, as Figure 5 shows, privacy seals are virtually non-existent outside of the US and Canada. Again, data for the use and adoption of privacy seals is incomplete and should be viewed with caution, but we would expect these deficiencies to play out evenly geographically, as all major certification agencies are US based. It is interesting to note that the only countries to use privacy seals in 2006 were the US, UK, South Africa, Canada and Belgium. Apart from the later, these are all countries where English is (one of) the official languages. In 2005, privacy seal use was restricted to the US, Canada, Japan, and Finland.

Another interesting finding is the skew in P3P adoption, with the US and Canada very much leading the way (Figure 6), with every other region showing a statistically significant decline. Determining why this is the case could be an interesting issue to investigate in the future, and would also require the analysis of the P3P policies themselves.

While other technologies could have been examined in this fashion, we decided to conclude this study by looking at two technologies which are particularly problematic for end-user privacy; webbug and 3rd party cookie use. Again, if regulation affects web-based data practices, this is where we should expect to see the biggest differences (see Figures 7 and 8). While the observed trends were in line with our expectations, the differences were not as marked as we had expected, nor were they uniform. The UK, a part of the EEA sample, consistently followed the patterns exhibited by the US rather than its European partners.

As noted at the beginning of this section, the impact of legislation on these practices remains a question which warrants further investigation. The short time spanned between the samples, the fact that at this point there are only 2 samples, and that no major piece of legislation was enacted which directly impacted online privacy practices, made it difficult for us to explore this use fo the data. With time however, we believe it will be interesting to investigate the long-term effect of legislation such as the GLBA on financial sites, or HIPAA on healthcare sites. This will however require a more longitudinal sampling method (given that both laws were in force when our first sample was taken), and a stronger focus on financial and healthcare sites.

6. DISCUSSION

The goals of this paper were to demonstrate the feasibility and value of using a system such as iWatch to study the current state of the art in terms of online practices and data collection techniques which may affect end-user privacy, and to provide a minimum set of current data about prevalent data practices. We believe we have demonstrated that the general approach is sound, though some fine-tuning is necessary. We have also generated a broad set of statistics which others may build on in their own research or system design. Having said this, a number of important lessons were learned as part of this study.

Given that we are using a web-crawler, following links as they appear on web-pages, our sample of domains is always going to be different from one crawl to the next. It is therefore difficult if not impossible to precisely control the distribution of sites. This presents two potential problems. The first is that it is difficult if not impossible to get a completely unbiased sample (at least in terms of geographic representation) by chance. Though for our purpose, some small adjustments are likely to be enough; those with a need for greater accuracy can enforce the distribution they desire by sampling from the dataset to achieve the right proportions of sites, though this would reduce the size of the overall dataset.

The second potential problem is that because of the dynamic nature of the web, any two samples are likely to deviate significantly in terms of the sites visited. If this deviation takes place early enough in the process, it may be difficult to directly compare samples. As an example, imagine that a significant number of the seed-list sites in instance A link to academic sites (due to some ongoing news story). In instance B, the same seed-list may instead point to a collection of e-commerce sites instead. In our samples, we had a seed-list of 100 items each time. Half that seed-list came from a public top-50 site list, and half the sites were manually picked to ensure a greater geographic distribution. Even though these samples were only separated by a year, there was only an 36% overlap in the top-50 site portion of the list. This likely lead to a significant divergence of the two samples, and possibly false inferences about changing practices, if the sample site is too small. With a large enough sample size, all things should even out.

This brings us to the question of whether a sample size of 0.02% of all domains is adequate for this kind of analysis. As a proof of

concept we were more than happy with this sample size, though for a production and archival system that may not be sufficient. While efforts to streamline data-collection, and thereby the resulting sample size can and will be made, the question of how much data must be collected and will need to be revisited.

One important area of bias which is not represented in Table 3, and for which we have no measure, but may nevertheless be of concern, is the likely under-representation of different market segments and domain types. Our seed-list was composed of the most popular websites of the day, all belonging to major corporations. Smaller "mom and pop" or non-commercial sites are therefore likely underrepresented. Previous research has shown that the web is not a completely connected graph. Rather, the web is a set of disconnected islands [18]. We therefore depend on a well-chosen seed-list to ensure that we can reach as many of these islands as possible, and have to accept that some sites will never be reachable. This is a possibility which concerns us, though the most popular websites are probably most important to most, a balanced, diverse sample would be more valuable overall.

We are also concerned about the difficulties we experienced in collecting full P3P policies, and the errors this could introduce into the analysis. We found that by trying to access full policies 3 times we got a significantly larger number of policies, but how many times should we try and access a server before giving up? Would we have found even more policies if we had checked back 5 times, 10, or 100? This instability is a problem which the community will have to address if P3P is to see further gains in adoption.

While there has been much debate about the value and shortcomings of P3P, the authors' perspective is that the adoption of technologies which communicate potential problems to the end-user (even if as some argue, flawed) can only be a positive thing. We were especially intrigued to find that the use of P3P policies coincided with the use of other, less desirable data collection practices such as 3rd party cookies and webbugs. Determining what the role of the policy was in that relation (smokescreen or explanation mechanism) is an interesting open question, one that would require us to parse the P3P policies.

Our inability to parse the P3P messages and compare their content to observed practices in time for this study is a significant shortcoming, and one which we will address in future work. Without knowing what P3P policies actually specify, and whether they contradict actual practices we cannot draw any solid conclusions as to the correlation between P3P adoption and things like 3rd-party cookies and webbugs.

We were reasonably pleased with our success with identifying sites using privacy seals (using official published lists from certifying agency). Early experiments trying to detect seals in the HTML stream yielded only a fraction of the sites found by matching against the seal providers lists, at a fraction of the cost. On the down-side side, our numbers are much lower than those reported by some others, leading us to conclude that in order for this to be a viable approach we need to broaden our list of seals. Search for seals in the HTML was appealing from the perspective of looking for misuse of seals, but this in retrospect turned out to be too difficult to do automatically. In [29], the reported detection of unauthorized seal use was performed manually, an approach which does not work with our intent of large-scale analysis. Automatically analyzing images unambiguously is very difficult, leading us to abandon these efforts.

Despite these shortcomings, our analysis also showed that there are interesting trends and patterns worth investigating with these datasets. One of the areas which we hope to expand into is the identification of best practices and guidelines to developers, legislators, and users. We also believe that these datasets could be of use to developers of privacy protection tools to either provide training or seed-date for more intelligent recommendation systems, or to inform where efforts are best spent.

One potential shortcoming to this geographic analysis is that our server is based in the US. In cases where we crawl multinational corporations or mirrored sites, our crawler is going to get directed to a US-based mirror. Given that we use GeoIP (www.maxmind.com) to map IP addresses to geographic locations for the servers, our results are necessarily be somewhat skewed, especially given that the sites most likely to engage in such behavior are the sites in our seed-list. Unfortunately, we do not at this time have a remedy for this problem.

7. FUTURE WORK

We have throughout this paper identified a number of shortcomings and caveats to our approach, discussing these where it seemed most natural. Our goals for the coming months are therefore relatively clear. We believe this study has validated the general approach, though some of the implementation details need to be refined. Our goal therefore is to address these as soon as possible so we can start to offer these data-sets to researchers, policy makers and tool developers on a regular basis (quarterly).

One of the areas of improvement identified in this study is the need for more careful balancing of the seed-list. We believe to have a strategy which will ensure a more balanced crawl, but acknowledge the fact that to a certain extent we are at the mercy of the tides. An intriguing possibility is to force the crawler to enforce the geographic proportions, but this would only work to ensure we do not over-represent any country or territory. There is however little we can do to ensure a minimum set of domains in a region short of stacking the seed-list.

In this study we also set an arbitrary cut-off point for countries (members of EEA, or the sighting of 10 unique domains in our sample). We now believe this policy may be less than desirable, and that instead a more reasonable policy would be to set a target for the number of domains to crawl, and close off countries or regions once their allotted quota of sites is reached. The list of links to crawl can quite easily be instrumented to keep track of the links' country of origin, which may then be used in the selection criteria.

We also need to reach out to more seal providers. While we have a short list of additional providers to contact, one difficult question is going to be again, when we have a complete enough set of seals, as well as ensuring that our list of seal certified sites remains up to date.

We believe that what we have been able to show in this paper is only the beginning of the kind of analysis which is possible with these types of data. The next steps includes looking at this data with more advanced statistical tools such as cluster analysis to look for patterns, either geographic or in terms of industries. We hope this kind of analysis can identify things like best practices, or industry conventions. This will helpfully help address some of the most serious points of criticism to this work, which is that though some of our analysis provides interesting insights, most of the data is rater superficial.

Working along these same lines we are currently trying to apply machine learning techniques to the datasets we have available, in combination with things like Netcraft's index of known phishing and malware sites, and their geographic locations. Using this data we hope to determine what meaningful risk indicators may be, in the hope of providing end-users with risk estimates before they follow a link or access a given site.

