

Digital Literacy and Privacy Behavior Online

Communication Research
40(2) 215–236
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0093650211418338
crx.sagepub.com



Yong Jin Park

Abstract

This study examined the impact of three dimensions of digital literacy on privacy-related online behaviors: (a) familiarity with technical aspects of the Internet, (b) awareness of common institutional practices, and (c) understanding of current privacy policy. Hierarchical regression models analyzed data from a national sample of 419 adult Internet users. The analyses showed strong predictive powers of user knowledge, as indicated by the three discrete dimensions, on privacy control behavior. However, the findings were mixed when accounting for the interaction between knowledge and Internet experiences. There were limitations on the extents of knowledge and action related to personalized information. Furthermore, those limitations divided with sociodemographic characteristics such as age, gender, income, and education. Ramifications for the current status of the FTC policy are discussed.

Keywords

Internet privacy, information control, surveillance, skills, digital divide

Introduction

The purpose of this study is to examine the role of digital literacy in the control of personal information online. The central question is whether and to what extent user knowledge functions as an enabler for active information control on the Internet. The investigation is twofold. First, the study examines the function of knowledge in new media use and privacy control behavior. Second, the analysis aims to identify the locus of any digital divide that may be prevalent among the public, with particular concern about user skills controlling inappropriate surveillance. On the deeper level, this study questions the Federal Trade Commission (FTC) Internet policy grounded on the assumption of an omni-competent user (FTC, 2009).

Corresponding Author:

Yong Jin Park, Howard University, 3735 Mazewood Lane, Fairfax, VA 22033, USA.
Email: yongjinp@hotmail.com

¹Howard University, Washington, DC, USA

The explicit consideration of knowledge is significant when considering the consequences of digital inequalities in users' ability to protect themselves from undue surveillance. Theoretically, it is valuable to contextualize Internet privacy in the digital divide debate by bridging the two fields that already moved beyond the concern of online access at the infrastructural level. In fact, as most civic lives migrate to online platforms, citizens are increasingly open to data surveillance, collection, and appropriation against which, under the FTC policy, the individual is in essence the sole guardian. Overall, the way in which to empower users against surveillance carries enormous practical and policy values.

The current study begins with a brief review of the literature on digital literacy, discusses prior studies, and poses research questions and hypotheses. Then, the study analyzes the extent of public knowledge in relation to the control of personal information and accounts for potential patterns of effects.

Digital Literacy and Privacy

Literacy is one of the most fundamental human conditions in diffusing democratic potentials (Pool, 1983). Neuman (1991) highlighted the centrality of literacy in promoting participatory orientation and developing a viable civic culture. The notion of digital literacy describes individual knowledge regarding computer-related functions (Bunz, 2004; Dutton & Anderson, 1989; Jenkins, 2006). According to Hargittai (2002, 2004), a second-level digital divide, the difference in user skills, impairs the democratic potential of the Internet as political and civic activities move online. In this sense, it is crucial to identify the genesis of digital divide at the user level, that is, what distinguishes differentiated uses in various aspects of Internet use. Scholars consistently point out that individual differences in cognitive ability may well explain different types of digital media skills (Freese, Rivas, & Hargittai, 2006; Hargittai & Hinnant, 2008). Furthermore, some research suggests that different levels of expertise can promote or inhibit users in specific domains, such as personalized data use and control (Hargittai, 2007).

Recently, the hotly contested privacy settings on social networking sites, such as Facebook, focused serious scholarly attention on personalized data control (e.g., boyd & Hargittai, 2010; Debatin, Lovejoy, Horn, & Hughes, 2009; Fogel & Nehmad, 2009). For instance, boyd and Hargittai (2010) reported a substantial number of young Facebook users were aware of and concerned about potential privacy threats, contrary to the wide misconception that young people do not care about privacy. However, Lewis, Kaufman and Christakis (2008) documented the behavioral patterns of publicly displaying personal profiles among college students. Solove (2007) suggested that given the delicate boundaries of social networking and online interaction, privacy control should be understood in terms of granular degree, not in absolute terms. Although advanced research in the field has emerged, few studies address the fundamental genesis of behavioral variations in predicting users' abilities to control personalized information. This gap warrants a systematic inquiry to explicate a predictive model of privacy behavior.

