Online Privacy and Consumer Protection: An Analysis of Portal Privacy Statements

Zizi Papacharissi and Jan Fernback

The advent of information technologies has raised public concern regarding privacy, as documented by the results of several surveys. Although extensive, online privacy statements seldom provide explicit reassurance that consumer information will be kept confidential and will not be exploited. This research examines these privacy statements to determine their overall utility. We evaluate the overall efficacy of privacy statements and focus on the language, format, privacy reassurances, complexity of legal and technical terms, and perceived statement credibility. A content analysis of privacy statements reveals that privacy statements do not always protect customer interests as much as they serve as legal safeguards for the companies involved.

In an 1890 Harvard Law Review article, Louis Brandeis and Samuel Warren had the foresight to argue that individual citizens should be free from having intimate information published by an increasingly powerful press. Basic human dignity, they claimed, gives individuals a “right to be let alone”—a right to privacy (Warren & Brandeis, 1890, p. 220). A century later, that powerful press has taken on new proportions with the growth of new media technologies, and privacy concerns are a pervasive part of public discourse regarding information technology. Accompanying the rise of the information society, the notion of privacy has expanded to encompass rights to control information about ourselves. Individuals have the right to inspect their own tax, medical, and other governmental records and to assume that sensitive personal information is not released by financial institutions, governments, doctors, or other businesses to third parties. But these rights to privacy conflict with the freedom of information that democratic societies need to function properly and that busi-
nesses use to economize their operations. The guarantee of free information allows technologies such as thermal imaging, satellite imagery, global positioning, face recognition software, and biometrics to flourish; however, these technologies are accompanied by public worry about loss of privacy and loss of the ability to control information about ourselves. Accordingly, privacy has many facets: individual privacy regarding the integrity of the body; privacy regarding individual behavior; privacy regarding personal communication; and privacy regarding individual data. All of these areas of privacy are increasingly threatened in the information age (Clarke, 1999).

In this study, we focus on online privacy and investigate how consumer information is protected or exposed by online portal sites. Specifically, we examine privacy statements featured in online portals to determine their efficacy for consumers. Through the increasing sophistication of data mining tools, consumer database creation and management has become a growing, profitable enterprise. Personal data is now a tradable commodity in capitalist societies (Hamelink, 2000), and thus, the free market economy and privacy are inherently at odds with one another. Because digitally stored data can have an indefinite life span, public concern over the ability to control our own information is evident in consumer reluctance to provide personal data to online businesses (Elgesem, 1996; Fox & Lewis, 2001). The information storage and retrieval capabilities of new media technologies can facilitate the collection and exchange of customer information, often without the knowledge or permission of the consumer. Companies frequently assemble databases of extensive consumer information that they use to market to specific target populations. As a result, individuals have become wary of disclosing personal information online (Fox, 2000; Fox & Lewis, 2001). Clarke (1999) argued that those concerns over online privacy reflect larger social concerns over “trust in the information society” (p. 60).

Whether or not consumer anxiety about information gathering is warranted, the online industry has responded to public concern and consumer advocacy efforts with voluntarily posted privacy statements to alleviate those concerns. Although frequently governed by suggested industry guidelines, as specified by TRUSTe or similar industry coalitions, these privacy statements seldom provide explicit reassurance that consumer information will be kept confidential and will not be exploited. Instead, they frequently outline how companies intend to use private customer information so that, in the event of consumer complaints, the companies are absolved of responsibility. Companies such as Microsoft Passport Services are known for exploiting consumer information and were finally pressured into revising their privacy policies and statements following a series of articles originating from Salon.com. Both Yahoo’s and Microsoft’s Hotmail e-mail services reportedly divulged customer information in opposition to their stated privacy policies not to share personally identifiable information (Gillis, 2002). Moreover, these privacy statements are usually placed inconveniently at the bottom of the page and are often tedious, complex, and replete with legal language the average Web user finds difficult to comprehend. Kandra (2001) found that many of the security statements of e-tailers sound reassuring but offer very little protection to the individual consumer. In addition, Web users often find privacy policies difficult to trust (Reagle & Cranor, 1999).
The Pew Internet & American Life Project (Fox, 2000; Fox & Lewis, 2001) reveals that consumer trust is a vital issue for Web users, arguing that, although they are gravely concerned about online privacy violations, Americans still engage in intimate and revealing acts online. Twenty-seven percent of online users are staunch believers in online privacy to the extent that they never willfully provide personal information to Web sites. Fifty-four percent of Internet users find online tracking of personal information to be harmful, and only 27% find tracking to be helpful because it provides personally tailored, user-specific information. Tellingly, 86% of online users favor “opt-in” policies that require Web sites to ask for permission before collecting or using personal data. But many users are not proficient enough with computers to employ the methods available to them to protect their privacy. For example, only 10% of Internet users have set their browsers to reject cookies; 5% use anonymizing software to mask their computer identity; and 24% have provided false personal data (like a fake name) to avoid revealing true information. Similarly, 94% of Internet users want disciplinary action taken against privacy violators (Fox, 2000).

Building on this evidence, we examine privacy statements posted by online companies to determine whether they effectively protect personally identifiable and nonidentifiable data. Even though privacy statements do not have the primary purpose of protecting consumers, they are used by online entities to secure the TRUSTe seal of privacy approval, which effectively communicates a privacy pledge to consumers. Companies frequently explicitly state their commitment to protecting private information in these privacy statements, which sets these statements up as privacy pledges. There is also the danger of assuming that privacy protection becomes a problem solved simply by companies offering explicit reassurances of personal information use. Moreover, the survey cited documents consumer uncertainty about privacy protection. In response to this trend, we examine privacy statements to determine whether they provide visitors with adequate information to decide for themselves if disclosing information to that site is a fair and safe exchange. In doing so, we look first at the prevailing regulatory framework regarding online privacy protection, consider the nature and structure of privacy statements, and finally consult relevant research on the overall efficacy of privacy statements.

