Policy and Internet

A Broken System of Self-Regulation of Privacy Online? Surveillance, Control, and Limits of User Features in U.S. Websites

Yong Jin Park

This article examines user control of privacy online as indicated by functional features of commercial websites. While prior studies have focused on what's written in privacy policy statements, systematic attention on the interactive aspects of the Web have been scant. This analysis, based on a sample of 398 commercial sites in the United States, shows that the more popular sites did not necessarily provide better privacy control features for users than sites that were randomly selected. In addition, there was no clear relationship between website characteristics and the functional features of privacy control. Implications are discussed for the current status of online privacy policy in the United States and the European Union.

KEY WORDS: surveillance, personal information, marketplaces, information policy, privacy

Introduction

The rise of commercial websites has introduced a widespread fear concerning the power of surveillance that monitors, processes, and records the digital footprints of citizens online. Yet one might posit a different reality in which the World Wide Web, by its interactive nature, can function as a tool of resistance and privacy control. These assessments, however, derive from technological determinism in which Web's properties will lead to predetermined outcomes of either interactive empowerment or fearful surveillance. Most problematic is that both perspectives ignore the actual practices of commercial websites in appropriating the possibility of active information control. Instead of resorting to utopian and dystopian visions of the future of privacy online, this study draws on empirical evidence to aid a systematic understanding of the likely realities of the potential of privacy protection online.

We analyze the condition of personal information control in U.S. commercial websites. The central question is whether and to what extent the website interface is constructed as an enabler for informed choice in managing personal information. Here information privacy is defined as the ability to control one's personal

1944-2866
Published by Wiley Periodicals, Inc., 350 Main Street, Malden, MA 02148, USA, and 9600 Garsington Road, Oxford, OX42 DQ.

data and associated identities; widely regarded as one of the most vulnerable aspects of online use (Nissenbaum, 2011). Accordingly, at the policy level, our task is to examine the voluntary provision by commercial sites of information privacy protection and control under the self-regulatory policy of the U.S. Federal Trade Commission (FTC). Note the differences in regulatory conditions in the European Union (EU) where the EU Data Protection Directive 95/46/EC provides at least a broad principle of privacy protection and data retention. That is, U.S. commercial websites are almost entirely left alone in the marketplace to define the contexts of privacy protection except in the limited cases of protection concerning children's data, health, and financial security. To date, however, little is known about the extent to which the marketplace functions in shaping the potential of the privacy control that the Web in its interactive nature might afford.

This article begins with a brief framework of the technological affordances of information control, and identifies prior studies. Two analyses proceed in this study; the first examines website interfaces for evidence of informed choice; the second scrutinizes the relationship between the condition of user control and the site/domain characteristics. Finally, the status of online privacy policy in the United States is examined, as an example of market-based voluntary provision of information protection (FTC, 2010).

Technological Affordance

The Web is inherently engaging, bi-directional, and empowering. The institutional use of Web technologies in commercial contexts, however, can either impose or curtail structural constraints for the users (Dimaggio, Hargittai, Neuman, & Robinson, 2001; cf. Norman, 1988). Scholars (e.g., Barber, 1998) consistently point out that the Web may permit us to go back to the dynamics of the face-to-face interaction. In terms of privacy, this means the possibility of websites maximizing information control by which users can get easily informed and interact to protect their digital identities; as well as the possibility to impose restriction and maximize surveillance.[1]

Here Erving Goffman's observation about social interaction provides a point of reference. Decades ago, Goffman (1959) posited that humans perform private–public boundary management by selectively revealing the self. A strong assumption is that the surrounding environment and its implicit rules can be set up by individuals for them to be able to take appropriate actions. In other words, privacy action is aided by the interactive condition. In Erving Goffman's (1959) terms, this is the construction of the stage/the theater in which the individuals can manage the presentation of self (and selves). Note the centrality of interactive design in providing a tool of control over personal identities for citizens online. The presence of such designs that allow users to get *informed* and *interact* is crucial for users to be able to monitor, protest, or rectify the use of personal information (Marx, 2003). That is, the interface can be designed to encourage or curtail informed, interactive, and voluntary action by users as users are enabled (1) to get informed and (2) to exercise control in their own interests.

Nevertheless, whether the self-regulating marketplace leads to the technological affordance—the ability of users to control their online information—remains largely unexplored. On the one hand, interface design can empower user control with the appropriate tools of resistance against surveillance in online commercial transactions. Yet it is also plausible to posit that there may be insufficient incentive to provide such idealized conditions in the marketplace (Agre, 1999; Marx, 2003) because personal information remains one of the most commercially viable aspects in online business practices. Empirical evidence has been scant in this regard, as only a few advanced studies explicitly apply the idea to test the marketplace provision of privacy protection. In sum, the function of marketplace in shaping or responding to the potential of the user control of online information warrants systematic investigation.

Previous Studies

Scholarly attention in this area can be divided into two phases: (1) earlier online privacy studies up to the mid-2000s; and (2) more recent studies, which include the development of Web 2.0 (significant for increasing privacy concerns as new online applications institutionalize an intensification of data collection practices). The earlier studies tend to fall into three main areas, namely: (1) the Fair Information Practice Principles (FIPPs), (2) privacy statements, and (3) surveillance practices.