This study puts forth a new measure of digital literacy that focuses on privacy and online privacy-related behaviors. Goffman (1959) theorized that individuals should be able to manage or control private—public boundaries by selectively revealing one's identities (see also Agre, 1998; Bellotti & Samarajiva, 1998). Note the critical role of knowledge assumed to be in operation, that is, awarenesses of institutional systems and social practices may well equip individuals to take appropriate actions. In the digital era, the idea encompasses critical understanding of data flow and its implicit rules for users to be able to act. Literacy may serve as a principle to support, encourage, and empower users to undertake informed control of their digital identities. In short, to exercise appropriate measures of resistance against the potential abuse of personal data, it may be that users should be able to understand data flow in cyberspace and its acceptable limits of exposure (Ball & Webster, 2003).

Prior Studies

Earlier Efforts

Earlier privacy studies attempted to identify the extent of consumer understanding regarding various aspects of data surveillance. Central to these efforts was the posited function of knowledge in exercising information control. For example, Culnan (1995) observed the low level of awareness among U.S. consumers regarding the removal process from direct mailing lists. Milne and Rome (2000) also found a lack of procedural knowledge for name removal process despite the fact that most respondents indicated their intention to "opt out." In addition, Nowak and Phelps (1997) indicated wide uncertainties and misinformation prevalent among consumers about the practices of direct mail marketers.

In the Internet era, the practices by database marketers came to the forefront of the scholarly and policy concern. Research efforts sharpened to observe the consequences of knowledge. Dommyer and Gross (2003), for instance, found significant associations between users' levels of awareness concerning privacy protection strategies and their telephone directory "opt out" status. Graeff and Harmon (2002) suggested an explicit connection between demographic variables and the level of knowledge. In addition, a Pew Survey (Fox, 2000) measured the knowledge of Internet users among different segments of the online population with its general finding, including a low level of familiarity with "cookies," confirmed in subsequent studies (e.g., Pew, 2007).

Culnan and Armstrong (1999) reported that user awareness of fair procedural practices in websites alleviated the levels of privacy concern. A study by Hoffman, Novak, and Peralta (1999) indicated that users, when explicitly aware of malpractices by sites, tend not to disclose information. These findings were significant in that they illustrated a dichotomy between the stated concern and behaviors (Park, 2008; Sheehan & Hoy, 1998) that could be potentially moderated through increased knowledge. However, most findings were limited because they measured a single variable, such as familiarity with "cookies," as a proxy variable for user knowledge (e.g., Pew, 2000). Furthermore, different measurement scales based on convenience samples (e.g., Lewis et al., 2008; Park, 2008) made it difficult to generalize the findings in any specific causal directionality.

Refined Studies

In a series of carefully designed surveys, Turow (2003; Turow, Feldman, & Meltzer, 2005) advanced this line of research. What Turow contributed is the sophistication of the measures that observed the public understandings of data practices by websites. In 2003, the first national sample survey found two alarming facts concerning (a) the widespread ignorance among the public regarding the fundamental aspects of data flow and (b) the lack of protective steps taken on the part of consumers. According to Turow (2003), this was particularly startling because the cognitive power of the users remained limited in contrast to the advances in institutional surveillance techniques. The second survey by Turow et al. in 2005 confirmed these findings. The updated study further identified the significant association between demographic characteristics, such as education, gender, age and income, and the lack of knowledge, attesting to the presence of a "knowledge gap" among users of different population segments. Some respondents reported falsifying information when they were explicitly aware of data surveillance. However, the level of misunderstanding remained wide. Most consumers misunderstood the mere presence of a privacy policy statement as data protection and performed few informed cost-benefit decisions about potential data misuse and surveillance.