**Online Privacy Statements and Regulation**

Privacy statements are a fairly new feature for online companies, although some businesses have always made responsible use of consumer information. They usually detail how the company intends to use the personal information collected from customers. Although numerous consumer privacy bills (including the Online Personal Privacy Act, the Consumer Privacy Protection Act, and the Consumer Internet Privacy Enhancement Act) have been brought before Congress, the United States is the only major trading nation that has not adopted blanket privacy protection legislation, instead opting for piecemeal legislation and private sector data protection measures. These measures are flawed, according to Fausett (2001), who claimed that privacy
policies tend to provide such large loopholes to companies that consumers’ rights to privacy are merely titular. One of the reasons for the trajectory of privacy policy in the United States, said Fausett, has to do with the unfettered nature of e-commerce and the Internet in general. Just as TV executives adopted the voluntary ratings systems to circumvent federal involvement in regulating program content, Web sites, too, have adopted their own privacy policies to ensure a lack of government involvement in regulating consumer privacy (with the exception of the Children’s Online Privacy Protection Act [COPPA]). Self-regulation informs consumers of how information on them is collected and used, ostensibly allowing the consumer to decide whether to do business with an online entity. Supposedly the consumer is empowered to make business/commerce decisions. Fausett pointed out that this scenario is rarely played out. Having been drafted by attorneys, online privacy statements often contain catchall stipulations that allow the online entity a high degree of flexibility regarding its uses of consumer information (Fausett, 2001).

In contrast, European Union (EU) member countries must follow strict and specific regulations that protect consumer privacy in accordance with the Directive on Data Protection of 1998. This privacy directive guarantees individual control over consumer data and insists that foreign trading partners adhere to the same level of equal protection (Lee, 2000). Thus, it prohibits the transmission of personal information from EU member countries to outside countries, including the United States, without adequate privacy protection. Other governments, such as Hong Kong, are writing privacy laws similar to the Directive on Data Protection, but the EU has nonetheless forged contractual agreements with U.S. companies to conduct business despite these differences in privacy policies (Lee, 2000). Such formulas do not tend to protect consumer interests as much as they serve as legal safeguards for the companies involved.

Before the EU’s Directive on Data Protection, the Organization for Economic Co-operation and Development (OECD) produced guidelines in 1980 entitled, “OECD Guidelines on Protection of Privacy and Transborder Flows of Personal Data” (McKenna, 2001). These guidelines inform the EU’s privacy policies and are accepted as foundational principles for safeguarding individual privacy in the information economy. The five recognized principles of fair information practices are (a) notice (what personal information is being collected and how it is used), (b) choice/consent (data collectors should obtain consent for uses of personal info), (c) access (data subjects should be granted access to their own collected info to guarantee its accuracy), (d) security (personal info should be kept secure and accurate), and (e) enforcement (these principles should be enforced). Because so many consumers are unaware of their rights or are unaware of the nature and extent of data collection practices, the weight of control over personal data is leaning toward the e-tailer as opposed to the consumer. McKenna suggested that a reevaluation of U.S. privacy initiatives is needed in order to set the pace for a binding “Global Agreement” (p. 347) on personal information privacy.

Lessig (1999) also argued that part of the problem with data mining and monitoring is that the burden is on the monitored person to establish innocence and/or independ-
ence. As more data are compiled, a consumer’s life becomes an ever-expanding record that can be accessed at any time. Even if the data collected about individuals are not misused, Lessig said, the ordinary uses of these data should concern all citizens. Although the data are used to help businesses market more efficiently, they are also used in insidious ways to manipulate consumers. Lessig claimed that TV advertising is so obvious in its manipulative nature as to be rendered less egregious in its manipulative tendencies. But the Web surfer who is not expecting to be the object of a marketing blitz may be “created” by that blitz itself. Consumers who make purchases based on those direct Internet marketing efforts alter their own behaviors such that they fit into the consumer profile created about them by the retailers themselves. Thus, “the system watches what you do; it fits you into a pattern; the pattern is then fed back to you in the form of options set by the pattern; the options reinforce the pattern; the cycle begins again” (Lessig, 1999, p. 154). Similarly, Huberman (2001) contended that “the capacity to personalize offerings and to study the surfing and shopping patterns of countless individuals can also work in a perverse way to monitor them, expose them, and even blackmail them, if necessary” (p. 98).

Perhaps lending support to Lessig’s (1999) and Huberman’s (2001) contentions, the U.S. Federal Trade Commission (FTC; 1998) found in a landmark report that few Web sites meet the FTC-suggested privacy criteria of notice, access, security, and third-party disclosure. Arguing that industry self-regulatory efforts have not adequately protected consumers, the FTC (2000) found in a follow-up study that only 20% of randomly selected Web sites had implemented its suggested four fair information practice principles of notice (of information collection practices), choice (regarding how personal data is used), access (to one’s own collected information), and security (of collected information). Based on this evidence, the FTC has urged Congress to enact legislation designed to protect consumer privacy online. Few such legislative restrictions on the collection and distribution of private information currently exist in the United States. For example, video rental records are private data (Video Protection Privacy Act of 1988); student loan information is private, as is driver’s license information; the Fair Credit Reporting Act of 1970 polices credit reporting agencies; the Electronic Communications Privacy Act of 1986 protects unauthorized access to e-mail and other electronic communication; and the government is limited in its ability to disseminate personal data such as tax records (Belgum, 1999). The Financial Modernization Act (Gramm–Leach–Bliley Act) of 1999 requires financial institutions to inform customers about their privacy practices, but the law provides limited control to consumers regarding the use and distribution of personal data (Privacy Rights Clearinghouse, 2002). The Act provides minimal privacy protection for consumers while still requiring them to take the initiative to request that their personal information not be shared with third parties. However, there are virtually no legal restraints on corporations regarding the collection and selling of private information. Pressman (2001) detailed a case in which the political Web portal voter.com (which has since failed) planned to sell a list of 170,000 e-mail addresses, including party affiliations. Voter.com’s privacy policy allowed subscriber data to be sold in the event that the
company was acquired. Junkbusters president Jason Catlett claims that privacy policies are ineffectual in protecting privacy because consumers do not have conventional rights (Pressman, 2001). Other failed Web sites, such as Toysmart, are attempting to thwart the ravages of bankruptcy by selling their customer lists despite posted privacy policies that specifically prohibited the company from sharing customer data with third parties (Pressman, 2001).