The first line of research concerns the observance of the FTC's FIPPs in commercial websites. This is perhaps the most comprehensive line of research in that the self-regulatory regime in the United States was to put to an explicit test with these concrete evaluative criteria. For instance, Culnan (2000) found that more than 67 percent of the sampled sites in her study ($n = 361$) collected personal data but only 14 percent of them provided any notice regarding data collection. It was also reported that most ecommerce websites failed to post "integrity" aspects of data uses, that is, whether data are used for the original purpose at the time of collection (Miyazaki & Fernandez, 2000). In this line, the most recent FTC policy report (2000) found that while 88 percent of the random sample sites ($n = 324$) disclosed data collection practices, only 10 percent of the random sites and 48 percent of the most popular sites ($n = 90$) voluntarily implemented all the elements of the FIPPs such as informing and allowing users to rectify incorrect information. Scholars (Park, 2011; Park & Jang, 2014) have raised consistent concern over the absence or presence of specific FIPP elements, as most consumer-oriented sites deviate from policy recommendations in their actual practices.

The second line of research up to the mid-2000s examines Internet privacy statements. The concern is the intentionality of the policy statement, that is, whether the written statement serves as a legal protection for the sites, rather than for consumers. This line of research has a long history, going back to consumer research examining the deceptive practices of fine prints in television and magazine (alcohol or tobacco) advertising (Barlow & Wogalter, 1991; Hackbarth, Silvestri, & Cosper, 1995). Numerous studies in this vein have focused on the

truthfulness of written statements and found that the policy statements in the U.S. commercial products often contained very few specifics about how to protect users, while most statements served to authorize business practices. This is supported by work by Privacy International (2006), which investigated the rhetorical strategies of the privacy policies of major Internet service providers (ISPs). Their study warned that most privacy statements were written "as little as possible" with the details of data collection seldom provided and "as confusing as possible" as the study characterized the online privacy statements of major ISPs as nothing but legal disclaimers guarding them against potential litigation.

The third line of studies in the pre-Web 2.0 era examines the data surveillance strategies as disclosed by individual sites in their privacy policies. Hong, McLaughlin, Pryor, Beaudoin, and Grabowicz (2005), for example, observed the scope of data collection and profiling among the news media sites (daily news, Internet media, magazines, and weeklies). A comprehensive study was also conducted by Privacy International (2006) which examined in the content analysis the extent of extraneous data retention and transfer to third parties (advertising broker) in such mega-sites as Google and Yahoo!. Schwaig, Kane, and Storey (2005) applied similar reasoning to the sites of Fortune 500 companies, asking whether offline credibility had any relationship with the scope of online personal data collection practices. The common thread among these studies is the discrepancy, rampant among the major sites across different commercial domains, between information surveillance and disclosure. That is, while data surveillance is becoming increasingly sophisticated through the use of third-party cookies and other collection strategies, a majority of commercial sites consistently fail to provide adequate information about how the data are used. Note that most content analyses in the earlier Internet privacy literature center on the integrity of policy statements. The FTC also cites the transparency of the written policies as an important criterion by which to measure the provision of privacy protection in the commercial sector. Unfortunately, systematic attention on either the interactive aspect of the Web (i.e., the ability of users to control their online information) or the function of marketplace in shaping such interactive features has been scant.

The recent studies undertaken since the development of "Web 2.0" technologies, however, have raised additional concerns. First, on the user side, empirical studies began to understand the depths of user demands and needs, and recent studies (Fogel & Nehmad, 2009; Hargittai & Litt, 2013; Park, 2013a, 2013b; Park, Campbell, & Kwak, 2012; cf. Popescu & Baruh, 2013) have found low levels of knowledge and self-presentation skills, indicating the need for systematic investigations on how commercial sites frame users' privacy control. Sheehan and Hoy (2000) previously pointed out that even when websites voluntarily followed FTC recommendations, most sites remained ineffective in responding to consumers' privacy concerns, indicating the need for better informative designs. Some evidence has also suggested that in an international context, standardized formats for disclosing online data collection practices may result in better protection as cultural differences may hinder users from exercising due privacy control (Park, 2008). Here it is particularly fruitful to question the underlying assumption of online privacy self-

regulation by which organizations are expected to provide settings for "informing" and "enabling" users to exercise control (Ashrafi & Kuilboer, 2005).

On the institutional side, recent empirical work has indicated the problematic settings of privacy preference and policy statement in commercial sites. Notably, this trend in the latter stage of privacy research marks a noticeable development as these studies highlight functional perspectives of privacy control from various disciplines: the interface design of user control from legal and policy studies (Hartzog, 2011; Massey, Eisenstein, Antón, & Swire, 2013; Park, 2011; Park et al., 2012), Platform for Privacy Preferences (P3P) compliance from information and computer science (Leon, Cranor, McDonald, & McGuire, 2010), perceived easiness of control from the business context (Kuo & Chen, 2011), as well as computer-mediated communication relational perspectives from interpersonal communication (Child & Petronio, 2010). Echoing all these efforts, Nissenbaum (2011) stressed the significance of privacy in context, with clearer privacy policies and fairer practices that can overcome a fundamental flaw in the U.S. policy assumption that users rationally understand all facts relevant to privacy choice. Her work is important in establishing common themes in the field that recognize the underlying concern of user control issues. Moreover, recent developments in surveillance practices and personalized advertising online highlight the need for clearer privacy control settings for users, as some of the works (Cecere & Rochelandet, 2013) have highlighted that consumers do not respond negatively to overly excessive data collection, suggesting the self-regulatory practices by websites would not be effective.

Still, in the U.S. context, systematic investigation is overdue concerning user interface settings for users to exercise control (Goffman, 1959; Marx, 2003). Note that the interactive condition alone does *not* determine cultural usages; however, it frames users' behavior and attention in personal information control. In this regard, the primary concern is the particularities of the site design embedded in the interface, as this is a manifestation of the deliberate choice on the part of the institution. In short, the institutional design of the participatory environment can either impose or curtail structural constraints for the users. To analyze how commercial websites are framed, channeled, and deployed to be accessible for informed choice is a natural extension of prior inquiries.