Acquisti and Gross (2006) narrowed the discussion to the interactive environment of a social networking site. In their Facebook study, they found that most member-users were unaware of internal data-collection rules, regardless of their different levels of concern and their frequency of site use. Although some managed their privacy, it was with limited (or misinformed) awareness of the visibility of their personal data. Furthermore, their levels of protective skills were highly limited despite the fact that most *did* adopt one or two strategizing behaviors (see also Acquisti & Grosslags, 2005; LaRose & Rifon, 2007; Metzger, 2004). This finding was critical because it linked low levels of knowledge and low levels of data management in a highly interactive environment. In addition, the reaffirmation of inattentive media use entails further examination of the individual decision-making process, that is, how inattentive user habits can turn into the practices of active information control.

A Step Further

It is important to note that Turow (2003; Turow et al., 2005) placed user knowledge and use in a broad context of social differentiation, whereas Acquisti and Gross (2006) analyzed information processing at the individual level. Combined, the contribution from both works offered significant understanding regarding the decision-making process in Internet privacy behaviors. In essence, the lack of knowledge about the extent of data flow is posited as a hindrance to the complex decision-making process, whereas the socioeconomic divide possibly remains the genesis of such limits.

Nevertheless, most prior studies rarely identified the consistent role of knowledge on empirical ground. First, knowledge assessments were limited to one-dimensional measures that relied on a single item (e.g., Fox, 2000) and failed to capture the diverse dimensions of cognitive structure. Second, analytically, prior studies did not explicitly test the relationship

between knowledge and differentiated uses with multivariate regression (e.g., Acquisti & Grosslags, 2005; Acquisti & Gross, 2006), despite the posited function of knowledge in the decision-making process. Third, drawing from the strategic marketing literature, a majority of earlier studies rarely advanced any consistent theoretical basis that explained the linkage between knowledge and new media behavior.

This is not to imply that prior research has not been useful. Indeed, the accumulation of the findings serves as a superb reference point to better predict the function of user knowledge. Warranted, however, is an analytical model that integrates individual cognitive differences into the broader context of the social divide and contributes advanced theoretical understanding of the second-level digital divide in terms of information privacy behavior (Hargittai, 2007).

Hypotheses and Research Questions

In sum, this study tests the explicit premise of digital literacy as applied to Internet privacy and identifies the function of knowledge in strategizing behaviors. In general, researchers theorized that critical understanding is required of citizens to participate in digital activities (Jenkins, 2006; van Dijk, 2005). Specifically, it follows that the more knowledge users have about data flow, the more equipped they will be to manage to control (Barnes, 2006; Turow et al., 2005). Conversely, the less aware users are, the more they are susceptible to manipulation and unable to act and control information flow in their best interest.

Hypothesis 1 (H1): Users with a high level of knowledge are more likely to exercise information control than those with a low level of knowledge.

Research Question (RQ1): To what extent are the users aware of online surveillance practices?

Research Question (RQ2): To what extent do the users exercise control over personal information?

When positing the function of knowledge, however, the discrete dimensional structure of knowledge should be beyond the simple bipolarity of presence or absence. In fact, scholars (e.g., Hargittai, 2004; Kwak, 1999; Neuman, 1986) investigated the different impacts of knowledge as multidimensional measures to capture the dynamics of human behaviors. Previous studies noted that technical familiarity had positive effects on digital media and Internet uses (Hargittai, 2004; Hargittai & Hinnant, 2008). Furthermore, in other domains, Turow (2003; Turow et al., 2005) recognized the centrality of awareness of behavioral marketing practices and related policy environment as empowering consumers. This led to the following hypotheses on the subtlety of knowledge dimensions:

Hypothesis 1a (H1a): Users with a high level of technical familiarity are more likely to exercise information control than those with a low level of knowledge.