One attempt at limiting corporate information mining practices which has demonstrated limited success is COPPA. COPPA requires online businesses that target children or have knowledge of collection of information from children under 13 to comply with its regulations (Cannon, 2001). Personal information is protected, including name, physical address, e-mail address, telephone number, social security number, cookies, or other persistent identifiers. Although noncompliance with COPPA has not resulted in many fines, the handling of children's data has improved since the implementation of the Act in April 2000. Some sites have ceased operation or have stopped catering to a child audience because the regulations have proven too arduous to comply with (Cannon, 2001).

Aside from COPPA, regulatory policy in the United States has declared Web operators responsible for the disclosure of information gathering, use, and protection practices. TRUSTe is a nonprofit, privacy-stamp program that grew out of federal efforts to help the Internet industry to establish privacy guidelines. TRUSTe (n.d.) explicitly states that

> our goal with these guidelines is to strike a reasonable balance between consumer privacy rights and expectations and the business need to realize the full value of asset portfolios. In an economy valued by information, customer data is like gold and, as such, deserves enhanced protection.

This statement presents the privacy statement as a medium for protection and is preceded by specific TRUSTe guidelines: that all certified online entities must be truthful, tailor the TRUSTe model privacy disclosure to the practices of their specific companies, be extensive and honest in this disclosure, revise and update the statement frequently, and communicate these practices to the entire online company. Web sites displaying the TRUSTe stamp indicate to users that data gathering and dissemination practices will be disclosed and are guaranteed by a third party (Benassi, 1999). Fausett (2001) hailed sites that use TRUSTe certification. This certification usually ensures that privacy policies are written in plain language and are prominently displayed on a Web site. For unsophisticated Web surfers, however, the TRUSTe guarantee can confuse users who may see the stamp as a guarantee of their privacy rather than a fair disclosure of that site's information retrieval and use policies. For example, Fausett noted that "a policy clearly stating that all private data is immediately turned over to telemarketers and thieves would still get a certification from the major privacy associations, so long as it were true" (p. 16). Compliance with the TRUSTe requirements is voluntary, and sites that do break policy have little retribution to fear (other
than some potential bad press). Although consumers may contact the FTC over online privacy violations, even those involving TRUSTe-certified sites, this has not been an effective strategy for consumer rights. Industry coalitions have formed in response to consumer concerns; the Online Privacy Alliance, a group of 50 online companies dedicated to promoting consumer data security, certifies companies that adhere to its recommended privacy policies (Lee, 2000). The American Association of Advertising Agencies also fosters consumer privacy rights in its recommendations for marketers, and some analysts claims that a few online businesses are restricting the sale of consumer lists, limiting monitoring, and discontinuing spam e-mail (Lee, 2000). More individuals are educating themselves about data collection practices and about mechanisms by which to thwart these practices. Some software solutions exist, however, for the more naive Web surfer. AT&T has developed its Privacy Bird software (www.privacybird.com), which enables individuals to set privacy preferences and determine whether a visited site corresponds with those preferences.

The Efficacy of Privacy Statements

Still, the burden for safeguarding individual privacy rights online is clearly on the consumer, who must read these statements to determine whether the portal presents a safe and fair environment for the exchange of information. Companies are not legally obliged to protect the consumer by means of a privacy statement; they simply have to state how they use personal information. In a study examining the disclosure procedures of privacy practices of various retail Web sites, Miyazaki and Fernandez (2000) found that consumers are more likely to purchase items from Web retailers whose privacy and security statements are present. It also found that 23.1% of retail Web sites disclosed customer identification practices. A later study by Miyazaki and Fernandez (2001) suggested that consumers with high levels of Internet experience perceive a lower risk to online shopping but that they demonstrate greater concern over online privacy. Sheehan and Hoy (2000) surveyed online consumers about their attitudes toward the FTC’s principles for online privacy, finding that consumers tend to trust online entities with whom they have prior relationships, exhibiting diminished privacy concerns. The survey also finds that consumer value of privacy is contextual; that is, certain types of information are more willingly submitted to online retailers, and more valuable information can be had by Web sites willing to provide something in exchange for it (e.g., contest entries). Despite this climate of demonstrated willingness to engage in online transactions, Miyazaki and Fernandez (2001) argued that the continued acquisition of consumer data will ultimately lead to a slowdown of online retail sales due to perceived infringements of privacy.

Some consumers have responded to the perceived invasion of privacy by presenting false information to sites or by using anonymizing or pseudonymizing software. These programs work to strip a user’s identity (www.anonymizer.com or www.PrivacyX.com) or to create fictitious identities (Lucent Personalized Web Assis-
The World Wide Web Consortium (W3C) is developing a Platform for Privacy Preferences Project (P3P), whose goal is to help users employ their own preferences over privacy practices while online (Reagle & Cranor, 1999). These strategies present some philosophical quandaries relating to individual responsibility, profit, and the concept of privacy. Some have suggested that the right to control information about oneself includes not only consenting to the sale of personal information but the profit from this sale (Reilly, 1999). Thus, the concept of privacy must change in accordance with the technological possibilities of the information age. Reilly suggested that policymakers can satisfy both consumers and information gatherers by restricting data collection techniques minimally in order to guarantee the empowering potential of online technologies to gather information and to protect information. But this perspective fails to consider the individual’s right to feel protected. Elgesem (1996) argued that the increasing sophistication of data collection methods represents a growing assumption of risk for consumers: We choose to give up some information about ourselves in exchange for convenience and for ease in achieving our own informational goals. The problem, said Elgesem, is that with the introduction of modern information technology, the processing of personal information becomes more complex and extensive. As a result, there is an increased cost in the form of increased risk of privacy violations. The point of the principles of modern privacy legislation is, in my view, to relieve the individual of this additional cost. (p. 52)

We can maintain our privacy by controlling the flow of information about ourselves, according to Elgesem (1996). This involves public knowledge about the way information is collected, the ability to inspect and correct that information, assurances that information is being used as it is intended to be, and assurances that information collection is justified in terms of the ends for which it will be used. Currently, many situations involving data collection do not provide consumers with the ability to choose whether or not to consent to the risk of revealing data (e.g., medical records). Because consumers must accept the obligatory risks that accompany institutional information practices, it seems unjust for individuals to assume the costs of those risks. Elgesem concluded that fair information management dictates that the costs of relinquishing individual privacy not outweigh the reasons for which the consumer initially chose to give up personal information.