Research Questions

This study focuses on the accessibility of site features as "engineered" into interface design. Poor usability for users to be informed and interact with commercial websites can be a major hurdle for individuals' data control. This study aims to fill gaps in the literature with (1) a focus on usability, and (2) further analyses of individual website and market domain characteristics. The investigation will be undertaken with a relatively big sample size. Accordingly, we ask the following research questions:

RQ 1.1: To what extent do commercial websites inform users of data collection practices?

RQ 1.2: To what extent do commercial websites allow users to interact with them to exercise control of personal data?

RQ 2: How do the personal information control conditions in commercial websites differ by site characteristics and intended market?

Methodology

Modeled after the 2000 FTC landmark "web sweep" content analysis (FTC, 2000), this study combines several methods of random and cluster samplings from prior studies. The sample selection of commercial websites was based on the combination of two groups: Group (1), the top websites with the highest traffic (as identified by Alexa.com), and Group (2), a random sample drawn from an AOL search log of 500,000 AOL customers. This combination was to overcome the shortcomings of each method alone. The use of the AOL log ensures the variance of the sampled sites in the externally valid Internet universe, while the inclusion of the top sites incorporates the most routinely visited venues as operated in daily context. The sample pool was created in two steps. In Group (1), the U.S. sites were identified from the top 500 global websites (Group (1): $n = 153/500$). In Group (2), 500 websites were randomly selected from the first 10,000 AOL search queries (www.gregsadetsky.com/aol-data/). This resulted in the creation of the sample pool of 1,000 websites; the top and the random samples ($500 + 500$) combined. Any replicates between the top and the random sites were excluded from the sample pool. From the sample of top websites, one government-operated site and three sites with the same policies were excluded. Finally, a site with a U.S. Internet Protocol address, but operating under foreign ownership was eliminated, creating a total of 148 sites for analysis.

For the random sample that resulted from the elimination of duplicate sites ($n = 250$), the following multistage cluster sampling was used. In the first stage, 500 clusters of individual search queries were identified by randomly selecting them from the 10,000 AOL user batches. In the second stage, an individual uniform resource locator (URL) within each cluster was randomly selected. Each cluster was mutually exclusive, consisting of 20–70 unique URLs. With a total of 500 clusters, this selection includes 10,000–35,000 sites from which to select the final samples. Note the advantage of this technique in increasing the chance of equal selection when it is impossible to locate all the elements within the sample frame. Three broken URLs were identified and eliminated during the coding process. In Group (2), the sampling rate was $(0.25) \times (0.01)$, with the confidence level of 95 percent and a SE $\pm$ 4.9.

Here some of the limitations are worth noting. First, the FTC sample setup was originally geared toward simple descriptive analyses in the U.S. context. Second, caution is necessary against overgeneralizing the findings from this composite sample due to the unequal sizes of different market domains. In other words, smaller sample sizes (and associated small variance) in certain domains make this study's findings harder to generalize, especially when those sites

stacked up against more common sites such as ecommerce. Finally, we should be cautious about direct comparison given the online and associated practices of Web 2.0 have undertaken a profound development since 1998.

Despite such shortcomings, however, content analysis on the combined sample as utilized in the FTC study is advantageous for observing existing field data in natural environments. In essence, this is to investigate the way in which most users end up divulging personal data in their daily routines online. For one, the top and the random sample setup, because it allows separate as well as combined analyses, best detects the relative performance of the most visited commercial websites. In this regard, careful attention on respective performances of market domains—albeit, with limited generalizable power—is much needed. Furthermore, the same setup as FTC provides an ideal baseline that can be used to measure the current status of privacy control in the Internet. Also noteworthy in this respect is that despite the development of Web 2.0 applications, the technical mode of data collection in commercial websites—and how users are situated to exercise control—remains remarkably consistent.

This study operationalized the website attributes in terms of the two functionalities: (1) inform and (2) interact. The extant literature (FTC, 2000; Turow, 2001; West & Miller, 2006) provided the base coding instruments. The aim was to establish the criterion validity, while advancing prior measures. Web in its interactive characteristics are the most fundamental architectural codes of the Internet. The key was then to measure the easiness of (1) access and (2) use/choice in managing information flow (i.e., interface design that supports informed choice). Inform (IF) dimension aimed to capture the extent to which users are to be informed of data practices by websites. The Interact (IT) dimension aimed to capture the extent to which users are able to manage/control information flow. For IF items, the presence of a link to a privacy statement, the link placement in a prominent place, font size and color differences from adjacent words and main text, clear labeling, and (Flesch–Kincaid) readability were coded (Massey et al., 2013). For IT items, the presence of a link in every page, the seal with a tagged link, an email link, the availability of a downloadable form, and a link to associated third parties were coded. In both dimensions, the sub-items were included to further specify each function.[2] These are discrete items within each function. In the IF dimension, the sub-items were text length for policy presentation, link placement in main menu for prominence, and other clarity for link labeling. In the IT dimension, the inclusion of other interactive features in the site, the availability of the P3P (i.e., a protocol for privacy protection code) function, the presence of opt-out options, any link to complain, and the number of clicks to privacy policies of the site and of associated third parties were noted.

Two coders were hired to code the individual websites. The coders underwent at least three training sessions. Intercoder reliability, based on a pilot sample (10 percent of the full sample; see Lacy & Riff, 1996), was calculated for Cohen's kappa ($\kappa = 0.84$ in total items).