Hypothesis 1b (H1b): Users with a high level of institutional surveillance awareness are more likely to exercise information control than those with a low level of knowledge.

Hypothesis 1c (H1c): Users with a high level of policy knowledge are more likely to exercise information control than those with a low level of knowledge.

In predicting information behavior, it is also critical to note that variations in Internet access experience, such as years of use and daily use, had significant impacts on the levels of skills in various aspects of the Internet (Hargittai, 2002, 2004). The freedom to use the Internet anytime, anywhere, and for any purposes was also one of the most significant single predictors for levels of online skills. As applied to personal information control, the effects of various levels of access experience may contribute to the digital divide, keeping levels of user knowledge constant.

Hypothesis 2 (H2): Internet access experiences will be positively associated with the levels of information control skills.

In addition, differences in socioeconomic status (SES) may negatively affect skills. Prior studies (Castells, 1996; DiMaggio, Hargittai, Neuman, & Robinson, 2001; Hargittai, 2002, 2004; Loges & Jung, 2001) consistently pointed out the role of SES in maintaining different levels of digital divide in skills. Also, there was evidence that in personal privacy (Turow et al., 2005; Turow & Hennessy, 2007) offline status, such as age, gender, income, and education, affects privacy protection behavior.

Hypothesis 3 (H3): There will be significant associations between users' sociodemographic status and their information control skills.

Hypothesis 3a (H3a): Income will be positively associated with the levels of information control skills.

Hypothesis 3b (H3b): Education will be positively associated with the levels of information control skills.

Hypothesis 3c (H3c): Age will be negatively associated with the levels of information control skills.

Hypothesis 3d (H3d): Gender (female: higher) will be negatively associated with the levels of control skills.

Finally, it is reasonable to assume the presence of interaction effects between the first-level predictor (Internet experience) and the second-level predictor (knowledge), while controlling for other demographic characteristics. These respective digital divide predictors, when intertwined, may deepen existing skill gaps. The effect of knowledge may or may not be present in moderating or accelerating influences of various aspects of Internet access experiences.

Research Question 3 (RQ3): Is there an interaction effect between the first- and second-level predictors?

In estimating the multivariate influences, a summary of the proposed model follows:

Y (information control) = $c + \beta 1$ (sociodemographic status) + $\beta 2$ (infrastructure-level Internet experience) + $\beta 3$ (individual-level knowledge) + $\beta 2*\beta 3 + e$

Method

The Study Population

The study examined a national probability sample of 419 adult Internet users (aged 18 and above). The Knowledge Networks (KN) recruited the panel respondents, using random digit dialing (RDD). The cross-sectional data included adult Internet users with online access at home, eliminating web-TV-based panel participants. The initial sampling frame included both listed and unlisted phone numbers and was not limited to computer owners. Once a household was randomly contacted by phone, KN recruited a household member(s) to the panel from which the survey participants were selected by chance. The panel participants were directed to a survey site and completed an online survey, which took about 10 minutes for completion. Administration of the survey occurred between October 31 and November 12, 2008.

The demographic characteristics of the KN panel were not far different from those of the general population. For the exclusive Internet user KN panel, however, a more appropriate baseline would be a nationally representative sample of U.S. Internet users. Table 1 presents descriptive statistics about the sociodemographic characteristics of the respondents in comparison to a Federal Communications Commission (FCC) 2010 wired and wireless Internet survey sample. The KN panel closely aligned with the Internet user profile of the FCC sample. However, the levels of income and age, on average, were slightly higher in this study sample. The time disparity between the two samples (2008 and 2010) may account for the differences as broadband became more affordable and widely diffused. Given the extent to which older and more affluent user groups are present in the sample, however, readers should use caution in making generalizations based on this study's findings. The total sample size was 456 from 663 initial contacts, with a completion rate of 69%. The final data set was limited to 419 after an item validity check.²