The individual’s agency in privacy concerns has also been examined by Clarke (1999). Clarke argued that privacy initiatives such as the W3C strive to recognize personal data as a form of intellectual property such that consumers are granted intellectual property rights and may license their personal information to whomever they choose. This treatment of privacy concerns as intellectual property amounts to an economic (or capitalistic) rather than a social–ethical solution to a very human concern. However, Caudill and Murphy (2000) suggested that an ethics-based model for privacy regulation would allay some consumer fears about their privacy in the online.
realm. For example, because online privacy is a global issue, not just a U.S.-led concern, international initiatives derived from an ethical foundation should be developed. Specific recommendations include, first, individual consent that recognizes the balanced needs of all parties (consumers, retailers, etc.). This means that individuals should be able to give consent to having their data collected but through a means that is less invasive to individual privacy. The second recommendation is innovative self-regulation. Silence does not mean consent, so the public must be educated about their rights and responsibilities (Caudill & Murphy, 2000).

McKenna (2001) also argued that it is somewhat unfair to expect consumers to wade through online privacy statements; one solution is to require notice on Web sites warning consumers about the site’s data collection practices. But true consumer control over data collection procedures is possible only with an opt-in device, McKenna claimed, allowing consumers to actively choose to visit or abandon sites whose data collection policies they do not agree with. However, if policies are implemented requiring e-commerce sites to offer consumers the ability to opt in before collecting individual data, e-tailers may find their costs of operation increasing (Farah & Higby, 2001). The loss of profiled consumer information would hinder Web businesses’ abilities to create huge databases of this information to be used for marketing purposes. Thus, e-commerce sites would contend with rising marketing expenses for less streamlined marketing practices. Some businesses complain that restrictive privacy mechanisms will result in the disappearance of free, high-quality service for the Web customer (Farah & Higby, 2001).

The corpus of relevant research highlights the complexity of protecting users’ private information under a legal regime that encourages self-regulation while maintaining profitability for online businesses and adjusting to technological advancements. Although previous research studied the legal complexities of providing such protection, we focus on the text of the privacy statement itself, examining the language through which online portals seek to guarantee privacy protection and explain how private information will be used. Therefore, we propose a content analysis of privacy statements that evaluates the overall efficacy of privacy statements.

We operationalize efficacy as the level of success with which the company explains and protects personal information. We attempt to capture consumer perceptions of privacy protection pledges because privacy statements are directed at consumers and because consumers are the group manifesting the greatest level of uncertainty regarding this information transaction. In doing so, we examine the usefulness of the statement for the average user and examine factors that may negate this efficacy, such as the complexity of the legal and technical language used; the length, organization, and readability of the statements; and the presence of explicit reassurances of privacy protection, among other statement characteristics. We also examine whether companies offer these reassurances in a credible manner, looking at whether the language used not only offers protection guarantees but does so in a convincing and persuasive manner. Because we have no way of actually checking whether these companies live up to their privacy promises, this is the closest measure we can have
of privacy statement believability. Thus, our research is guided by the following research questions:

RQ1: What is the perceived credibility of portal privacy statements?
RQ2: Which other factors influence overall privacy statement efficacy?

Method

Sample and Procedures

For the purposes of this study, we used the search engines of Google, Yahoo, AltaVista, Netscape, and CNet to obtain listings of portals and conduct a content analysis of their privacy statements. We used these popular search engines to obtain as diverse and comprehensive a sampling pool as possible. The random sampling employed facilitated the use of a sample that could be generalizable to the Web as a whole, although follow-up studies could work with larger samples and include foreign language portals. Using a random sampling interval, we obtained 200 portals from each engine, thus working with an initial sample of 1,000 listings. Out of those, we used a sampling interval to obtain a sample of approximately 100 portals, the privacy statements of which would be coded. We defined a portal as “a directory of general information.” If the Uniform Resource Locator (URL) led to a site that no longer existed, had switched focus, or had mistakenly been identified as a portal by the search engine, we skipped that listing and sampled the next one, based on the sampling interval. When we came upon several sites that did not feature privacy statements, we marked that absence and the URL of the site, and sampled the next site, as our goal was to code as many privacy statements as possible. Therefore, our sample was expanded to include 169 sites, 97 of which actually featured privacy statements that were coded.

The coding sheet and codebook were pretested by a group of students and colleagues, who used them to evaluate privacy statements of their choice. These allowed us to obtain feedback on how these statements are perceived from a diverse population, with varying levels of legal and computer expertise. The goal was to operationalize terms employed in the codebook in a manner that the average user could relate to. In addition, we tested the coding instrument on a smaller subsample of 10 privacy statements. After some additional clarification and wording adjustments, the privacy statements were analyzed by two coders, who had been properly trained and had practiced on a number of privacy statements. The coders did not have substantial expertise in computers or privacy law. They were trained to evaluate the statements based on the operationalizations provided by the study codebook; therefore, their own background was irrelevant. We coded for a number of descriptive characteristics and also included a few evaluative measures of overall statement clarity and credibility. For the 97 privacy statements coded, the average length was 1,362 words \((SD = 1,752, \text{ mode } = 1,245)\), with the statements’ word count ranging from 11,
from nwilink.com ("Privacy Pledge: we will never give your contact information to anyone"), providing a two-sentence-long privacy assurance, to 12,615 words from Real Networks. Coder agreement on this item was complete. Most privacy statements featured some hyperlinks (M = 6.89, SD = 11.74), very few of which linked to other organizations containing privacy information (M = 1.51, SD = 4.54, mode = 0), some linking to the same organization (M = 4.42, SD = 8.30), and even fewer linking to other pages containing information on privacy subtopics (M = 0.63, SD = 2.36, mode = 0). Terms of use statements were also featured on 60% of these sites, and children’s privacy policy statements were posted on 39% of the sites. The two coders reached complete agreement on all these items.

Finally, we also noted whether the site had been independently reviewed and certified by a privacy-related agency or resource. Intercoder agreement on this item was also complete. Several sites were certified by TRUSTe (8.3%), an online independent privacy monitor and certification agency. Because TRUSTe is a widely known and used service, we included an item that measured how closely privacy statements followed TRUSTe specifications, measured with .86 intercoder reliability. TRUSTe provides extensive privacy guidelines and specific privacy statement templates on its site; we evaluated how closely a privacy statement followed the TRUSTe template. Only 6.2% of the statements coded followed the TRUSTe template closely, with 10.3% adopting a slightly divergent format while still covering key TRUSTe areas, 80.4% covering the privacy policy but with a format and focus independent of TRUSTe specifications, and a few (3.1%) providing very brief privacy reassurances that did not constitute a privacy policy but rather a privacy pledge. Only one site was certified by an agency other than TRUSTe (BBonline), in addition to securing a TRUSTe certification.