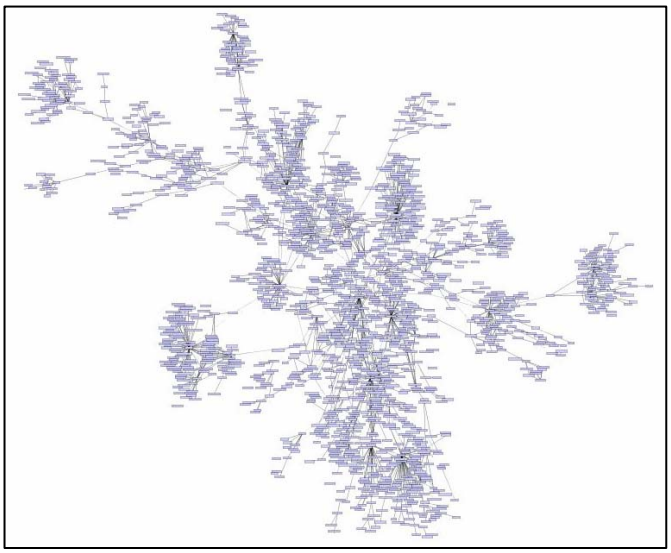


Figure 9: Data-sharing network based on cookies

Along these lines we are performing more sophisticated types of analysis, such as detecting and tracking information sharing networks composed of cookies, webbugs, banners and similar technologies. Initial explorations are promising. Figure 9 shows a network, visualized from real data, of sites connected to each other through 3rd party cookies. Each of the blue rectangles represents a domain, and each line a cookie. We were surprised by the number and size of the networks detected, and believe this can be a useful way of examining the spread of information.

We are also interested in looking for policy pages in order to try and combine our data with goals extracted from natural language policies, forming pseudo-machine readable policies, as explored by [3, 25]. These could then be compared to observed practices, and P3P policies to try and detect inconsistencies. Detecting inconsistencies between stated policy and observed practices will probably be one of the most valuable pieces of data in terms of identifying sites which put end-users' privacy at risk, either through malice or negligence.

Finally, we are naturally seeking to receive and incorporate feedback from other researchers on what practices to track and ways to track or improve the accuracy and value of our data. This could also potentially extend to accepting recommendations on new practices or technologies to track.

8. ACKNOWLEDGMENTS

This work was in part funded by NSF ITR Grant #0113792. We thank Yi Han Bae, John O. Ndukuba, Robert Marinski, and Leandro Taberner for their help in developing the iWatch crawler, as well as the support staff at The Georgia Institute of Technology and Oregon State University for their assistance. We also thank the students, colleagues and reviewers who have helped us with revisions of this paper.

9. 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] Anderson, R.E. "Social impacts of computing: Codes of professional ethics." *Social Science Computing Review*, 2 (Winter 1992), 453-469.

[3] Antón, A.I., Earp, J.B., Bolchini, D., He, Q., Jensen, C., and Stufflebeam, W. "The Lack of Clarity in Financial Privacy Policies and the Need for Standardization." *IEEE Security & Privacy*, 2(2), pp. 36-45, 2004.

[4] Arshad, F. "Privacy Fox - A JavaScript-based P3P Agent for Mozilla Firefox." *Privacy Policy, Law, and Technology*. 17-801

[5] Ashley, P., and Schunter, M. "The Platform for Enterprise Privacy Practices." *Information Security Solutions Europe*, Paris France, October 2002.

[6] Belanger, F., Hiller1, J.S., Smith, W.J. "Trustworthiness in electronic commerce: the role of privacy, security, and site attributes." *Journal of Strategic Information Systems* 11 (2002) 245–270.

[7] Campbell, A.J. "Relationship marketing in consumer markets: A comparison of managerial and consumer attitudes about information privacy." *Journal of Direct Marketing* 11, 3 (Summer 1997), 44–56.

[8] Cranor, L.F, Langheinrich, M., Marchiori, M., Presler-Marshall, M., and Reagle, J. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification. Retrieved Nov 10, 2004. http://www.w3.org/TR/P3P.

[9] Cranor, L.F. *Web Privacy with P3P*. O'Reilly & Associates, Inc., Sebastopol, CA, USA, 2002.

[10] Cranor, L.F., Bayers,S., Kormann,D. "Automated Analysis of P3P-Enabled Web sites" Proceedings of the 5th *International Conference on Electronic Commerce, ICEC2003*

[11] Culnan, M.J. "Georgetown Internet Privacy Policy Survey: Report to the Federal Trade Commission." Washington, DC: Georgetown University, McDonough School of Business.

[12] Culnan, M. J. and Milne, G. R. "The Culnan-Milne Survey on Consumers & Online Privacy Notices: Summary of Responses." Washington DC: FTC, December 2001.

[13] Dhamija, R.,and Tygar, J.D., ."The battle against phishing: Dynamic Security Skins." Proceedings of the *2005 Symposium on Usable Privacy and Security*.

[14] Dhamija, R., Tygar, J.D., and Hearst, M. "Why Phishing Works." In *Proceedings of CHI 2006*, April 22–27, 2006, Montréal, Québec, Canada.

[15] Earp J.B. and Meyer, G. "Internet Consumer Behavior: Privacy and its Impact on Internet Policy", *28th Telecommunications Policy Research Conference*, Sept. 23-25, 2000.

[16] Egelman, S., Cranor, L.F., and Chowdhury, A. "An Analysis of P3PEnabled Web Sites among Top 20 Search Results." *ICEC'06*, August 14—16, 2006, Fredericton, Canada.

[17] European Union (EU). *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.*

[18] Flake, W.G., Pennock, D.M., and Fain, D.C. The Self-Organized Web: The Yin to the Semantic Web's Yang IEEE Intelligent Systems, 2003.

[19] Heydon, A., and Najork, M. "Mercator: A scalable, extensible Web crawler." *World Wide Web* Volume 2, Number 4, December, 1999.

[20] Javelin Strategy & Research, *2005 Identity Fraud Survey Report*, January 2005. http://www.javelinstrategy.com/reports/2005IdentityFraudSurveyReport.html.

[21] Jensen, C., Potts, C., and Jensen, C. "Privacy practices of Internet users: Self-report versus observed behavior." *International Journal of Human-Computer Studie*s Volume 63, Issues 1-2, July 2005, 203-227.

[22] Jensen. C. and Potts, C. "Privacy Policies as Decision-Making Tools: A Usability Evaluation of Online Privacy Notices" *Proceedings of CHI'04* Vienna, Austria, April 2004

[23] Jupiter Research. "Security and Privacy Data." Presentation to the *FTC Security Workshop*, May 20, 2002

[24] Juvenal. *The Sixteen Satires*, Satire VI, verse 347. Penguin Classics; 3rd edition 1999.

[25] Karat,J., Karat,C.M.,and Brodie,C.A. "An empirical study of natural language parsing of privacy policy rules using the SPARCLE policy workbench" Proceedings of the second *Symposium on Usable Privacy and Security, SOUPS*

[26] Kuner., C. *European Data Privacy Law and Online Business.* Oxford University Press., 2003.

[27] Meinert, D.B., and Peterson, D.K. "Would Regulation of Web Site Privacy Policy Statements Increase Consumer Trust?" *Information Science Journal* Volume 9, 2006

[28] Miyazaki, A. D., Krishnamurthy, S. "Internet Seals of Approval: Effects on Online Privacy Policies and Consumer Perceptions" *Journal of Consumer Affairs*, Volume 36 Issue 1 Page 28-49, 2002.

[29] Moores, T.T., and Dhillon, G. "Do Privacy Seals in E-Commerce Really Work?" *Communications Of The ACM* December 2003/Vol. 46

[30] Moshchuk, A., Bragin, T., Gribble, S.D., Levy, H.M. "A Crawler-based Study of Spyware on the Web" in *Proceedings of the Annual Network and Distributed System Security Symposium.* San Diego, February 2007.

[31] Plato. *The Republic*. Penguin Classics; 2nd edition 2003

[32] Provos, N., McNamee, D., Mavrommatis, P., Wang, K., Modadugu, N. "The Ghost In The Browser Analysis of Web-based Malware." *First Workshop on Hot Topics in Understanding Botnets* April 10, 2007, Cambridge, MA.

[33] Schwartz, P.M., and Reidenberg, J.R. *Data Privacy Law: A Study of United States Data Protection.* Michie, 1996.

[34] United States (US) *Children's Online Privacy Protection Act of 1998*, Public Law No. 105-277, October 21, 1998.

[35] United States (US) *Gramm-Leach-Bliley Financial Modernization Act of 1999*, Public Law No. 106-102, November 1, 1999.

[36] United States (US) *Health Insurance Portability and Accountability Act of 1996*, Public Law No. 104-191, August 21, 1996.

[37] Whitten, A. and Tygar, J.D. "Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0," *Proceedings of the 8th USENIX Security Symposium*.

[38] Weirich D. and Sasse, M.A. (2001) "Pretty Good Persuasion: A first step towards effective password security for the Real World." *Proceedings of the New Security Paradigms Workshop 2001* (Sept. 10-13, Cloudcroft, NM), pp. 137-143. ACM Press.