Measures

Knowledge: Independent variable. Digital privacy literacy was operationalized as user awareness in three dimensions: (a) technical familiarity, (b) awareness of institutional practices, and (c) policy understanding. Technical familiarity was rated with five items on a 6-point scale ($1 = not \ at \ all$, $6 = very \ familiar$). Eight true-false knowledge items were used for surveillance awareness and seven true-false items were measured for policy understanding, later coded 1 for correct answers with 0 assigned to all other responses. Each dimension of user knowledge was combined to create an index, adapted from prior studies (e.g., Hargittai & Hinnant, 2008; Pew Internet, 2007; Turow, 2003; Turow et al., 2005; $\alpha = .82$, technical knowledge; *Kuder-Richardson* 20 reliability = .79, surveillance awareness;

			FCC broadband 2010 (N = 3,005)			
	KN panel 2008 (N = 419)		Internet user		Total	
	М	SD	М	SD	М	SD
Education	2.97	0.93	2.96	0.93	2.63	1.02
Age	46.34	16.24	42.52	15.83	46.69	17.99
Income	6.07	1.86	5.48	2.26	4.84	2.44
Female	male 53.6%		50.2%		51.7%	
Internet experi	ences					
Number of Internet use	nternet use:Yea	rs of	1	1.06		4.41
Number of Internet use: Minutes of daily Internet use		297.51		303.54		
Number of Internet access sites		2.32		1.31		
Type of Inter	net connection					
Dial up				8.5%		
Cable	Cable		37.7%			
DSL			35.3%			

Table 1. Main Characteristics of Study Participants (N = 419).

Note: FCC = Federal Communications Commission. For gender, male was coded as 1 and female as 2. Education in both surveys was measured in four categories. Income in the KN panel was recoded into 9 categories to be equivalent to FCC 2010 May wired and wireless Internet survey.

Kuder-Richardson 20 reliability = .73, policy knowledge). This was to capture a whole dimension of data flow in the context of institutional surveillance practices. Two additional items (privacy-specific risk and protection awareness) specified technical familiarity to capture subtle knowledge structures that may be present in user behavior.

Information control behavior: Dependent variable. One of the main purposes in this study was to identify information control behavior as currently in daily routine. Information control was operationalized as user behavior in strategizing data release—that is, whether to opt out or not. Thus it was central to capture how users systematically manage/control personal data and its flow (that can be associated with one's identity). Information control is multifaceted in nature, requiring a combination of social and technical skills as intertwined in Internet uses (Marx, 2003; Resnick, 2002, for "sociotechnical" capital). Following this, preexisting survey items were elaborated into (a) social and (b) technical dimensions. Within the social dimension, we made a distinction between active and passive control to capture the subtlety of user behavior.³

Respondents were asked to report the extent to which they were involved in each of the information control behaviors on a 6-point scale, ranging from *never* to *very often*. Eight items were used for the social dimension and we measured four items for the technical dimension, modified from the extant literature (e.g., Acquisti & Gross, 2006; Marx, 2003;

Metzger, 2004; Pew Internet, 2007; Turow, 2003; Turow et al., 2005). Informed by the preestablished items, the survey established the criterion validity of each item. Each item was a question that asked (a) the type(s) of information strategies adopted and (b) the intensity, as indicated in the frequency of use, of such strategies. The composite index (summation of items) was created to construct a continuous scale for each dimension ($\alpha = .80$, social; $\alpha = .70$, technical dimensions).

Internet experience. Two items measured online experiences in daily routines as they were related to differentiated uses of the Internet (Hargittai, 2004, 2005, 2007). First, we asked how long (in minutes) Internet was used. Second, we measured the number of years of experience with Internet (Kwak, Skoric, Williams, & Poor, 2004). Hargittai and Hinnant (2008) also noted the predictive power of autonomy of use on user skills. A single item measured the number of Internet access locations for each respondent on a 6-point scale (1 = one, 6 = more than six), adapted from Hargittai and Hinnant (2008).