Measures

We included several items that aimed at evaluating the effectiveness of the privacy statement, specifically relating to the perceived clarity and credibility of the document. We began by asking the coders to evaluate the readability of the document, using two items, on a 5-point semantic differential scale, ranging from 1 (understandable) to 5 (dense) and from 1 (organized) to 5 (unorganized). An understandable statement used simple writing, short sentences, unsophisticated language, and was overall easier to read through. A more dense statement, on the contrary, made use of longer sentences, employed more confusing sentence structure, adopted more sophisticated language, and was overall more difficult to browse through. An organized statement, similarly, utilized logical flow and rational progression, delineated clearly between different sections of the document, both design-wise and in terms of language use, and was easily navigable. In contrast, an unorganized statement appeared more cluttered, did not follow a logical procession in presenting the information, and did not distinguish between different sections of the document clearly. We decided against employing an automatic
readability method or algorithm because we were interested in gauging both read-
ability and understandability of statements. These required that both coders not
only examine the readability of the statement but also evaluate the clarity with
which sophisticated terms were explained. In addition, the text of privacy state-
ments with hyperlinks that frequently directed readers to pages with additional and
more specific information was not easily transferable into such a program. Finally,
it was essential to the validity of the research that both coders read these privacy
statements in their entirety. Both judgments were based on the statement as a
whole. Most documents were fairly understandable, with the majority (92.8%, M =
2.11, SD = 0.85, mode = 2) of sites obtaining scores between 1 and 3. Similarly,
most statements were clearly organized into manageable sections, with most sites
obtaining scores between 1 and 3 (78.4%, M = 2.56, SD = 1.06, mode = 2).
Intercoder reliability for both items was .79. Reliability for all content analysis vari-
ables was calculated using the Perreault and Leigh (1989) reliability index: 
\[ I = \frac{[(F_o/N) - (1/k)]}{k/(k - 1)} \]
0.5, for \( F_o/n > 1/k \), where \( F_o \) is the observed frequency of
agreement between coders, \( N \) is the total number of judgments, and \( k \) is the
number of categories. This index accounts for coder chance agreement, the number of
categories used, and is sensitive to coding weaknesses. Reliability scores can range
from 0 to 1, with higher scores indicating greater intercoder agreement.

In the same vein, we measured how infused with legal and computer terms these
documents were, in the event that use of such terms could produce a document vague
and unfriendly to the average user. To measure this aspect of user-friendliness, we
asked how frequently legal terms were used, how clearly legal terms were explained,
how extensively computer terms were used, and how clearly these terms were ex-
plained. For all four items, responses were measured on a 3-point scale, ranging from
1 (low) to 3 (high). Examples of legal terms included confidentiality, subpoena, en-
fforcement, court order, personally identifiable information, exigent circumstances,
and edicts of law, among others. Examples of computer terms included log files, IP
(Internet Protocol) addresses, cookies, Web beacons, third-party advertising, data
mining, and others. We looked at the entire statement and also considered text that
was employed to clarify the meaning and use of computer and legal terms. The fre-
quency rating was assigned based on the occurrence of these terms, keeping in mind
the length of the statement too. For the clarity item, a judgment was made on how
clearly these terms were explained, looking at the entire statement and not consider-
ing frequency of use. We found that most (78.4%, M = 1.28, SD = 0.57) statements
made very low use of legal terms, whereas the majority of statements made low or
medium use of computer terms (93.8%, M = 1.68, SD = 0.59). Legal terms on most
(63.5%) statements were not explained clearly at all (M = 1.45, SD = 0.65). The two
coders reached .84 agreement on the frequency of legal terms and complete agree-
ment on the clarity of legal terms. Computer terms were frequently not clearly ex-
plained either, with approximately half (50.5%) the statements receiving the mini-
imum rating on this item (M = 1.56, SD = 0.61). Intercoder agreement for the
frequency of computer terms used was .93, and for their respective clarity, .84.
We included one statement intended to measure the privacy statement effectiveness in providing user privacy assurances, which asked the coders whether they felt, having read the statement, that the site offered low (1), medium (2), or high (3) protection, measured with .87 intercoder reliability. Usually, sites that received a low rating (39.2%), offered absolutely no details of how any information collected by the site would be protected, or specified that both personally and nonpersonally identifiable information would be collected and shared with third parties. Sites receiving a medium protection rating (54.6%) offered assurances that personally identifiable information would be protected, but specified, at varying degrees of vagueness, that nonpersonally identifiable information would be collected, used further, and shared with third parties. Sites receiving a high protection rating (6.2%) assured users that both personally and nonpersonally identifiable information would be protected (M = 1.67, SD = 0.59).

Eight statements gauged the credibility of the overall statement, that is, how successful the statements were in persuading users that they were truthful and committed to protecting consumer privacy. There is no previous instrument that measured the perceived credibility of privacy statements, so we looked at how credibility in general has been measured for other media, and consulted the McCroskey (1966) and Berlo, Lemert, and Mertz (1970) Source Credibility Scales. Both of these scales are frequently used to measure source credibility within the context of interpersonal communication, measuring concepts such as character, honesty, authoritativeness, and safety. We were influenced by those elements conceptually closer to our objective and adapted the statements that were most relevant to the context of this study. Coders indicated agreement with these statements on a 5-point Likert-type scale, ranging from 1 (strongly agree) to 5 (strongly disagree). It must be understood that these credibility scales were created to capture the audience’s or reader’s overall evaluation of credibility. When employed in the newspaper and TV news setting, respondents are typically asked to rate sources on overall credibility. In this particular context, coders were asked to rate the believability of privacy statements using the scale. Because this presents an atypical use of the scale, we supplemented its use with the aforementioned items, which assess the protection rating (r = .65, p < .001) and legal (r = .24, p < .01) and technical (r = .36, p < .001) clarity of the document-measured concepts relevant to those implied by the credibility statements. Statistically significant correlations between those items and the credibility statements confirmed the conceptual precision of the credibility scale employed. Moreover, the coders were provided with explicit operational explanations and examples for each item, in the content analysis codebook and in training, so as to ensure consistency and accuracy in coding.