[39] Wu, M., Miller, R.C., Garfinkel, S.L. "Do Security Toolbars Actually Prevent Phishing Attacks?" in *proceedings of CHI 2006*, April 22-27, 2006, Montréal, Québec, Canada.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Int. J. Human-Computer Studies 63 (2005) 203–227
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Privacy practices of Internet users: Self-reports versus observed behavior

Carlos Jensen[a,*], Colin Potts[a], Christian Jensen[b]

[a]*Graphics, Visualization and Usability Center, Georgia Institute of Technology, Atlanta, GA 30332, USA*
[b]*Department of Economics, Southern Methodist University, Dallas, TX 75275, USA*

Abstract

Several recent surveys conclude that people are concerned about privacy and consider it to be an important factor in their online decision making. This paper reports on a study in which (1) user concerns were analysed more deeply and (2) what users said was contrasted with what they did in an experimental e-commerce scenario. Eleven independent variables were shown to affect the online behavior of at least some groups of users. Most significant were trust marks present on web pages and the existence of a privacy policy, though users seldom consulted the policy when one existed. We also find that many users have inaccurate perceptions of their own knowledge about privacy technology and vulnerabilities, and that important user groups, like those similar to the Westin "privacy fundamentalists", do not appear to form a cohesive group for privacy-related decision making.

In this study we adopt an experimental economic research paradigm, a method for examining user behavior which challenges the current emphasis on survey data. We discuss these issues and the implications of our results on user interpretation of trust marks and interaction design. Although broad policy implications are beyond the scope of this paper, we conclude by questioning the application of the ethical/legal doctrine of informed consent to online transactions in the light of the evidence that users frequently do not consult privacy policies.

Keywords: Privacy; Design; Survey; Economic models; E-commerce; Decision-making; Policy

*Corresponding author. Tel.: +1 404 385 1102; fax: +1 617 373 5121.
E-mail addresses: carlosj@cc.gatech.edu (C. Jensen), potts@cc.gatech.edu (C. Potts), christia@mail.smu.edu (C. Jensen).

1071-5819/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2005.04.019

1. Introduction

Several recent surveys conclude that people are concerned about privacy and consider it to be an important factor in their online decision making (Cranor et al., 1999; Culnan, 1999; Earp and Meyer, 2000; Culnan and Milne, 2001; Jupiter, 2002). According to one study, privacy concerns are the most frequently cited reason for not engaging in e-commerce (Jupiter, 2002). Indeed, the increasing prevalence of data collection, sharing and storage mean that this may be an increasingly prudent position for consumers to adopt (FTC, 2000; Adkinson et al., 2002).

Most studies of user concerns about privacy have been done using a survey methodology. These studies report surprisingly high rates among respondents of such behaviors as reading a privacy policy when visiting a site or taking concrete steps to protect their privacy. Informal analysis of log-file data, however, suggests that the true rates are much lower (Jensen and Potts, 2004).

This paper presents the results of an empirical study comparing users' self-reported with their observed behavior in a simulated e-commerce scenario. In particular, we examined which visible indicators of privacy invasions or privacy guarantees were effective in swaying consumers' purchase decisions. We also examined what effects gender, level of experience, and other demographic variables have on reported and observed behavior. Finally we investigated the salience of categorization schemes for users privacy concerns based on survey responses. One such scheme used in Internet-based market research is the Westin privacy segmentation (Harris et al., 1998), in which people are classified into one of three groups; "privacy Fundamentalists", "privacy pragmatists", and "privacy unconcerned." Such schemes imply that users can be classified systematically and that a user's category helps predict the user's online behavior. Only by comparing self-reports with online behavior can such assumptions be verified.

2. Method

This study was conducted online, with subjects recruited through email announcements to mailing lists, and advertisements on academic websites. Over 175 volunteer subjects, predominantly from the United States, participated in the study. Subjects came from diverse backgrounds, though approximately two thirds were currently involved in education (students, faculty and researchers). Subjects were asked a series of multiple choice demographic questions. Subjects were anonymous; they did not need to, and were not given an opportunity to give any personally identifying information. Subjects did not receive compensation for their participation, there was no deception in this study; subjects knew the purpose of the study when they decided to participate.

The study was divided into four separate but interrelated sections: (1) A basic demographic survey. (2) A survey of privacy values and attitudes. (3) A set of questions challenging users' knowledge of specific technologies and how they affect privacy. (4) An experiment presenting subjects with a series of pair-wise comparison

tasks to determine the effect privacy indicators have on actual behavior. Subjects typically completed all four sections in one sitting, though they had the option to interrupt the study and return to it later. In all, subjects spent between 45 and 60 min on this study.

2.1. Demographic survey

In addition to collecting information on age, gender, and geographic location, we asked subjects about their educational and computer experience. Subjects, on average, were more highly educated (16.2 years of education) than the general Internet population (14.4 years of education) (NTIA, 2002; Jensen and Potts, 2004). Because of the self-selected nature of survey participation, we expect this population to be somewhat more concerned and knowledgeable of online privacy issues than the norm.

While 8.6% of the survey participants claimed never to buy things online, the majority purchasing things online at least once per month. On average, users reported their maximum online purchase to have been around $1000.00. These statistics indicate a survey population comfortable and familiar with e-commerce.

Our sample contained a larger group of men (74%) than women (26%), and subjects' ages averaged 30. Computer experience was high; the average respondent reported 7 years of online experience. Over 90% of participants reported having access to the Internet both from home and work, spending 25 h online per week.

The only statistically significant gender difference in the demographics was that women reported lower levels of computer expertise. The population average was 4.2 on a 5-point Likert scale, women averaged 3.6, while men averaged 4.4 ($p = .01$). Women consistently reported higher levels of concern with privacy, and online privacy in particular, though none of these differences proved statistically significant.

In terms of exposure to fraud and identity theft, our sample (see Table 1) matched data reported for the general population by the Federal Trade Commission (Synovate, 2003). Consistent with the findings of a recent study (Javelin, 2005), the majority of reported cases of identity theft and credit-card fraud originate offline rather than online.

Table 1
Victimhood and self-protection

	All (%)	Women (%)	Men (%)
Victims of identity theft	6.8	10.5	6.3
Victims of online identity theft	2.3	10.5	0.0
Victims of credit card fraud	14.3	26.3	11.9
Victims of online credit card fraud	4.6	5.6	4.6
Have installed software to protect online privacy	37.9	43.8	39.4
Have taken other steps to protect online privacy	42.7	58.8	40.3

Percentage of survey participants who claimed to have been the victims of identity theft or fraud, or taken steps to protect themselves.

One interesting finding from this section was the surprisingly large number of subjects claiming to have taken steps such as installing some form of privacy or security software. In the results section of this paper we explore the relationship between the installation of such software and the Westin privacy segmentation of users.

2.2. Privacy values

The second part of this survey consisted of a number of 5-point Likert-scale questions relating to attitudes and expectations on privacy, both online and offline. This section also focused on subjects' use of privacy policies, asking them to rate their likelihood of reading a sites' privacy policy based on the type of site, the activities they were engaged in, and their familiarity with the site. The exact questions and subjects' responses are reported in the results section.

This section of the survey was used to map our subjects to the Westin privacy segmentation (Harris, 2003). This index divides the population into three groups based on their level of concern with regards to privacy. This segmentation has been widely adopted, and is widely used to direct marketing and research efforts. Subjects are categorized based on their answers to three questions:

- Consumers have lost all control over how personal information is collected and used by companies.
- Most businesses handle the personal information they collect about consumers in a proper and confidential way.
- Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

Subjects giving privacy-oriented answers to all three questions are classified as "Privacy fundamentalists," those giving no privacy-oriented answers are classified as "Privacy unconcerned", while those in-between are classified as "Privacy pragmatists".

We chose not to use the three Westin classification questions because we wished to ask questions more directly related to online privacy. We placed our subjects into one of the three Westin categories based on the general pattern of their responses. We picked questions which corresponded closely to those used in the Westin surveys. Because we did not ask the same questions as Westin, the mapping is imprecise, but corresponds sufficiently for the purposes of our analysis. To avoid confusion we refer to our mapping as the Westin equivalence.

We chose to use five questions to map to the Westin groups rather than three, resulting in a more robust definition of the three categories. These questions were chosen to match what we considered to be essential properties of the three categories of users. These questions were as follows:

- I am concerned about online identity theft.
- I am concerned about my privacy online.

- I am concerned about my privacy in everyday life.
- I am likely to read the privacy policy of an ecommerce site before buying anything.
- Privacy policies accurately reflect what companies do.

We classified a participant as a "Fundamentalist" if he or she gave a privacy-oriented response to four of these five questions (and no negative answers). A participant was classified as "Unconcerned" if he or she gave no privacy-oriented responses (and at most one neutral response) to these five questions. The remaining participants were classified as Pragmatists.

Ninety-three participants completed the attitude survey and were classified as shown in Table 2. The rightmost column gives the proportions of respondents in a recent poll who were classified according to the corresponding Westin categories (Harris, 2003). The results from our classification are in line with the results of the Harris-Westin privacy polls conducted in recent years. The only slight difference is that our classification led to a more evenly divided population, with greater percentages falling in the Fundamentalist and Unconcerned categories and fewer in the Pragmatist category. The values we observed for each category was within the range of what has been reported in surveys in the past three years. It is not clear how much of this effect can be attributed to our defining questions as opposed to the way we selected participants for this study.