In analyzing the condition of information control, various market-domain and site factors were included in line with prior literature. Note the two levels in the explanatory variables: (1) the market domain, and (2) the site factors (Table 1). The

Table 1. Characteristics of Sample Sites ($n = 398$)

Levels	Descriptions	Mean	SD
Market factors	Type of market domain		
M 1: Online	Whether the site operation is confined online (1 = yes, 0 = no)	0.63	0.48
M 2: New media	Search engine or directory sites (1 = yes, 0 = no)	0.10	0.29
M 3: Sensitive	Whether a site deals with sensitive data (health or financial information) (1 = yes, 0 = no)	0.08	0.26
M 4: Younger	Whether a site is targeted toward children, teenagers, or younger users (1 = yes, 0 = no)	0.08	0.27
Site factors	Characteristics of an individual site		
S 1: Publicly listed[a]	The site (or its parent company) in public stock market (1 = yes, 0 = no)	0.30	0.46
S 2: Ranking	Traffic ranking in September 2008 (000,000)	56,978.3	302,953.7
S 3: Years	Number of years of operation	10.24[b]	3.50
S 4: Seal member	Whether a site is a member of Truste, BBBonline, or Safe Harbor (1 = yes, 0 = no)	0.29	0.22
S 5: U.S. percent	Percent of U.S. users (%)	62.38	26.38

Notes: Data are for 2008, unless otherwise indicated. Data sources are (1) alexa.com for ranking, publicly listed, year, and U.S. percent; (2) truste.com and bbbonline.com for seal member; and (3) coders identified domain characteristics through corporate info (e.g., About us) in each site. For instance, in identifying the websites targeting children or teenagers (M 4), the coders referred to the site information to see whether their products are toys, children-teen games, and other related services such as the Cartoon Network. In any case of ambiguity, we made the most conservative decision not to classify the sites as M 4.

[a]Publicly listed is a proxy value for revenue.

[b]The medium is 1998, indicating most sampled sites are well established in marketplace.

market factors served to measure whether the sites in a specific market domain are in fact more inclined to provide privacy protection functions. The site factors were included to measure the influences of individual website attributes, such as financial resources (Schwaig et al., 2005) and the number of years in site operation (Palmer, Bailey, & Faraj, 2000), in incentivizing further provision. We examine whether better-resourced sites, as indicated in revenues, traffic ranking (Hindman, 2008; Schwaig et al., 2005), seal membership or broader business scopes (Danna & Gandy, 2002), are more responsive to the demand from the public in the marketplace.

Data Analysis

In line with these strategies, two analyses proceeded. The first analysis was the comparison between the top and the random sites. A series of 2×2 tables were constructed for mean comparison among individual items that described each function of IF and IT. The second line of analysis took into account domain and site characteristics. For this, additive indexes in IF and IT dimensions were calculated with the presence or absence of content criterion (0 = absence; 1 = presence). For readability and text length, a dichotomous value was assigned (0 for the sites higher than the medium value). Multivariate ordinary least squares (OLS) regressions were run on the IF Index (range 0–12, mean = 5.74, SD = 2.47),

the IT Index (range 0–10, mean = 3.50, SD = 1.85) and then, the combined score (range 0–18, mean = 9.25, SD = 3.37). OLS regression is appropriate, given variances were not significantly larger than the mean scores, in which case count models may be applicable.

Results

Descriptive Findings

Research question 1 asked the extent to which the sampled sites were designed for informed control, as indicated by the discrete items in each dimension of IF and IT. Tables 2 and 3 display the findings. In the IF dimension (Table 2), the mean comparison showed no significant difference between the top and the random sites. The item with significant difference was IF 4.2 (other features for clarity), with more random sites in such provision (31.4 vs. 45.4 percent). Further, the random sites were more likely to have accessible policy statements than the top sites (13.3 vs. 11.6 Flesch–Kincaid grade levels). The texts in the top sites tended to be far longer than those in the random sites. The baseline of overall provision was found to be extremely low, as some items reached less than 10 percent. Note the three items in Table 2: IF 1 (presence of link to statement), IF 1.1 (presence of one clear policy), and IF 4 (clear label). More than 50 percent of the top and random sites provided these functionalities. However, the difference did *not* reach a significance level when the random and the top sites were compared. In the IT dimension (Table 3), significant differences between the top and the random sites were found, as the top sites were more likely to score

Table 2. Content Analysis: Inform

Items	Total	Top	Random
IF 1. Presence of link to privacy statement in front page	87.9	89.9	86.7
IF 1.1. Presence of one clear policy statement	84.1	88.4	81.5
IF 2. Placement: link placed in a clear prominent place	4.5	4.7	4.4
IF 2.1. Link placed in main menu	2.0	2.0	2.0
IF 3. Font size and color of the link to privacy statement			
IF 3.1. Font size is larger than adjacent words	4.5	7.1	2.9
IF 3.2. Font size is larger than main text	7.1	4.3	8.8
IF 3.3. Font color is different from adjacent words	7.9	6.4	8.8
IF 3.4. Font color is different from main text	35.5	34.3	36.3
IF 4. Link clearly labeled as "privacy policy"	73.4	77.9	70.8
IF 4.1. Other clarity in labeling	4.2	5.0	3.8
IF 4.2. The link has other features (italics; highlighted; underlined) that make it stand out	40.3	31.4	45.4*
IF 5. Readability (Flesch–Kincaid grade level)[a]	12.2	13.3**	11.6
IF 5.1. Text length[b]	1,786.4	2,221.6**	1,522.4

Notes: * Significant at 0.05 level; **significant at 0.01 level. Entries are percents based on the number of sites in each category.[a]Grade level indicates the accessibility of policy statement.[b]Unit is the number of word count.