In this particular context, the coder, trained to represent the average consumer, was asked to read the overall statement and complete this scale, relying on the impression
of credibility the statement generated. Because average consumers are not law experts or computer experts, they are unable to evaluate the statement from a legal or technological perspective. Average consumers, however, can form personal opinions of exactly how convincing a privacy account or pledge is, which is why we coded for the overall impression of credibility, or perceived credibility. In responding to these statements, coders were asked to consider whether they got the sense that the statement existed for the sole purpose of explaining how private information would be used to legally protect the company, whether the statement specified the steps taken to protect personal information collected, if the statement was mostly about detailing how personal information would be exploited, and whether the language and assurances were clear or vague. For instance, for the “legal safeguard” item, coders were asked to determine whether the statements existed for the sole purpose of explaining precisely how the customer’s privacy will be taken advantage of, or whether they felt, having read the statement, that the company took serious steps to protect the personal information collected. For the “serious protection” item, coders were instructed to consider specific steps taken by the company to ensure privacy and were pointed to such examples. For the “company trusted with your information” item, coders were instructed to consider whether the language and explanations offered by the company were reassuring enough of privacy protection. Similar definitions were offered for the remaining items.

The specific statements used, with descriptive statistics, are provided in Table 1. A factor analysis and reliability information are reported in response to RQ1 in the following section. Intercoder reliability for these items ranged from .79 to .87.

Results

RQ1: Perceived Credibility of Privacy Statements

The first research question involved the perceived credibility of privacy statements and was evaluated using a credibility measuring instrument, adapted to this context from other media research. A factor and reliability analysis was conducted to fine-tune this measure further. These items were new, so the factor analysis was conducted to determine any overlap or similarities between certain items and yielded two primary factors. The first factor contained four items, which all directly inquired about privacy protection and personal information collection methods. The mean score for this factor was 2.44 (SD = 0.78), and the coefficient alpha was .92. This factor was named “protection for consumer” because it included statements that described the degree of privacy protection the portal in question provided. The second factor, named “purpose of statement,” included three items, which contained three statements that tackled the same topic but in a less direct manner. The mean score for this factor was 2.10 (SD = 0.56), with an alpha reliability of .53. One item (“This statement serves as a legal safeguard for the company”) did not load on either factor. Given
the low reliability of the second factor, and the fact that there was little conceptual distinction between these two, we assumed that the first factor contained the most valid and reliable items, especially because the items within the second factor evaluated credibility in a vaguer manner. The mean for the entire scale was 2.23 (SD = 0.50), with an alpha reliability of .79. Because reliability for the entire scale was higher, and given the vague conceptual distinction between the two factors, we decided to use the entire scale in subsequent data analyses. No significant differences were noted in the analyses when they were run using the two separate factors, which further supported our decision to use the entire scale. The results of the factor analysis are detailed in Table 2.

**Table 1**

<table>
<thead>
<tr>
<th>Items</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>This statement serves as a legal safeguard for the company</td>
<td>1.76</td>
<td>0.67</td>
</tr>
<tr>
<td>This statement offers no serious protection for the consumer</td>
<td>2.28</td>
<td>0.84</td>
</tr>
<tr>
<td>This statement focuses more on explaining how information will be used by the company</td>
<td>1.93</td>
<td>0.73</td>
</tr>
<tr>
<td>This statement explains how private information collected will be protected(^{a})</td>
<td>3.67</td>
<td>0.89</td>
</tr>
<tr>
<td>Having read the statement, you feel this company can be trusted with your information(^{a})</td>
<td>3.47</td>
<td>0.90</td>
</tr>
<tr>
<td>This company will not engage in improper use of private information(^{a})</td>
<td>3.36</td>
<td>0.89</td>
</tr>
<tr>
<td>The purpose of the statement is to inform the consumer</td>
<td>2.23</td>
<td>0.68</td>
</tr>
<tr>
<td>This statement is used as a promotional tool for the company’s stance on privacy</td>
<td>2.15</td>
<td>0.92</td>
</tr>
</tbody>
</table>

\(^{a}\)These statements were reverse coded for subsequent analysis.

**RQ2: Privacy Statement Effectiveness**

The second research question investigated relationships between characteristics of the privacy statement, aimed at evaluating overall statement effectiveness. Pearson correlations were used to take a closer look at these relationships where appropriate, and several significant findings emerged. The presence of a TRUSTe certification led to lengthier \((r = -0.33, p < .001)\) and less organized \((r = -0.44, p < .001)\) statements, in which legal terms were explained with greater clarity \((r = 0.32, p < .001)\). Understandably, the presence of TRUSTe certifications was related to a greater likelihood of adopting the TRUSTe format \((r = -0.73, p < .001)\). The length of the privacy statement was negatively related to the tendency to adopt the TRUSTe template \((r = -0.35, p < .001)\).
and positively related to frequent use of legal terms \( (r = .30, p < .01) \). Given the nature of the variables, this indicated that lengthier statements tended to follow the TRUSTe format closely and that frequent use of legal terms was usually encountered in longer privacy statements. The degree to which a statement was understandable was positively related to how organized it was \( (r = .39, p < .001) \) and how frequently legal terms were used \( (r = .28, p < .01) \), was negatively related to clarity of legal terms \( (r = -.37, p < .001) \) and the overall impression of whether the statement offered low, medium, or high protection \( (r = -.26, p = .01) \). This implied that understandable statements were more likely to be better organized and use more legal terms that are explained more clearly. In addition, the most understandable statements tended to give the impression of offering low protection, or perhaps, low protection was usually explained in simple and unsophisticated terms. Fairly similar findings were noted for the level of organization of privacy statements, which was positively related to the tendency to adopt the TRUSTe format \( (r = .36, p < .001) \), and negatively related to clarity of legal terms \( (r = -.35, p < .001) \), use of computer terms \( (r = -.21, p < .05) \), clarity of computer terms \( (r = -.23, p < .05) \), and the overall impression of whether the statement offered low, medium, or high protection \( (r = -.27, p = .01) \). The less closely a