2.3. Knowledge challenge

One of the consistent problems with privacy surveys is the tendency subjects have of over-reporting their understanding of privacy-related issues and their willingness to act in order to protect their privacy. In order to test users and determine how big this perception gap is, we included a set of knowledge challenges in our survey. These challenges were focused on three commonly used and discussed technologies which may impact user privacy: Cookies, Web-bugs and P3P privacy policies. We chose these technologies in particular because they are parts of the vocabulary users are frequently assumed to be familiar with when setting privacy preferences (for instance in Microsoft's Internet Explorer 6.0).

Table 2
Population privacy classification

	Harris-Westin Polls				Survey—2004 (Count)
	1999 (%)	2000 (%)	2001 (%)	2003 (%)	
Fundamentalist	25	25	34	26	34% (32)
Pragmatist	54	63	58	64	43% (40)
Unconcerned	22	12	8	10	23% (21)

Percentage of the population as classified by the Westin Privacy Segmentation, and our Westin equivalence test.

When participants claimed to know what these technologies were, they were asked to rate their level of concern as well as select a reason why this technology may impact their privacy. Subjects were given a list of five possible reasons, two of which were correct, and three of which were incorrect. Users could select any number of these reasons, and we counted any answer which contained at least one correct option as a correct answer. We used these responses to gauge what percentage of subjects was truly familiar with a technology.

2.4. E-commerce experiment

To further test reported behavior against actual behavior, we included an e-commerce experiment to complement the survey sections. Each participant was presented with eight pairs of simulated e-commerce web-pages, one pair at a time, and asked to select which site they would prefer to buy from. Fig. 1 shows an example testing the difference between the use of the TRUSTe symbol and credit card icons. Subjects knew these were not real e-commerce sites, and that no money was being exchanged. The contents and design of the pages were in all cases similar

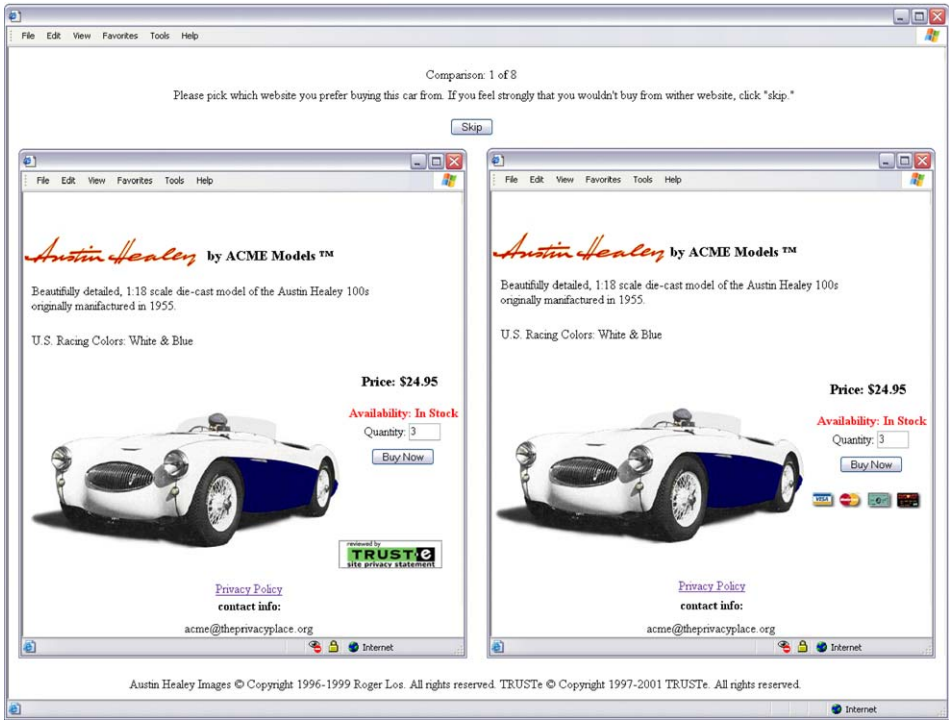


Fig. 1. Screenshot from e-commerce experiment: In each scenario subjects were asked to identify which site they preferred buying from. The websites were identical except for two factors. In this case, one site uses a TRUSTe logo and the second uses credit card icons.

and involved a controlled variation of twelve factors commonly cited as affecting e-commerce decision-making.

The independent variables were: (1) Price of item; (2) visible indication of Secure Socket-Layer (SSL) encryption; (3) use of third-party cookies and P3P; (4) providing an e-mail address, (5) a telephone number, or (6) a postal address for the company; (7) the presence of a privacy seal (TRUSTe), (8) the presence of credit-card symbols (Visa, Mastercard, America Express, and Discover), and (9–12) four distinct types of privacy policies.

Given these 12 factors, there were 66 possible experimental conditions. Subjects were asked to indicate their preference in eight randomly selected scenarios. This resulted in a total of 1521 responses, or an average of 23 observations for each cell.

In addition to tracking selections, this part of the experiment also tracked subjects' use of policies in their decision-making. When a policy was present in one or both pages, we tracked whether the subject opened the policy page. We cannot, however, ascertain how much of the policy was read or how carefully.

2.4.1. *Description of e-commerce factors*

For the manipulation of the price we chose to use a fixed amount in order to reduce variability. Subjects in this condition were offered a 20% discount ($5.00) on their purchase.

To indicate the use of third-party cookies and P3P we used the Internet Explorer icon for blocked third-party cookies, which is placed in the lower right-hand corner of the browser (see Fig. 2a). While this is a feature unique to Internet Explorer, more than 90% of survey participants used this browser, and therefore were assumed to be familiar with the icon. This is of course a negative factor, the presence of this icon means the website attempted to do something which many users would be opposed to, so the dependent variable was defined as the absence of this icon.

Secure communication in the form of SSL was simulated in a similar fashion, the use of the familiar and ubiquitous SSL "lock" (see Fig. 2b). This icon also resides in the lower right-hand corner of the browser, but is common to all browsers, and has been in use for a number of years.

Finally, the four policies used were defined along two axes, whether they addressed issues considered important by users or by companies, and whether they were policies which would have a positive or negative impact on the users' privacy. In this way we derive the following four policies:

- User centered—Good.
- User centered—Bad.

Fig. 2. (a& b) Cookie blocked (P3P) and SSL encryption icons used in the Microsoft Internet Explorer browser, respectively. The overwhelming majority of study participants used Internet explorer and should be familiar with these icons.

- Company Centered—Good.
- Company Centered—Bad.

The definition of user and company centered policies comes from a survey of user concerns and privacy policy inventory (Earp and Meyer, 2000). In that study the authors found that users were most concerned with notices disclosing information transfer/sharing practices, notification practices, and information storage practices, in order of importance. Corporations, as evident in examinations of privacy policies, are most concerned with disclosing information on security practices (assurances), collection mechanisms, and consent/assent policies, in order of importance. To simplify matters, our policies only addressed the relevant three issues (notification, data collection, and data sharing, or data collection, security, and data use). Furthermore, the "good" policies gave assurances to users on all three issues, and the "bad" policies admitted adverse practices on all three issues.

3. Results

The results of the demographic survey are reported above in the description of the study participants. In the following subsections, we present the results of the attitude survey, the self-assessment of knowledge of, and attitudes toward, technology and the simulated e-commerce scenario.

3.1. Attitudes toward privacy

Attitudes toward privacy were assessed by means of the 5-point Likert-scale attitude survey. "Agree/Strongly agree" and "Disagree/Strongly disagree" responses were pooled in Table 3, so there are three major columns, not five. These are broken down into three figures: the total responses in that category, the female responses, and the male responses. Since some participants did not indicate their gender on the demographic survey, the total may not match the average of the two sub-groups.

Both an independence test ($p<.10$) and an analysis of a logistic regression model ($p<.10$) agree that females are trending towards under-representation in the unconcerned group, but that this does not reach statistically significant levels. This means that women tend to report higher levels of concern than the men in the five key questions used to define the different user categories.

Independence tests suggest that males and females rate their concerns about privacy (general), threats posed by cookies, and their predisposition to rechecking policies differently ($p<.05$). The logistic regression models suggest that females tend to score these higher, but this trend is not statistically significant.

We also note that the level of concern for online identity theft and credit-card fraud is marginally higher than concerns for offline identity theft and credit-card fraud. This goes against expectations, as more subjects reported having experienced problems offline than online. This was consistent with recent findings indicating that most identity theft occurs offline (Synovate, 2003).