Sociodemographic characteristics. As noted above, this study aimed to assess the potential influence of offline sociodemographic characteristics on online skills. Four items (income, education, age, and gender) were used.

Analytical Strategies

The analyses proceeded as follows. First, descriptive data identified the overall trends in user knowledge and behavior. Second, a series of hierarchical (moderated) regressions tested the hypotheses for each of the knowledge dimensions, accounting for multilevel influences. Hierarchical regression is useful to identify explanatory powers of the variables in each level, while considering the order of the predicted causal priority. Age, gender, education, and income were included in the first block, with Internet experiences in the second block. A total of nine interaction terms between knowledge and Internet experiences were created for the final equations. The variables were standardized prior to entry in each block to reduce potential problems of multicollinearity (see Kwak et al., 2004).

Results

Identifying the Overall Trends

Table 2 shows the limited extent of user knowledge in all three dimensions. The users did possess basic understanding of acquisition and use of personal information online (M = 4.73, SD = 2.40). Yet what the result indicated was that more than 40% of the respondents misunderstood the most basic aspects of institutional data practices. Only eight respondents (1.9%) scored correctly on all of the policy-related knowledge questions, with a miniscule mean score of 1.96 (SD = 1.86). Furthermore, a majority of the respondents reported low levels of familiarity with basic technical terms (M = 15.05, SD = 6.25).

The second block in Table 2 shows (a) the type and (b) the intensity for each dimension of information control. Overall, the sample respondents adopted one or more types of control strategies. Nevertheless, the levels of personal information control were consistently

	М	SD	Min	Max
Digital privacy literacy				
Technical familiarity	15.05	6.25	3	30
Awareness of institutional surveillance	4.73	2.40	0	8
Policy understanding	1.96	1.86	0	7
Information control skill				
Social dimension	24.81	9.18	5	48
Tech dimension	13.12	5.18	I	24

Table 2. Descriptive Statistics of Main Variables Used in Analyses.

low. In the technical dimension, the mean score was 13.12, indicating most users rarely adopted or used technology through either web browser or privacy enhancing technologies (PET). In the social dimension, the public involvement remained moderate. The users exercised relatively high levels of information control in terms of (a) withdrawal, (b) hiding, and (c) avoidance (M = 13.36, SD = 5.13). However, in the dimension of (a) complaint, (b) rectification, and (c) multiple account use, most respondents reported low levels of involvement (M = 11.45, SD = 4.99). Tables 3 and 4 present the distributions of individual knowledge and behavior measures.

The extent of knowledge and behavior divided across different segments of the user population. In terms of knowledge, education was a consistent predictor for higher scores (r=.22, p<.01, technical familiarity; r=.14, p<.01, surveillance practice; r=.11, p<.05, policy understanding). Economic status, measured by income level, showed significant correlations (r=.12, p<.05, technical familiarity; r=.12, p<.01, surveillance practices). Older users scored consistently low in technical familiarity and surveillance practice (r=-.16, p<.01; r=-.07, p<.10), whereas female users also scored low in all three knowledge dimensions (r=-.21, p<.01; r=-.18, p<.01; r=-.23, p<.01).

There was no difference in the level of information control in terms of economic status. Furthermore, the level of education had no clear impact in either information control dimension. There was a gender difference in technical skills (r = -.16, p < .01); however, the consistent impact appeared in age (r = -.14, p < .01, social; r = -.15, p < .01, technical dimensions), displaying the most persistent presence of the age gap in information control behavior among sociodemographic factors.

Testing the Hierarchical Regression Model

In estimating multivariate influences, the proposed model hypothesized the effects of knowledge, Internet access and use, and sociodemographics for each block. Tables 5 and 6 show the results of analyses in the social and technical dimensions.