### Table 2

**Perceived Privacy Statement Credibility Factor Analysis**

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor 1: Protection for consumer</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This statement offers no serious protection for the consumer</td>
<td>.87</td>
<td>-.00</td>
</tr>
<tr>
<td>This statement explains how private information collected will be protected*</td>
<td>.81</td>
<td>-.00</td>
</tr>
<tr>
<td>Having read the statement, you feel this company can be trusted with your information*</td>
<td>.91</td>
<td>.14</td>
</tr>
<tr>
<td>This company will not engage in improper use of private information*</td>
<td>.88</td>
<td>.20</td>
</tr>
<tr>
<td><strong>Factor 2: Purpose of statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This statement focuses more on explaining how information will be used by the company</td>
<td>.82</td>
<td>.18</td>
</tr>
<tr>
<td>The purpose of the statement is to inform the consumer</td>
<td>.67</td>
<td>-.20</td>
</tr>
<tr>
<td>This statement is used as a promotional tool for the company’s stance on privacy</td>
<td>.61</td>
<td>.00</td>
</tr>
</tbody>
</table>

*These statements were reverse coded for subsequent analysis.
statement followed the TRUSTe template, the more frequent the use of legal and computer terms; the less clearly it explained computer terms and the lower protection it offered, the more organized it appeared.

Related to these findings, the tendency to adopt the TRUSTe format was also negatively related to extensive use ($r = -.36, p < .001$) and low clarity ($r = -.22, p < .05$) of computer terms, indicating that those statements trying to follow the TRUSTe template (not necessarily TRUSTe-certified) were more likely to engage in greater use of computer terms, without necessarily explaining these terms clearly. The clarity with which legal terms were explained was positively related to the clarity of computer terms ($r = .37, p < .001$) and the overall impression of whether the statement offered low, medium, or high protection ($r = .27, p < .01$). This indicated that clarity of legal terms was frequently accompanied by clarity of computer terms and that this commitment to clarity gave the impression of higher privacy protection. The clarity of computer terms was also positively related to the overall impression of whether the statement offered low, medium, or high protection ($r = .49, p < .001$), thus reinforcing this finding.

Finally, the perceived credibility of the statement, measured by the corresponding scale, was positively related to clarity of legal terms ($r = .24, p < .05$), clarity of computer terms ($r = .36, p < .001$), the overall impression of whether the statement offered low, medium, or high protection ($r = .65, p < .001$), and negatively related to extensive use of computer terms ($r = -.23, p < .05$). This indicated that the perceived credibility of a privacy statement was higher when legal and computer terms were clearly explained, and not when computer terms were simply frequently used. Understandably, higher perceptions of credibility were likely to correspond to impressions of higher privacy protection, thus validating the perceived credibility measuring instrument. These results are discussed further in the following section.

**Discussion**

When planning this study, we expected that most portals would feature privacy statements prominently, given the heightened privacy concerns of consumers and the media coverage online privacy issues have received. Instead, we found that even though privacy statements were prevalent and extensive in most major commercial portals, like Google, Yahoo, and AOL, several smaller circulation portals did not feature them at all, or at least featured much less extensive and reassuring versions of them. The major portals were much more likely to feature lengthier, TRUSTe-certified statements, probably because they had the personnel to devote to this task, and they could afford the legal cost of providing a full privacy disclosure. Several smaller portals featured amateur attempts at providing a shorter, less detailed privacy statement, which rarely was certified by a third party like TRUSTe.

The ratings that sites received on overall impression of whether the statement offered low, medium, or high protection, were low, indicating that the presence of a privacy statement did not ensure privacy protection. The ratings received using our per-
ceived credibility measuring instrument ratified these initial ratings, revealing that privacy statements frequently do not guarantee the protection of personal information but rather serve as a legal safeguard for the company by detailing how personal information collected will be used. Even though sites frequently distinguished between personally identifiable and nonidentifiable information, the average low rating revealed that most statements did not successfully guarantee to protect this information either. Indeed, for several sites, an initial promise to protect personally identifiable information was countered by vague language on how that would be accomplished, especially as the sites were bound to sharing agreements with third parties and co-owners, as revealed later on in the statements. Sites also frequently warned that the privacy statements would be subject to change at any time, at which event users may or may not be notified. Nevertheless, several sites, like Real Networks, provided very extensive and reassuring statements, which were also perceived to be quite credible. The perceived credibility, of course, is no guarantee of actual protection, but as these privacy statements become more prevalent, it is worth looking into the factors that may influence such perceptions of credibility for potential users, who actually might take the time to read these privacy statements. Our intent, with using the perceived credibility measure, was to come as close as possible to evaluating whether these sites actually abided by what their privacy statements said.

Frequently, privacy statements offered conflicting or unconvincing reasons for collecting personal information, which tended to affect their credibility rating in a negative manner. For example, Dotplanet.org’s (n.d.) statement offered:

We use your IP address to help diagnose problems with our server, and to administer our Web site. Your IP address is used to help identify you and to gather broad demographic information. We use cookies to deliver content specific to your interests and to save your password, so you do not have to re-enter it each time you visit our site. Our sites’ registration forms require users to give us contact information (like their name and e-mail address) and demographic information (like their zip code, age, or income level).

Even though the statement may truthfully identify the type of information collected, the use of this information is justified in a rather ambiguous manner, pointing to site maintenance and to obscure user customization options. Similarly, About.com (2003) claimed that collecting personal information was undertaken merely for the advantage of the customer:

So that we can provide you with the most efficient and enhanced service, we request information about you. We collect personally identifiable information such as names, e-mail addresses and demographic information such as age, gender and zip code. We also may collect your IP address, browser type, domain name, access times and referring Web site address. This information is collected both during registration on a site or in response to specific requests, for example when signing up for a newsletter, en-
tering a sweepstakes or contest or answering a survey. Information collected at one About Web site will be available within the About family of Web sites.

These claims reduce the privacy statement to a company promotional tool and represent unconvincing attempts to portray personal information use as a beneficial service for the customer while obfuscating profits generated by the collection and trading of private data.

Therefore, language use was influential in establishing a conception of credibility. Specifically, the findings of this study revealed that perceived credibility of privacy statements was related to clear explanations of computer and legal terms, probably because the presence of such specific language helped create the impression of a more sincere approach to privacy protection. Vague or contradictory language, on the other hand, frequently challenged credibility perceptions. The extensive use of computer terms, conversely, was negatively related to perceptions of credibility, perhaps because the more frequently computer terms were used, the more extensive and sophisticated technology it seemed that the company had in place to collect and monitor personal information. These findings clearly correspond with the previous research that demonstrates public worry about Internet privacy.