Table 3
Participant privacy attitudes and concerns

	Agree			Neutral			Disagree		
	All (%)	F (%)	M (%)	All (%)	**F (%)**	M (%)	All (%)	F (%)	M (%)
I am concerned about online identity theft*	61.3	79.0	55.9	20.4	10.5	22.1	18.3	10.5	22.1
I am concerned about online credit card fraud	66.7	84.2	60.3	16.1	5.3	19.1	17.2	10.5	20.6
I am concerned about my privacy online*	72.0	89.5	69.1	15.1	0.0	17.7	12.9	10.5	13.2
I am concerned about my privacy in everyday life*	59.1	73.7	52.9	23.7	15.8	26.5	17.2	10.5	20.6
I am likely to read the privacy policy of a site I visit for the first time	23.7	47.4	17.7	15.1	21.1	14.7	61.3	31.6	67.7
I am likely to read the privacy policy of a site which does not ask me for information	7.5	15.8	2.9	6.5	5.3	7.4	86.0	79.0	89.7
I am likely to read the privacy policy of an ecommerce site before buying anything*	43.0	79.0	35.3	25.8	21.1	25.0	31.2	0.0	39.7
I am likely to re-check the privacy policies of sites I frequently visit	7.5	10.5	4.4	9.7	10.5	8.8	82.8	79.0	86.8
What privacy policies say frequently influences my decision whether to visit or use a websites	19.4	26.3	16.2	37.6	31.6	36.8	43.0	42.1	47.1
Privacy policies accurately reflect what companies do*	16.1	15.8	14.7	50.5	52.6	50.0	33.3	31.6	35.3
Privacy policies are easy to find	36.6	21.1	38.2	35.5	52.6	32.4	28.0	26.3	29.4
It is important to me that websites publish privacy policies	68.8	63.2	69.1	19.4	31.6	17.7	11.8	5.3	13.2

Response rates to privacy attitudes survey items. Questions used to map participants to the three Westin categories are marked with a "*".

Looking at the three Westin-classes of users, we do find a high internal consistency in the answers across this section. Pragmatists rate their concern about online credit-card fraud, online identity theft, and privacy in everyday life significantly lower than Fundamentalists ($p < .05$ for all). These users also differed in their ratings of how much policies influence their decisions, how trustworthy policies are, whether policies are read on the first visit to a site, whether to check the policy when buying something online, and the likelihood that policies will be re-checked (again significantly lower than Fundamentalists ($p < .05$ for all)). Unconcerned users rate the same questions significantly lower than Pragmatists ($p < .03$ for all).

3.2. Knowledge of, and attitudes toward, privacy-relevant technology

A number of differences emerged when participants were asked if they knew about certain privacy-relevant technologies and then asked a follow-up question to probe their knowledge. The results are summarized in Table 4 as "claim" and "demonstrate" knowledge for the three technologies in question: P3P, cookies, and web-bugs. False report shows the percentage of subjects who claimed knowledge but failed to demonstrate it. The demonstrate row shows what proportion of the total population actually proved knowledgeable about these technologies. Thus, of the 21.5% who claimed to know P3P; only 25.0% could answer the probe question correctly, or 5.4% of all participants. It is important to remember that our subjects were more highly educated about computers and privacy than the average user, and that we set our threshold for knowledge pretty low. These numbers are therefore likely upper-bounds.

According to this survey, claiming knowledge about a technology does not mean much. Across the board, less than a quarter of participants who claimed to know a technology could answer simple questions about it. For P3P and Web-bugs, we find that on the whole, only 5–6% of subjects actually understand these technologies.

Only in the case of cookies do we see the majority of subjects (over 90%) claiming knowledge. Though significantly more people know about cookies than the other two technologies, the disparity between claimed knowledge and proven knowledge is

Table 4
Key technology familiarity

	P3P			Cookies			Web-bugs		
	All (%)	F (%)	M (%)	All (%)	F (%)	M (%)	All (%)	F (%)	M (%)
Claim knowledge	21.5	21.7	23.4	90.3	95.7	89.1	34.8	34.8	36.5
False report (of those who claim)	75.0	80.0	73.3	84.5	90.9	80.7	82.8	75.0	84.0
Demonstrate knowledge (overall)	5.4	4.3	6.3	14.0	8.7	17.2	5.4	8.7	5.9

Percentage of survey population to claim to understand technologies, miss-judge their understanding, and percentage of knowledgeable participants over-all.

actually larger than in the other cases, especially for women, who perform poorly on these questions. In this light, it may be argued that the lower computer experience scores reported by women might reflect accurate self-assessments, and not differences in confidence. The interesting exception to this is that more women seem to know what web-bugs are then men do.

Of the technologies examined here, cookies have, by far, received the most publicity, something evident by the very high recognition rate. These results show that, the vast majority of users do not have any real knowledge why or how cookies pose a risk to them. Despite this lack of knowledge, participants registered moderate to high levels of support for the adoption of these technologies, or in the case of cookies, the ability to control their use.

In terms of risk or benefit evaluations (Table 5), P3P and cookies were viewed as moderate risks, whereas web-bugs were viewed almost unanimously as a high-risk to personal privacy. Throughout this section we find that there are no statistically significant gender differences (the number of women who demonstrated knowledge of cookies is so small that statistical tests are inconclusive).

The only difference found in terms of the Westin equivalence was that the Unconcerned rated their concern about web-bugs significantly lower than Fundamentalists, ($p < .005$). This means that while privacy Fundamentalists are no more knowledgeable than Pragmatists or the Unconcerned, they do worry more about the risks. It is therefore possible that this segmentation is not so much based on the subjects' knowledge of risks, but on other risk estimates and sensitivities.

Table 5
Technology and risk perception

		Yes/agree			Don't know/ neutral			No/disagree		
		All (%)	F (%)	M (%)	All (%)	F (%)	M (%)	All (%)	F (%)	M (%)
P3P	It is important to me that sites adopt p3p policies	66.7	75.0	73.3	23.8	25.0	13.3	9.5	0.0	13.3
	p3p can help protect my privacy	47.6	50.0	46.7	47.6	50.0	46.7	4.8	0.0	6.7
Cookies	It is important to me to know about and control the use of cookies	72.6	73.3	73.4	17.9	26.7	14.1	9.5	0.0	12.5
	Cookies are a threat to my privacy	45.2	60.0	42.2	35.7	20.0	42.2	19.1	20.0	15.6
Web-bugs	It is important for me to know about and control the use of web-bugs	71.9	100.0	68.0	18.8	0.0	20.0	9.4	0.0	12.0
	Web-bugs present a threat to my privacy	71.9	75.0	72.0	18.8	25.0	16.0	9.4	0.0	12.0

Rate of survey participants expressing interest in and concern about key privacy technologies.

The Westin index is often used in marketing and deployment decisions with regards to privacy tools. It is assumed that the Fundamentalists are the drivers of this market, while the other two categories of users are largely uninterested. We found that the only statistically significant relationship between those claiming to have downloaded and installed countermeasures and the Westin index was that Unconcerned users reported to have done so to a less extend ($p < .05$). There were no significant differences between Fundamentalists and Pragmatists.

3.3. A further look at demographics and experience

There were some significant relationships between survey items. There was a significant correlation between self-reported frequency of online purchases and the maximum purchase amount (those reporting more frequent purchases also reported spending more money ($p < .0001$). This makes sense, as both numbers say something about their comfort and confidence with e-commerce.

We also found a correlation between the maximum online purchase amount and the reported frequency of victimization in credit card fraud or identity theft, online or offline. Victims tend to have spent more money online than non-victims (hundreds of dollars versus tens) ($p < .05$). This also makes sense, those with the highest online purchase amounts were also those with the most frequent transactions, and therefore the highest level of exposure. In both cases, differences emerge when comparing those who are moderately frequent buyers with those who rarely buy, and between those who have made moderately large purchases with those who have only made small purchases. Very frequent purchases and the purchase of very big-ticket items do not appear to be associated with victimization, possibly because we have few subjects in these categories.

Having stratified the participants according to our approximations to the Westin categories, it was possible to investigate whether their attitudes toward privacy as indicated by Westin category was associated with online experience or expertise. This was not the case, as shown by Table 6. Note that, as in most of the tables in this paper, not all subjects were classified as Fundamentalists, Pragmatists or

Table 6
Westin equivalence and e-commerce

	Frequency of purchase					Maximum purchase amount			
	Never (%)	Less than month (%)	Monthly (%)	More than monthly (%)	Weekly (%)	N/A (%)	Tens (%)	Hundreds (%)	Thousands (%)
Fundamentalist	6.3	43.8	21.9	18.8	9.4	6.3	18.8	40.6	34.4
Pragmatist	0.0	42.9	33.3	14.3	9.5	0.0	28.6	38.1	33.3
Unconcerned	2.7	54.1	21.6	13.5	8.1	0.0	13.5	59.5	27.0
All	8.6	43.4	28.0	13.1	6.3	6.9	18.3	48.0	26.3

Mapping of e-commerce activity to Westin equivalence.

Unconcerned, therefore the statistics for the total sample in the study may not match the sums of the subgroups.

There is a rather puzzling lack of association between Westin category and subjects reports of victimization (fraud or identity theft). One would expect victims to be more concerned about privacy and therefore be classified as Fundamentalists or Pragmatists. Again, this lack of correlation may be caused by the low number of victims in this survey, or it may be the case that victims accept this as the risk of doing business online. What we do find is that women tend to be underrepresented in the Unconcerned group, though this not statistically significant ($p < .10$).

3.4. Inspection of privacy policies

In the simulated e-commerce scenarios, 97% of the trials made at least one privacy policy accessible. Thus in almost all trials, a participant could check a privacy policy if he or she wanted to. In fact, Table 7 reveals that policies were only consulted in 25.9% of cases where a policy was available. This number is similar to the rates at which subjects self-reported they would (23.7%) in the general case, but much lower than the 43.0% reported for e-commerce scenarios. Over half of the participants (58.2%) consulted at least one policy. The mean number of policy look-ups for these participants was 3.18 in eight trials.