Table 3. Content Analysis: Interact

Items	Total	Top	Random
IT 1. Privacy policy is linked from each page	84.4	87.2	82.7
IT 1.1. Number of clicks away[a]	2.0	2.1*	1.9
IT 1.2. Others: privacy blog, discussion lists, etc.	8.4	14.9**	4.5
IT 2. Privacy seal or safe harbor visible with a tagged link	16.5	23.0*	12.6
IT 3. Active email link to make inquiries	57.2	62.2	54.3
IT 3.1. Out-links to complain or make inquires (e.g., FTC or other associations)	12.2	19.6**	7.7
IT 4. Availability of downloadable form to request, correct, or confirm data uses	17.5	16.9	17.8
IT 4.1. Edit function, e.g., preferences or profile	25.8	39.2**	17.8
IT 4.2. The option of opt out	33.0	45.9**	25.3
IT 4.3. P3P embedded	77.7	83.1*	74.5
IT 5. Link to privacy policies in third-party sites associated	20.3	35.8**	10.9
IT 5.1. Number of clicks away[b]	2.7	3.2**	2.4

Notes: *Significant at 0.05 level; **significant at 0.01 level. Entries are percents based on the number of sites in each category.

[a]Unit is the average number of mouse clicks from the front page to the policy page.

[b]Unit is the average number of mouse clicks from the front page to the policies in third-party sites.

well in the provision of IT items. The biggest difference (24.9 percent) was present in IT 5 (link to third parties associated). While 35.8 percent of the top sites had the links, only 10.9 percent of the random sites had them. Among the sealed sites, the top sites were more likely to provide tagged linked seals than the random sites (23.0 vs. 12.6 percent). Further, the top sites tended to contain more IT 3.1 (links to complain) than the random sites (19.6 vs. 7.7 percent).

Despite the differences in favor of the top sites, it is critical to see the provision of IT items in the absolute sense. Noteworthy is the low level of voluntary provision of user control in most sites. In fact, more than 50 percent of both top and random sites provided only three functionalities; IT 1 (policy linked from each page), IT 3 (active email), and IT 4.3 (P3P). In other words, while the top sites performed better in this dimension, the provision of most IT items remained particularly limited as shown in Table 3. The greater number of third-party links in the top sites may derive from the top sites having a greater number of third parties associated with them than the random sites, many of which are small-scale business sites. Interestingly, the number of clicks away from the front page to the policy page was significantly greater in the top sites. That is, the site features for control were deeply embedded among different pages within the site, indicative of limited usability.

Explaining the Findings

Research question 2 asked the extent to which the self-regulating commercial sites, as indicated by site and market characteristics, were related to the presence

Table 4. OLS Regression Analysis

	Full Model, IF + IT		IF Dimension, Inform		IT Dimension, Interact	
	Coef.	*t*-Value	Coef.	*t*-Value	Coef.	*t*-Value
Market factor						
M 1: Online	0.00	0.01	−0.03	−0.37	0.03	0.60
M 2: New media	0.04	0.74	−0.03	−0.52	0.12	2.32*
M 3: Sensitive	0.04	0.71	0.05	0.77	0.01	0.23
M 4: Younger	0.07	1.12	0.05	0.86	0.04	0.93
Site factor						
S 1: Publicly listed	0.05	0.78	−0.05	−0.68	0.15	2.67**
S 2: Ranking	−0.09	−1.34	0.08	1.11	−0.24	−4.48**
S 3: Years	−0.15	−1.95^	−0.10	−1.24	−0.13	−2.04*
S 4: Seal member	0.29	4.45**	0.01	0.19	0.47	8.78**
S 5: U.S. percent	−0.00	−0.07	0.01	0.12	−0.02	−0.34
R^2	0.17		0.03		0.46	
SE		2.72		2.08		1.33

Notes: ^Significant at 0.10 level; *significant at 0.05 level; **significant at 0.01 level. Entries are standardized regression coefficients.

or absence of information control. Table 4 presents the results from OLS regression analyses. In the IF Index, regression coefficients indicated no significant impact of the market factors. Furthermore, none of the site factors had any effect on the IF Index. That is, no clear pattern of market and site characteristics was found to be conducive to different extents of such provision. In the IT Index, there was a positive impact of M 2, indicating the new media sites performed better ($\beta = 0.12$, $P < 0.05$). Also, the impacts of S 1 (public), S 2 (traffic ranking), and S 3 (year) were found to be significant ($\beta = 0.15$, $P < 0.01$, $\beta = -0.24$, $P < 0.01$, $\beta = -0.13$, $P < 0.05$) with the biggest contribution from S 4 (seal member) ($\beta = 0.47$, $P < 0.01$). The findings indicate that at least in the provision of IT items, the site characteristics made significant contributions to variations.

However, when the IT and the IF dimensions were combined into the full model, the significance disappeared except with respect to S 4 (seal member) ($\beta = 0.29$, $P < 0.01$). The impact of S 4 should *not* be overinterpreted, as the function of IT is to be understood in the continuum of IF. In other words, it seems less meaningful to have the full control attribute when its elements were deeply embedded (i.e., hard to locate) within the site itself. Significance was found for S 3 (year) ($\beta = -0.15$, $P < 0.10$), indicating the new startup sites did *not* necessarily perform better in the combined measure. The full model, as a whole, contributed to an R^2 of 0.17 (SE = 2.72). This result remained robust as the truncated model, with no high provision item, provided almost the identical results. In none of the models was the level of tolerance for multicolinearity above 0.5, as measured by variance inflation factor. Also, the correlations among independent variables were not found to be prohibitively high (see Table 5). Given the limits of the secondary data, alternative predictors were employed in separate analyses when available. Most notably, the variable of "publicly listed" (parent companies in public listing

Table 5. Correlation Coefficients for Independent Variables

	M 1	M 2	M 3	M 4	S 1	S 2	S 3	S 4	S 5
M 1: Online	—								
M 2: New media	0.20**	—							
M 3: Sensitive	−0.10*	−0.09	—						
M 4: Younger	0.21**	−0.03	−0.08	—					
S 1: Public	−0.02**	0.01	0.18**	−0.15*	—				
S 2: Ranking	−0.08	−0.04	−0.01	−0.00	−0.06	—			
S 3: Years	0.45**	0.09	−0.00	0.10*	−0.38**	0.05	—		
S 4: Seal member	−0.05	0.07	0.03	−0.04	0.20**	−0.06	−0.25**	—	
S 5: U.S. percent	−0.37**	−0.12*	0.23**	0.02	0.10	0.11*	−0.22**	−0.01	—

Note: *Significant at 0.05 level; **significant at 0.01 level.

or not) was used as a proxy variable of revenue (of which the data reliability is questionable, given its extremely wide SD). Yet the OLS results remained consistent, indicating that noneffect of market and site factors on the provision of information control was the result of robust statistical tests.