In addition to our perceived credibility instrument, we included one overall rating on whether the statement offered low, medium, or high protection, based on the evaluation of coders who had read the entire statement. As previously indicated, sites that received a low rating offered no protection of personally identifiable or nonidentifiable information, sites receiving a medium rating protected identifiable information but not nonidentifiable information, and sites receiving a high rating promised to protect both. This rating was related to the understandability and organization of the statement, indicating that as the level of promised protection increased, statements became less understandable and organized. Sites that received these high ratings did tend to produce very extensive statements, which is a possible explanation for this finding. This rating was also related to clarity of computer and legal terms, revealing that high privacy ratings were met by clarity in expression when the discussion involved legal and technological issues. These findings are understandable and corroborate the relationship noted for the perceived credibility measure.

Evaluating privacy statements in this manner has permitted the analysis of privacy statement advantages and shortcomings. Limitations associated with this study, however, warrant additional research that could further validate these findings and lead to specific recommendations regarding online privacy protection. A larger sample of Web sites, a survey of the entire population, or the possible inclusion of several foreign language Web sites could further inform and validate these findings. In addition, a survey of average Web users employing the credibility measured adapted for this study could serve as an estimate of how average users perceive privacy statement credibility. This survey would have to work with a smaller sample of privacy statements that respondents would read and then evaluate. The purpose of this study was
to work with as large a sample as possible, which is why a smaller number of evaluators worked with a greater number of portal privacy statements. Additional instruments for evaluating privacy protections, perhaps by devising software or more objective means of measuring the privacy offered could be developed, even though intercoder reliability obtained for the instruments in this study was satisfactory. Finally, the use of qualitative methodologies would allow a more in-depth analysis of privacy statement language and pledges.

Future research could work on adding more items to the perceived credibility measure and investigate whether additional factors can be developed. The measure itself can be fine-tuned by improving the wording and modifying statements that behaved ambiguously in the factor analysis. A survey of users could be combined with the content analysis, so as to provide additional evaluations of privacy statements and understand how users respond to privacy pledges made by portals. The sample itself could be expanded to include additional portals, even several foreign language ones, which were excluded from this sample. The privacy statements of other organizations, including online retailers, financial institutions, insurance agencies, pharmaceutical companies, and other institutions could also be analyzed and compared. Terms of use statements and children’s policy statements could also be analyzed to determine what type of personal information is collected and/or protected by these sites. Finally, a discourse analysis of privacy statements could provide in-depth analysis of privacy statement content and examine how portals use statements to guarantee personal information privacy and/or legally protect themselves.

The study results are indicative of the need to not simply state how personally identifiable and nonidentifiable information will be used but to do so with specificity and clarity. Even though the etymology of the term privacy statement primes the user for a guarantee of privacy protection, the vocabulary of the statement itself frequently creates a legal safeguard for the company that seldom offers explicit reassurances of privacy protection. The overall usefulness of the privacy statement to the user may be compromised by the need to legally protect the business practices of the company in question. Stipulating precisely how information collected will be protected or shared leads to higher evaluations of credibility for the privacy statement and, consequently, the portal. Even though independent organizations, like TRUSTe, provide general format guidelines, specify which topics should be covered, and suggest language to be used, they do not emphasize clarity and specificity equally. Moreover, the focus so far has been to pressure companies to state how they use information rather than set certain privacy protection principles and ensure they are upheld. Companies are primarily rewarded for being truthful rather than protective. Nevertheless, online entities feel obliged to emphasize their commitment to protecting privacy through these statements, even though their reported use of personal information contradicts this pledge. It is this inconsistency between the promises offered and the practices divulged that prepares the consumer for privacy protection while offering an account of personal information use not worthy of consumer trust. Therefore, the privacy statement becomes an inadequate mea-
sure of privacy protection, designed to sustain an illusion of concern over information collected.

This study underlines the vulnerability of the privacy statement as a method of privacy protection. Our most important finding indicated that the majority of portals utilized the privacy statement to make vague promises of how personally identifiable information would be protected and ascertain their right to collect and trade nonpersonally identifiable data. Even by definition, the privacy statement itself is a flawed guarantee of privacy protection. Stating information collection practices does not guarantee that they will be maintained or that privacy is ensured. Constant technological changes and the interlocked nature of new media ownership combined with the minuscule number of users who ever notice or even read these statements further challenge the effectiveness of this instrument and indicate that privacy protection should involve more direct and legislative measures.

This research supports the FTC’s recommendation for a legislative solution to online privacy concerns rather than industry-driven privacy initiatives. Specifically, the monitoring of privacy statements and privacy practices, combined with retributions for privacy violations, would be a more effective method of personal information protection. Nevertheless, the regulatory structure and mentality present in the United States cannot readily support such measures, which, while foolproof, sound unrealistic. Perhaps the disparate nature of these statements lends support to the notion that federal legislation that enables consumers to opt in is the most comprehensive way to ensure the FTC-suggested privacy criteria of notice, access, security, and third-party disclosure. This would of course require a reversal of the present opt-out trend, which assumes that personal information can be traded and prompts users to take some form of action to ensure this does not occur. An opt-in approach could involve treating privacy protection as more of a business transaction between consumers and the private sector. Therefore, consumers would possess the right to negotiate their private information with greater precision in exchange for online benefits, thus conceptualizing private information as a type of private (intellectual) property.

Every approach to privacy protection should recognize that the notion of privacy involves a sense of basic human dignity, which can be compromised by the proliferation of sensitive information about oneself. Trust, respect, and personal integrity are at issue—consumers want to trust retailers and the Web but they also do not want to be defined by what they purchase and who they are on paper (as objects of marketing). Moreover, mistakes are made in the collection and handling of information that may result in a misrepresentation of the individual consumer; this may have consequences in that consumer’s financial transactions, credit health, and even personal relationships. Discriminatory practices may also result from the mishandling of information, such as being denied a loan. In a country that relies only on industry self-policing, citizen projects, and judicial enforcement of existing privacy laws, the ability of consumers to understand their privacy protections (such as they are) is of the utmost magnitude.
References