The likelihood that a participant would consult at least one policy was unrelated to the participant's Westin category and gender. In other words, women and Fundamentalists are no more likely to read policies than men or the privacy Unconcerned, respectively.

We believe that the policy consultation numbers are inflated because subjects knew they were being observed, and what the purpose of the experiment was. They therefore likely took more care and were more thorough in their decision-making process than they normally would.

3.5. Factors influencing simulated purchases

The data from the e-commerce experiment were analysed using a binary logistic regression technique in which a best-fitting regression model was constructed for a subset of independent variables measuring the differences between the two

Table 7
Policy consultation rate

	Policy consulted (where available) (%)
Fundamentalist	29.7
Pragmatist	26.2
Unconcerned	26.8
All	25.9

Percentage of scenarios or trials where the subject consulted a privacy policy, by Westin equivalence.

e-commerce web-pages compared. In this section we present several models for different sub-groups we define.

A binary logistic model assigns coefficients to the independent variables so as to match the behavior of the dependent variable as closely as possible across all observations. The coefficients in this model can be used to determine the estimated relative importance of each factor in the decision-making process. The higher the coefficient assigned to a factor, the more it influenced the decision-making.

One of the twelve independent variables, the use of 3rd party cookies and P3P, was found not to be significant in any model, and is therefore dropped from further consideration. There are several potential explanations for why this variable proved to be insignificant across the board. It could be that subjects are not concerned about the use of 3rd party cookies and P3P in websites. In the challenge section of this survey, this was the technology users expressed least concern with. It is also possible that subjects were not familiar with the indictor used, that it brought about the wrong associations, or that it was simply not sufficiently visible.

The best-fit model including all independent variables as factors (except 3rd party cookies and P3P) leads to a fit of 8.4% (Table 8). This means that this model accounts for 8.4% of the variance in the sample. While this is not a great fit for such a model, it is not unexpected given the large number of factors which we do not control for in this experiment, and the natural variance in peoples decision-making strategies and sensitivities with regard to privacy.

The tables in this section are normalized so the coefficient of the most important variable in each model is set to 100% and the others drop according to their relative coefficient. For instance, in this case, the inclusion of a contact email contributes 56.7% less to the users' decision than the inclusion of a TRUSTe seal. The Logistic

Table 8
Best fit model with all factors

Variable	Contribution (%)	Rank	Probability
TRUSTe	100.0	1	$p<0.0001$
Policy-User-Good	93.5	2	$p<0.0001$
Policy-Corp-Good	86.2	3	$p<0.0005$
Policy-Corp-Bad	74.7	4	$p<0.0001$
Policy-User-Bad	55.4	8	$p<0.0001$
Contact Phone	74.6	5	$p<0.0001$
Contact Address	69.5	6	$p<0.0001$
Price Cut	62.3	7	$p<0.0001$
Credit Card	50.9	9	$p<0.0001$
SSL	48.8	10	$p<0.0005$
Contact Email	43.3	11	$p<0.001$
McFadden R^2	8.4		

Ranking of experimental factors by order of contribution towards explaining user actions. Contribution measured as percentage of most influential factor. Note that ''Policy-User-Bad'' is presented out of sequence to illustrate relative importance of policies.

regression model produces actual numbers for the coefficients, but one cannot use these to compare between models as with the percentages.

Note that Policy-User-Bad is placed out of order from the other factors (to group it with the other policy options). The order in terms of significance of the policies was as expected, users preferred policies which addressed their concerns, or in the negative case, did not confirm their fears (Policy-User-Bad ranks significantly lower than Policy-Corp-Bad). It is interesting to note what strong effect policies have despite being inspected in only a quarter of the cases. This indicates that in many cases it is that the presence of a policy has a positive effect on users, not its content.

To get a better sense of what is going on, and to study different decision-making priorities and processes, we divided the sample into relevant sub-groups. The sub-groups examined will be men versus women, the three Westin categories of users, and the way subjects use privacy policies.

The contribution of each factor to the model in each sub-sample is given in Tables 9–11. Empty cells represent cases where the factor did not prove to be statistically significant in the decision-making model. Note that values cannot be directly compared across columns, only their relative values and ranks.

When looking at the differences between men and women's decision-making (Table 9), we were not surprised to see that men followed the pattern of the overall population (with some local swapping of factors). Men constituted 74% of the survey sample; it was therefore natural that they greatly influence the global model.

The model for women proved interesting because it offered a far better fit than the global or male logistic regression model (16.1% versus 8.4% globally). Women also eliminated two factors from their model, "Contact Email" and "Credit Card". Furthermore, women exhibited clustering behavior in terms of the ranking of factors. TRUSTe was a very influential factor, with the next most influential factor, "Policy-Corp-Good", contributing 27.3% less to the decision. At the bottom of this

Table 9
Decision-making by gender

Variable	All (%)	Rank	Men (%)	Rank	Women (%)	Rank
TRUSTe	100.0	1	91.0	2	100.0	1
Policy-User-Good	93.5	2	100.0	1	63.3	3
Policy-Corp-Good	86.2	3	83.8	3	72.7	2
Policy-Corp-Bad	74.7	4	80.2	4	41.5	5
Policy-User-Bad	55.4	8	49.2	10	36.8	8
Contact Phone	74.6	5	76.2	5	40.7	6
Contact Address	69.5	6	73.2	6	29.4	9
Price Cut	62.3	7	61.4	8	55.8	4
Credit Card	50.9	9	64.9	7		
SSL	48.8	10	46.6	11	37.5	7
Contact Email	43.3	11	54.9	9		
McFadden R^2	8.4		7.8		16.1	

Decision-making and relative importance of factors by gender.

Table 10
Decision-making by Westin equivalence

Variable	All (%)	Rank	Pragmatists (%)	Rank	Unconcerned (%)	Rank
TRUSTe	100.0	1	100.0	1		
Policy-User-Good	93.5	2	83.6	2	100.0	1
Policy-Corp-Good	86.2	3			47.5	2
Policy-Corp-Bad	74.7	4	69.1	5		
Policy-User-Bad	55.4	8				
Contact Phone	74.6	5	71.4	4		
Contact Address	69.5	6	77.7	3		
Price Cut	62.3	7	49.6	9		
Credit Card	50.9	9	63.1	7		
SSL	48.8	10	69.0	6		
Contact Email	43.3	11	51.8	8		
McFadden R^2	8.4		13.9		12.7	

Decision-making and relative importance of factors by Westin Equivalence. Note that no model emerged for the Fundamentalists.

scale we see a tight cluster of factors, all within 5 percentage points of each other (rank 8–5).

The next way to divide up and examine the population was according to the Westin privacy classifications (Table 10). We present two models, one for the Pragmatists and one for the Unconcerned. None of the variables were statistically significant for the Fundamentalists, and no model could be found.

The simpler of the two models was that for the Unconcerned, consisting of only two variables, yet accounting for 12.7% of the variability of the sample. This is significantly better than what we accomplish for the general population. What we find is that the presence or absence of the ''Policy-User-Good'', and to a lesser extent the ''Policy-Corp-Good'', determined the users' choice. For the Pragmatists, the model was much more complex, including 9 of the 11 factors, and apart from a marked preference for TRUSTe indicators, there were no dramatic jumps in the weight of one factor to its nearest neighbor.

The final way of dividing the sample which we examined in this paper was according to the users' relationship with, and use of, the privacy policies (Table 11). We have seen clear indications that privacy policies greatly influence users' choice in most models, yet we know that almost half of the users never looked at a policy, and that policies were only consulted in a quarter of the trials. We therefore examined the sample two different ways. First we compare the trials in which the user checked a policy with the ones where no policy-check was conducted (*Per-Trial Policy Behavior*). Then we divided the sample based on user behavior, the group of users who checked a policy at least once were compared to the group of users who never checked policy (*Per-User Policy Behavior*). The first resulted in a roughly 25–75% split of the sample, while the second resulted in a roughly 50–50% division.

Here we see how strongly policies influence the decision-making process. In the model of trials where users consulted the policy, it accounted for 22.6% of the

Table 11
Decision-making by policy-related behavior

Variable	General case		Per-trial policy behavior				Per-user policy behavior			
	All (%)	Rank	Checked (%)	Rank	Unchecked (%)	Rank	Some checks (%)	Rank	Never checks (%)	Rank
TRUSTe	100.0	1	54.9	3	96.2	2	91.3	3	100.0	1
Policy-User-Good	93.5	2	100.0	1	73.6	6	100.0	1	48.0	8
Policy-Corp-Good	86.2	3	83.8	2	71.8	7	91.5	2	49.6	7
Policy-Corp-Bad	74.7	4	43.3	4	82.4	3	66.9	4	76.2	2
Policy-User-Bad	55.4	8	38.0	5	55.3	11	59.8	6		
Contact Phone	74.6	5			100.0	1	64.2	5	73.8	3
Contact Address	69.5	6			78.3	4	56.8	8	73.2	4
Price Cut	62.3	7			57.2	9	59.8	6	45.9	9
Credit Card	50.9	9			78.1	5	35.1	10	69.6	5
SSL	48.8	10			58.2	8	51.2	9		
Contact Email	43.3	11			57.2	9	30.7	11	49.6	6
McFadden R^2	8.4		22.6		6.9		12.4		7.2	

Decision-making and relative importance of factors by per-user and per-experiment policy-reading behavior.

variance in the sample. Furthermore, we saw clear evidence of discrimination between the good and the bad policies, with users showing a strong preference for the good. Interestingly, the only non-policy factor to remain in this model was the TRUSTe seal, coming in between the good and the bad policies. This shows the strength of positive associations users have with these types of trust-marks.