Discussion

The two analyses examined (1) the features of website interface for informed choice, and (2) the relationship between the condition of user control and site/domain characteristics. The results suggest the dubious function of the marketplace in harnessing technological affordance of information control. First, the analyses based on the top and random samples of this study showed the limited extent of user control embedded in the sites. While the top sites performed better in the IT dimension, the provision of control features by both top and random sites was limited in the absolute sense. Furthermore, there was no critical difference between the top and the random sites in most IF items. Second, the regression analyses showed that although a few site characteristics had significant impacts on the IT dimension, the influence of such factors was limited in the combined measure. In addition, the market factors had no or limited impact on the extent of such provision in either the dimension of informing or interacting for control.

In idealized digital spheres, users should be able to exercise privacy control by easily finding how information about them is collected, retained, and processed. Further, they should be able to post questions or responses to the policy they find questionable and engage with others through interactive links and features connected to the sites. Yet at present, the commercial websites function as a one-way surveillance platform, largely closed in interface constraints, with limited provision of interactive links and features. Simply put, the findings from this study's sample indicate that the platform with the interface channel that allows users to freely exercise control does not exist.

In this light, the poor design of the websites oriented toward young users and sensitive data (health and finance) raises a particular policy concern.[3,4] Indeed,

the fact that those sites were to be found low in the provision of control links and features points out a grave policy problem in these sectors, for which potential data misuse may have tangible consequences. Younger users with heavy online use will also remain more vulnerable as they engage in political or social online activities through such sites as Facebook and MySpace. For health and finance websites, users may well expect privacy, given the sensitive nature of the personal data identified in such sites. Granted that we need a longitudinal study with a larger sample size from these sectors, it is still alarming that the interface is far from transparent in health- and finance-related sites because those sectors are likely to have the greater extent of data mining (Danna & Gandy, 2002). This concern resonates with the recent FTC investigation about the misuse of personal data by data brokerage firms in the financial sector (FTC, 2014). Also in the health sector, the prior findings by West and Miller (2006) have indicated inadequate readability of policy statements in e-health websites in the United States Here the incongruence between supply (i.e., privacy protection) and demand (i.e., the public concern) appears to exist as the market segments do not meet the expectations of parental or patient concerns.

When these results are compared with findings from prior studies, we can see that no clear improvement has been made in the commercial spheres since 1998 when the FTC undertook its first investigation. For instance, in 1998, 82 percent of the sites[5] posted a policy notice. In 2008, the number remained more or less the same, with 88.4 percent of the top sites providing one. Neither has there been any apparent improvement in the prominence of the policy (linked from main menu), given the sites with no seal membership (10.6 percent) in 2002 provided more direct menu options than either top or random sites in 2008 (2.0 percent). In terms of the readability of the statement, we can speculate that the trend appears to be regressing, as this study's findings hint that statements may be getting even longer and more complicated, perhaps reflecting more extensive data collection to be described as part of disclaimers.[6] Here it should be noted that commercial data collection has become far more common in today's digital environments. More importantly, the condition in which users are to get informed has not been much improved, making it harder to conclude that there have been dramatic changes in commercial websites.

Some of the findings appear to suggest that the marketplace improved. At least seal member sites perform better than nonseal member sites in overall provision of IF and IT dimensions. The improved privacy features in the IT dimension in new media sites also suggests that large search engine and directory sites such as Google or Yahoo! may be more receptive to a growing public concern. Granted this may be the case; it presents the most optimistic view in a small subset of the sample sites. Moreover, this study cannot discern the clear benefit of more provision of IT items in these sites, given they tend to be bigger than the random sites in the number of internal pages, which creates a burden for users to control with no clear signposting. Thus, this study cautiously eliminates the alternative interpretation that commercial websites are realizing the potential of usability for informing and interacting for privacy.

Policy Discussion and Conclusions

This study's findings suggest that policymakers have daunting tasks. First, how do we make use of a wide range of technological capabilities to translate the goal of privacy protection and control against surveillance in commercial sites? And how do we ensure that the technological affordance of information control does not disappear in the marketplace due to failures from business practice or government policy? The findings of this study should inform policymakers in the United States and the EU of the potential pitfalls when privacy protection is entirely left to the marketplace alone.

In fact, in the context of the United States, the FTC has maintained its faith in market-based self-regulation to this day. In 2010, the FTC proposed a "Do Not Track List" through which users can request a list of specific websites not to track user behavior. Despite its promises, this proposal concerns third-party behavioral advertising, effectively leaving market self-regulation (by the FTC regarding data retention and collection in most online transactions) as a de facto policy. The findings of this article pose a serious question on the continuing market-driven policy in the United States (see the Consumer Privacy Bill of Rights for a similar policy stance; The White House, 2012). It should also be noted that while the EU policy stance is more rigorous than that of the United States, it does not mean that the institutional practices in commercial websites are also better implemented in the European context (Robinson, Graux, Botterman, & Valeri, 2009). This is particularly so, given the low provision of the site features and links found in a majority of the sampled sites that have a global market reach and even operate for commercial transactions with EU consumers. At the end, it appears that the function of commercial entities in pursuit of economic efficiency may be at odds with the policy rationale behind the voluntary implementation of the FIPPs. Subsequently, this study urges that the focus of privacy protection should be on how users can actually exercise control over personal information, not on what is written in the policy statement. In this regard, specific policy recommendations can include the implementation of website features that handle basic "opt-out" functions. In informing users, the standardization of privacy policy statements in terms of readability and length may be a viable option. Surely, these types of functional features must stand out in design salience and ease of use.

As in any empirical endeavors, this study is not immune from shortcomings. First, a caution is necessary because in employing cross-sectional data, this study could not compare the provision by the sites before and after the site characteristics, such as revenue or traffic ranking, changed. In addition, the study sampled only U.S. commercial websites. Thus, the inferences from this study's findings may not be generalizable to noncommercial websites or sites operating in other nations. Future studies must discern the potential effects of the interactive interface on user learning and protective behaviors because content analysis does not detect actual user action. In this regard, in-depth observations of user interaction (see Park & Jang, 2014) are also needed as empirical reports of user frustration or action will inform policymakers of the value of usability. Still, until those works are to be

done, policymakers will be better guided by this study's findings. That is, the policy remedies of the specific interface guidelines should be implemented, as the current marketplace practices inhibit the potential of user informing or control over the use, collection, and retention of personal information.

Yong Jin Park, Ph.D., is an Associate Professor at School of Communications, Howard University, Washington, DC [yongjinp@hotmail.com].

Notes

The author feels very grateful to the anonymous reviewers for their kind insights. The author also wishes to note his gratitude to the members of Howard Media Group at Howard University.

1. In the privacy literature, the term surveillance usually indicates data monitoring by the government or law enforcement agencies. Given the public and the private sectors increasingly overlap in data activities, we use this term broadly to encompass data collection in commercial contexts. Still, a caution is needed for readers to discern the distinction between the private and the public sectors in their respective focuses of surveillance.
2. Here these sub-items can be understood as more direct measures that complement the IT dimension, while the IT items in general should be construed as the interface conditions that potentially enable active privacy-related actions from users. Still, the future studies should invoke behavioral experiments in which researchers are allowed to discern actual actions from the part of users in their interaction with sites.
3. Special locus of health and finance-related websites was noteworthy as coders referred to corporate info of each site to define health and finance-related websites. First, in the U.S. context, domain-specific regulations such as the Health Insurance Portability and Accountability Act, Financial Institution Privacy Protection Act, and the Gramm-Leach-Bliley Act add the weight to the two domains. Still under these laws, data collection activities are left unregulated, unlike data sharing which is subject to regulations. Also in the user sides, the existing survey data indicate much needed attention on these types of websites as most users remain particularly sensitive about the release of financial and health-related personal data.
4. Children's Online Privacy Protection Act (COPPA) is perhaps the strongest example of U.S. law regulating websites' abilities to collect personal information. COPPA, however, concerns the websites targeting the children under the age of 13, leaving a vast majority of websites unregulated.
5. The number was weighted for top and random sample websites.
6. The comparison is among the top sites in Flesch–Kincaid readability. Milne, Culnan, and Greene (2006) reported 11.2 in 2001 and 12.3 in 2003, in their sampled sites.

References

Agre, P. 1999. "The Architecture of Identity: Embedding Privacy in Market Institutions." *Information, Communication and Society* 2 (1): 1–25.

Ashrafi, N., and J.P. Kuilboer. 2005. "Online Privacy Polices: An Empirical Perspective on Self-Regulatory Practices." *Journal of Electronic Commerce in Organizations* 3 (4): 61–74.

Barber, B. 1998. "Three Scenarios for the Future of Technology and Democracy." *Political Science Quarterly* 113 (4): 573–89.

Barlow, T., and M.S. Wogalter. 1991. "Alcohol Beverage Warnings in Print Advertisements." Paper presented at 35th Annual Meeting of the Human Factors Society: Human Factors and Ergonomics Society, September 2–6, San Francisco, CA, 51–55.

Cecere, G., and F. Rochelandet. 2013. "Privacy Intrusiveness and Web Audiences: Empirical Evidence." *Telecommunications Policy* 37 (10): 1004–14.

Child, J.T., and S. Petronio. 2010. "Unpacking the Paradoxes of Privacy in CMC Relationships: The Challenges of Blogging and Relational Communication on the Internet." In *Computer-Mediated Communication in Personal Relationships*, eds. K.B. Wright and L.M. Webb. New York: Peter Lang, 21–40.

Culnan, M.J. 2000. "Protecting Privacy Online: Is Self-Regulation Working?" *Journal of Public Policy and Marketing* 19 (1): 20–26.

Danna, A., and O. Gandy. 2002. "All That Glitters Is Not Gold: Digging Beneath the Surface of Data Mining." *Journal of Business Ethics* 40: 373–86.

Dimaggio, D., E. Hargittai, R. Neuman, and J. Robinson. 2001. "Social Implication of the Internet." *The Annual Review of Sociology* 27: 307–36.

Federal Trade Commission (FTC). 2000. *Self-Regulation and Privacy Online Before the House Commerce Subcomm. on Telecom., Trade, and Consumer Protection*.

Federal Trade Commission (FTC). 2010. *Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers*. http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

Federal Trade Commission (FTC). 2014. *Two Data Brokers Settle FTC Charges That They Sold Consumer Data Without Complying With Protections Required Under the Fair Credit Reporting Act*. http://www.ftc.gov/news-events/press-releases/2014/04/two-data-brokers-settle-ftc-charges-they-sold-consumer-data.