When we looked at the unchecked policy model, we saw that policies are still important in the decision-making, though the ordering of the good and bad policies is arbitrary since they were not actually read. Interestingly enough, this is the only model which rated the presence of "Contact Phone" as the most significant factor. What we saw was clear evidence of how people used these factors to determine "trustworthiness", not based on fact but rather on appearance and first impression. Policies are important, not just because of what they say, but because they are there. As we saw, this model offered a far worse fit for the data, demonstrating the finality that policy checking brings to decision-making.

When we look at the second category we see similar behavior. For those who never check policies we find that TRUSTe dominates over the other factors, and that policies, though never read, have a fairly powerful effect. Policy-checkers naturally closely match the behavior of the policy checked group, as they on average checked policies in half the trials.

4. Discussion of results

In this section, we discuss the results and possible threats to their validity. Broader implications are discussed in the next section.

4.1. Privacy classification

The Westin privacy segmentation is a way of dividing and thinking about privacy sensitivities which has been widely adopted and embraced by industry and academia. In our survey we did not use the same questions as Westin, but we were able to find very similar, very cohesive groups, especially in the privacy questionnaire. It is therefore likely that we identified the same groups as the Westin surveys have identified in the past.

We think that it is interesting and important to demonstrate that these groups are significant and identifiable outside the context of the three traditionally posed questions, and that they employ very different decision-making strategies. It is also interesting to see where these groups started to lose their significance, especially in the analysis of experiment. One of the most surprising findings was that we were unable to find a logistic regression model for the Fundamentalist group, where none of the twelve independent variables were significant.

There are several potential explanations for the lack of a model for the Fundamentalists. This group may itself be a collection of very different sub-groups, all highly concerned about personal privacy, but with very different decision-making strategies. Or perhaps the Fundamentalists are not really influenced by any of the

factors we included in this study. This would mean that this group, though as likely as any other to read privacy policies, is not influenced by them, or does not trust what policies say. It is also possible that Fundamentalists are looking for different types of information in policies than the rest of the population.

Another interesting finding with regards to these three user groups is that Fundamentalists are no more likely to install or use software to protect themselves than the Pragmatist group, though they are different from the Unconcerned. This means that the Pragmatists should not be ignored as consumers or early adopters of privacy enhancing technologies.

4.2. Gender

There are interesting indications throughout the study of gender differences. Women were underrepresented in the study and therefore some gender-specific questions could not be answered.

The Westin classification may confound gender, since a smaller proportion of women were classified as Unconcerned than were men, a result that approached statistical significance. We do not know whether this means that women tend to be more risk-averse, more pessimistic or skeptical about the motives and honesty of online vendors, less knowledgeable about the technology in question, or more knowledgeable about the risks of online transactions (e.g. having suffered more from the consequences of identity theft or fraud or being familiar with the fate of friends and associates who have been). Our study design and the number of people falling in some of these categories (particularly victims of fraud or identity theft) are such that we cannot investigate the reasons further.

The stratified model for women was by far the best fitting model for the experimental scenario. Although the presence of the TRUSTe seal was one of the most significant variables in all models, it is noteworthy that for women, the presence or absence of the seal was a much more significant factor than any other.

4.3. Indicators

Third-party cookies and P3P was the only indicator not to prove significant in any model. There are several potential explanations for this. It could be that subjects are not concerned about the use of third-party cookies, or are not familiar enough with P3P to make decisions based on this indicator. In the challenge section of this survey, cookies were the technology users expressed least concern with. However, is also possible that subjects were not familiar with the indictor used on the web page to signal the presence of third-party cookies. We did not test subjects' knowledge of indicators. However it is consistent with this interpretation that cookies were the technology for which subjects' self-reported knowledge diverged most from their ability to answer the challenge question.

Perhaps users were unable to process information about cookies coherently. This is an unlikely explanation. A more likely one, given the low effect of the SSL

encryption indicator is that these icons are too inconspicuous, and that many users do not notice or pay attention to these indicators.

4.4. Policies

The order of significance of the policy variables in the regression analysis was as expected: users preferred policies which addressed their concerns rather than the company's; in negative cases, they preferred policies that did not confirm their fears (that is Policy-User-Bad ranks significantly lower than Policy-Corp-Bad).

It is interesting to note that the presence of a visible link to a privacy policy has a major effect on purchasing behavior, even though only a quarter of the policies were consulted. In most cases, users had more confidence in a site simply because it had a policy.

It was the Unconcerned users who were most influenced by the content of the policies. The picture that emerges here is of users who take a more casual approach to the evaluation of privacy risks, yet are strongly swayed by the assurances made in policies. Since one of the questions used to categorize users referred to the trustworthiness of policies, it is not surprising that the Unconcerned were more affected by the policies alone than other, potentially more skeptical users.

5. Implications

We conclude with a discussion of the implications of our results in the following four areas:

- The design of user interface indicators so that users understand and may act on privacy-relevant information.
- The wisdom of classifying users into categories along a single dimension of privacy concern.
- Implications for research methodology of the contrast between the results obtained from self-reports and those obtained through experimental economic scenarios.
- Public policy implications, such as the regulation and legislation of how and when users must be notified of privacy practices and policies, together with limits to the notion of informed consent.

5.1. Design

While this work provides important guidance for business, policymakers, and management, it also provides important insights for interface designers. In our experiment there is a set of variables which we can call "trust-marks", factors which may not say anything about the site's privacy practices, but which are interpreted as such by users.

One such factor, the TRUSTe marker, very important in most models, actually says nothing about the practices of the site. This marker serves as a guarantee that the site discloses a minimum set of information in its policy, and that it follows the practices it claims rather than what the policy says about these practices. The TRUSTe marker should have been a powerful indicator, but only in conjunction with a privacy policy.

Privacy policies themselves serve as "trust-marks", evident we think from the impact they have on users who never consult them. What we are seeing is that the presence of a policy has a significant effect on decision-making regardless of whether the policy was read or not. The impact a policy has is of course more powerful when it is read, but it is not negligible when it is not.

Other factors which can be classified as "trust-marks" are the credit card icons and the contact information variables. The credit card icons are interesting because they do not in fact imply any promise of fraud prevention or privacy protection. Just about every e-commerce site accepts some form of credit card payment (some operate on electronic payment systems such as Paypal), and it is therefore not clear why consumers should find these icons reassuring.

As to the inclusion of contact information, there was a strong preference for phone information over mailing or email information. This means that users were looking for ways of holding companies accountable, or indicators for a company's willingness to dealing directly with them should they have any problems as a result of this transaction. One interesting question is whether users would actually test to see if the phone number was valid before buying from a site. In these tests, the phone number was plainly invalid.

The impact these "trust-marks" have on decision-making, across all user groups, shows a clear need for designers to develop privacy-enhancing technologies which give users simple and clearly visible trust indicators. If these markers are not clearly visible they may be ignored, as we saw with the SSL encryption icon and Cookie-blocking and P3P. These indicators may be too inconspicuous for users to notice. While it is possible that users do not place a lot of value in these factors, we believe that the reason that relatively meaningless indicators, such as the credit-card icons, are preferred is because they are more clearly visible.

What we found in this study, like others, was that only a minority of subjects read policies with any frequency. The information contained in these policies is considered highly significant, and highly influential in users' decision-making, but is rarely sought out. In this experiment, where the rate of policy consultation was likely inflated by subjects knowing the purpose of the experiment, we found that subjects only consulted policies in a quarter of trials. Other studies have shown this rate to be much lower in real life, by as much as a factor of ten (Jensen and Potts, 2004).

These findings all argue for the development of policy simplifications, standardization, or machine readable policies. Based on this data we can also make a strong case for the need to develop and implement standardized, simple visual indicators for the practices and technologies websites use, and the risks users are exposed to.

5.2. User classification

When we ran the logistic regression analysis on the Fundamentalist sample we were surprised to find that none of the variables were statistically significant, and that no model could be found. This is especially surprising because the Fundamentalists were the second largest user group with 34% of the sample population, surpassed by the Pragmatists (43%) and outranking the Unconcerned (23%), both of whom provided models accounting for over 12% of the variance. This lack of a model for the Fundamentalists is interesting, because the Westin categories are used in marketing particularly to isolate those consumers who are most likely to embrace privacy-protecting products and services. Our results, however, indicate that while Fundamentalists have a consistent concern with privacy, they do not form a cohesive group with respect to decision-making. Only the Unconcerned and Pragmatists are internally cohesive groups.