Fogel, J., and E. Nehmad. 2009. "Internet Social Network Communities: Risk Taking, Trust, and Privacy Concerns." *Computers in Human Behavior* 25: 153–60.

Goffman, E. 1959. *The Presentation of Self in Everyday Life, University of Edinburgh Social Sciences Research Centre*.

Hackbarth, D., B. Silvestri, and W. Cosper. 1995. "Tobacco and Alcohol Billboards in 50 Chicago Neighborhoods: Market Segmentation to Sell Dangerous Products to the Poor." *Journal of Public Health Policy* 16 (2): 213–30.

Hargittai, E., and E. Litt. 2013. "Facebook Fired? The Role of Internet Skill in People's Job-Related Privacy Practices Online." *IEEE Security & Privacy* 11 (3): 38–45.

Hartzog, W. 2011. "Website Design as Contract." *American University Law Review* 60 (6): 1635–70.

Hindman, M. 2008. *The Myth of Digital Democracy*. Princeton, NJ: Princeton University Press.

Hong, T., M. McLaughlin, L. Pryor, C. Beaudoin, and P. Grabowicz. 2005. "Internet Privacy Practices of News Media and Implications for Online Journalism." *Journalism Studies* 6 (1): 15–28.

Kuo, H.-M., and C.-W. Chen. 2011. "Application of Quality Function Deployment to Improve the Quality of Internet Shopping Website Interface Design." *International Journal of Innovative Computing, Information and Control* 7 (1): 253–68.

Lacy, S., and D. Riff. 1996. "Sampling Error and Selecting Intercoder Reliability Samples for Nominal Content Categories." *Journalism and Mass Communication Quarterly* 73 (4): 963–73.

Leon, P., L. Cranor, A. McDonald, and R. McGuire. 2010. "Token Attempt: The Misrepresentation of Website Privacy Policies Through the Misuse of P3P Compact Policy Tokens." Paper presented at the ACM Workshop on Privacy in the Electronic Society, WPES 2010, October 4, Chicago, IL.

Marx, G. 2003. "A Tack in the Shoe: Neutralizing and Resisting the New Surveillance." *Journal of Social Issues* 59 (2): 369–90.

Massey, A.K., J. Eisenstein, A.I. Antón, and P.P. Swire. 2013. "Automated Text Mining for Requirements Analysis of Policy Documents." *21st IEEE International Requirements Engineering Conference (RE), 4–13*. http://www.cc.gatech.edu/~jeisenst/papers/re13rt-p085-p-18125-preprint.pdf.

Milne, G.R., M. Culnan, and H. Greene. 2006. "A Longitudinal Assessment of Online Privacy Notice Readability." *Journal of Public Policy & Marketing* 25 (2): 238–49.

Miyazaki, A.D., and A. Fernandez. 2000. "Internet Privacy and Security: An Examination of Online Retailer Disclosures." *Journal of Public Policy & Marketing* 19 (1): 54–61.

Nissenbaum, H. 2011. "A Contextual Approach to Privacy Online." *Daedalus* 140: 32–48.

Norman, D.A. 1988. *The Psychology of Everyday Thing*. New York: Basic Books.

Palmer, J.W., J.P. Bailey, and S. Faraj. 2000. "The Role of Intermediaries in the Development of Trust on the WWW: The Use and Prominence of Trusted Third Parties and Privacy Statements." *Journal of Computer-Mediated Communication* 5 (3). http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2000.tb00342.x/full.

Park, Y.J. 2008. "Privacy Regime, Culture and User Practices in the Cyber-Marketplace." *Info* 10 (2): 57–74.

Park, Y.J. 2011. "Market Philosophy and Information Privacy." *Javnost-The Public* 18 (2): 87–100.

Park, Y.J. 2013a. "Digital Literacy and Privacy Behavior Online." *Communication Research* 40 (2): 215–36.

Park, Y.J. 2013b. "Offline Status, Online Status: Reproduction of Social Categories in Personal Information Skill and Knowledge." *Social Science Computer Review* 31 (6): 680–702.

Park, Y.J., S. Campbell, and N. Kwak. 2012. "Affect, Cognition and Reward: Predictors of Privacy Protection Online." *Computers in Human Behavior* 28 (3): 1019–27.

Park, Y.J., and S.M. Jang. 2014. "Understanding Privacy Knowledge and Skill in Mobile Communication." *Computers in Human Behavior* 38: 296–303.

Popescu, M., and L. Baruh. 2013. "Captive But Mobile: Privacy Concerns and Remedies for the Mobile Environment." *The Information Society* 29 (5): 272–86.

Privacy International (PI). 2006. *A Race to the Bottom: Privacy Ranking of Internet Service Companies. A Consultation Report*.

Robinson, N., H. Graux, M. Botterman, and L. Valeri. 2009. *Review of the European Data Protection Directive*. Information Commissioner's Office (ICO), RAND.

Schwaig, K., G. Kane, and V.C. Storey. 2005. "Compliance to the Fair Information Practices: How Are the Fortune 500 Handling On-Line Privacy Disclosures?" *ACM* 36 (1): 49–63.

Sheehan, K.B., and M.G. Hoy. 2000. "Dimensions of Privacy Concern Among Online Customers." *Journal of Public Policy & Marketing* 19 (1): 62–73.

The White House. 2012. *Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy. The Consumer Privacy Bill of Rights*. February 23. http://www.whitehouse.gov/sites/default/files/privacy-final.pdf.

Turow, J. 2001. *Privacy Policies on Children's Websites: Do They Play by the Rules*? Report No. 38 of the Annenberg Public Policy Center, March.

West, D., and E.A. Miller. 2006. "Digital Divide in Public e-Health Sites: Barriers to Accessibility and Privacy in Health Department Websites." *Journal of Health Care for the Poor and Underserved* 17 (3): 652–66.