Colloquially speaking, it seems that while it may be more difficult to push the other groups' buttons, they do at least have some; the Fundamentalists, in contrast, don't seem to have a single set of buttons to push. Perhaps market researchers should turn their attention to how the concerns of the less concerned groups can be mobilized rather than concentrating on the more diverse concerns of the Fundamentalists.

The main alternative explanation for the lack of a model for the Fundamentalists, as discussed in the previous section, is that Fundamentalists were not influenced by the factors we used in this experiment. This seems an ad hoc explanation, however, given that the other groups behaved as expected and that there were twelve plausibly variables under investigation. Also worth noting is the finding from our survey that Fundamentalists were no more likely than others to install privacy-protecting technology. The most parsimonious explanation appears to be that Fundamentalists are not really that "fundamentalist" about privacy at all.

5.3. Research methodology

In general, the study demonstrates that users do not do what they say, and they do not know what they claim to know. Although the subjects of this survey consulted online privacy policies more often than previous log-based studies indicate users "in the wild" do, their behavior did not match their survey statements. Subjects were also generally not able to answer questions about privacy-related technology that they claimed to know, a trend particularly noticeable in the case of cookies, where they reported the highest knowledge.

Such results call for a reevaluation of the role of surveys in the study of Internet behavior. Surveys appear to be best suited to the evaluation of attitudes and opinions rather than behaviors or experience. Where issues arise, such as the role of perceived competence in decisions-making, the use of self-reports is invaluable as a baseline against which actual behavior can be compared. The self-reported data should not be taken as evidence of behavior, however. Indeed, the shakiness of our subjects' self-reports and judgment of knowledge leads us to wonder whether their

experience reports should be taken seriously. How many years experience a user has and how intense the usage during that period may have been may be very different from what the user reports in a survey. In addition to demand characteristics of the survey situation (albeit in a situation of personal anonymity) simple forgetting and salience effects are likely to skew the user's recall and categorization of his or her experience.

In place of a full reliance on survey data, we see the need for more concerted attempts in the area of experimental economics in which users are put in realistic situations and required to indicate preferences or make economic decisions such as bidding in auctions or purchasing items with simulated funds. To capture the effect on decision-making of the context of previous decisions and of current affairs (e.g. news about technology vulnerabilities or protections) such studies should ideally be longitudinal rather than one-shot experiments. Such a shift in methodology would likely increase the ecological validity of research instruments and settings but would require a wholesale shift in how we plan studies, recruit subjects, and standardize instruments. As an example of standardization, consider the role played in the current study of the various online policies. Research into the effect of policy content on online behavior clearly requires that the policies used in different studies be systematically comparable if not identical.

The experiment reported here was not intended to be a full-fledged study in experimental economics. We did not investigate systematically the trade-off between privacy indicators and a range of price points or product attributes for the items for sale. Instead, we presented a 20% difference in price of an identical item available from two vendors. Different price differentials, or tradeoffs between price and product quality could interact with privacy awareness in complex ways. Nor did we vary the items for sale in a way that would assess the sensitivity of the models to the nature of the product. (It is unlikely, for example, that privacy indicators would have the same impact relative to price when the consequences of disclosure are more sensitive than model cars—such as pharmaceuticals or erotica. Although when users would behave more cautiously and when more recklessly and how these behaviors might interact with demographic and personality variables is hard to predict.)

In connection with the distinction between professed and actual knowledge, it would be interesting to know whether knowledge really is power and whether a little knowledge is a dangerous thing. Are users who have more knowledge about the privacy implications of Internet technology better able to make more effective discriminations among web sites and services? How does user knowledge relate to user confidence in making online purchasing decisions, and does a little knowledge (for example, of cookies in the case of our study) lead to overconfidence and/or reduced effectiveness? Do general demographic factors, such as amount of online experience, education, age and gender play a role in modulating the answers to these two questions? Unfortunately, our study design does not permit these analyses. It does suggest that they should be fascinating questions to answer in future research.

5.4. Public policy

Public trust in technology rests on the policies that regulate technology and on the doctrine of informed consent. This is not the forum to discuss the broader issues of policymaking and regulation in any depth. However, the notion of agreement and informed consent obviously relies heavily on user's understanding of both technology and the consequences of online behavior. It appears from the results presented above that many users have an incorrect understanding of their own knowledge of technology, their online behavior, and potential consequences. Not only do users frequently fail to consult online privacy policies, when they do, the policy may not help them make informed decisions. Recent studies have found that online privacy policies are difficult to find and demand levels of reading skill to understand that are not typical of the Internet population (Jensen and Potts, 2004; Antón et al., 2004) In view of recent court rulings in the US insurance industry stating that policies must be worded in plain language in order to be enforceable as contracts, the obscurity of privacy policies may call their validity into question.

To consult policies regularly would be very inefficient and dysfunctional unless the likely consequences of not doing so were punitive. Users therefore seem to adopt a strategy of sporadic checking, possibly triggered by the presence of suspicious indicators, in conjunction with the heavy use of proxies or surrogates.

The most significant of these proxies are trust markers. For these to serve as surrogates for detailed inspection of policies, trust marks need to be quality indicators, and not merely presence indicators. In the case of the TRUSTe mark, users appear to take its presence as evidence of the quality of the privacy policy, not merely that the vendor has a privacy policy and follows it.

Trust marks that are presence indicators but not quality indicators do not encourage deception by vendors, but they do make it possible. Unscrupulous vendors could use such marks as camouflage for policies and practices that users would not willingly agree to.

The results of this study show that even self-selected volunteers in a survey on online privacy, who are therefore likely predisposed to think about privacy issues, and who know that their online behavior is being monitored, still show remarkable ignorance and inappropriately placed trust in their actions. To avoid exploitation and consequent reduction of that trust, greater public awareness of privacy issues, the capabilities and limitations of privacy-enhancing technologies and the significance of policies and trust indicators are all necessary.

Acknowledgments

This research was supported by NSF ITR Grant 0325269. The authors thank TRUSTe for permission to use their mark in these experiments, and Roger Los of http://www.los.com/ and http://austinhealey.com/ for permission to use his artwork. We also thank Ji Han Bae and William Stufflebeam for their help in designing and running this survey.

References

Adkinson, W.F., Eisenach, J.A., Lenard, T.M., 2002. Privacy online: a report on the information practices and policies of commercial web sites. Progress and Freedom Foundation, Washington DC. Online: http://www.pff.org/issues-pubs/books/020301privacyonlinereport.pdf.

Antón, A.I., Earp, J.B., Bolchini, D., He, Q., Jensen, C., Stufflebeam, W., 2004. The lack of clarity in financial privacy policies and the need for standardization. IEEE Security & Privacy 2 (2), 36–45.

Cranor, L.F., Reagle, J., Ackerman, M.S., 1999. Beyond concern: understanding net users' attitudes about online privacy, AT&T Labs-Research Technical Report TR 99.4.3, http://www.research.att.com/library/trs/TRs/99/99.4/99.43/report.htm.

Culnan, M.J., 1999. Georgetown Internet Privacy Policy Survey: Report to the Federal Trade Commission. Georgetown University, McDonough School of Business, Washington, DC Online: http://www.msb.georgetown.edu/faculty/culnanm/GIPPS/gipps1.pdf.

Culnan, M.J., Milne, G.R., 2001. The Culnan-Milne survey on consumers & online privacy notices: summary of responses. Proceedings of Get Noticed: Effective Financial Privacy Notices. A Federal Trade Commission Workshop. Washington, DC. Online: http://www.ftc.gov/bcp/workshops/glb/supporting/culnan-milne.pdf.

Earp, J.B., Meyer, G., 2000. Internet consumer behavior: privacy and its impact on internet policy. Proceedings of the TPRC 28th Research Conference on Communication, Information and Internet Policy, Alexandria, VA.

Federal Trade Commission (FTC), 2000. Privacy online: fair information practices in the electronic marketplace: a report to congress. Online: http://www.ftc.gov/reports/privacy2000/privacy2000.pdf.

Harris et al., 1998. E-commerce & privacy: what net users want. Privacy & American business and price water house coopers LLP. Online: http://www.pandab.org/ecommercesurvey.html.

Harris Interactive, 2003. The Harris Poll® #17: most people are 'privacy pragmatists' who, while concerned about privacy, will sometimes trade it off for other benefits. Online: http://www.harrisinteractive.com/harris_poll/index.asp?PID=365.

Javelin Strategy & Research, 2005. 2005 Identity Fraud Survey Report. Online: http://www.javelinstrategy.com/reports/2005IdentityFraudSurveyReport.html.

Jensen, C., Potts, C., 2004. Privacy policies as decision-making tools: a usability evaluation of online privacy notices. Proceedings of CHI'04. Vienna, Austria, pp. 471–478.

Jupiter Research, 2002. Security and privacy data. Presentation to the Federal Trade Commission Consumer Information Security Workshop. Online: http://www.ftc.gov/bcp/workshops/security/020520leathern.pdf.

National Telecommunications and Information Administration (NTIA), 2002. A nation online: how Americans are expanding their use of the Internet. Washington, DC. Online: http://www.ntia.doc.gov/ntiahome/dn/.

Synovate, 2003. Identity theft survey report, prepared for the Federal Trade Commission (FTC). Online: http://www.ftc.gov/os/2003/09/synovatereport.pdf.

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