The findings revealed robust support for Hypothesis 1a, the positive role of knowledge as measured by technical familiarities, in both social and technical dimensions (β = .26, p < .001; β = .46, p < .001). The support was strong and the knowledge block alone

 Table 3. Distribution of Individual Knowledge Measures.

	М	SD
Technical familiarity		
Generic Internet		
HTML	3.35	1.86
Preference setting	4.10	1.44
ISP	3.11	1.88
Cache	1.60	0.90
BCC (on email)	2.93	1.79
Privacy risk		
Phishing	3.12	2.16
Privacy protection		
p3p	1.47	0.88
Surveillance practices		
Companies today have the ability to place an online advertisement that targets you based on information collected on your web-browsing behavior	0.75	0.43
A company can tell you that you have opened an email even if you do not respond	0.57	0.49
When you go to a website, it can collect information about you even if you do not register	0.65	0.47
Popular search engine sites, such as Google, track the sites you come from and go to	0.66	0.47
E-commerce sites, such as Amazon or Netflix, may exchange your personal information with law enforcement and credit bureau	0.45	0.49
What a computer user clicks while online surfing can be recorded as a trail	0.72	0.44
Most online merchants monitor and record your browsing in their sites	0.68	0.46
When a website has a privacy policy, it means the site will not share your information with other websites or companies	0.25	0.43
Policy understanding		
Government policy restricts how long websites can keep the information they gather about you	0.20	0.40
It is legal for an online store to charge different people different prices at the same time of day	0.22	0.41
A website is legally allowed to share information about you with affiliates without telling you the names of the affiliates	0.40	0.49
By law, e-commerce sites, such as Amazon, are required to give you the opportunity to see the information they gather about you	0.14	0.35
Privacy laws require website policies to have easy to understand rules and the same format	0.20	0.40
U.S. government agencies can collect information about you online without your knowledge and consent	0.56	0.49
When I give personal information to an online banking site such as citibank.com, privacy laws say the site has no right to share that information, even with companies it owns	0.22	0.41

Table 4. Distribution of Individual Skill Measures.

		М	SD
Social dimension			
Avoidance	Stopped visiting particular websites because you fear they might deposit unwanted program on your computers	3.21	1.85
Hiding	Given false or inaccurate email address or fake name to websites because of the privacy concern	2.54	1.73
Withdrawal I	Decided not to make an online purchase because you were unsure of how information would be used	3.42	1.72
Withdrawal 2	Chose not to register on a website because it asked you for personal information to get into the site	4.28	1.63
Complain	Complained to a consumer or government agency about marketing practices of particular websites	1.50	1.07
Rectify I	Asked a website to remove your name and address from any lists used for marketing purpose	3.51	1.82
Rectify 2	Asked not to share your personal information with other companies	3.58	1.97
Multiple accounts	Used an email address that is not your main address, in order to avoid giving a website real information about yourself	2.89	1.97
Tech dimension	,		
Clearing history	Cleared your web browser history	3.49	1.81
Filtering emails	Used filters to block or manage unwanted email	4.56	1.90
Erasing cookies	Erased some or all of the cookies on your computer	3.68	1.90
Using PET software	Used software that hides your computer's identity from websites you visit	1.41	1.48

(incremental R^2) accounted for .051 and .147 in each dimension. However, the support was mixed when knowledge was specified in terms of privacy-specific familiarity. In a separate hierarchical regression that accounted only for privacy risk awareness (phishing), the support was evident. Yet the level of privacy protection knowledge (that is, familiarity with p3p) alone offered no support in the combined social index measure.⁴ Furthermore, in the tech dimension, familiarity with p3p was found relatively weak in effect size and significance ($\beta = .15, p < .05$).

There was support for the posited association between knowledge and levels of information control when knowledge was indicated by users' awareness of data-surveillance practices (Hypothesis 1b). The support was robust across both social and tech dimensions $(\beta = .32, p < .001; \beta = .27, p < .001)$. As a block, the knowledge accounted for .091 and .066 of the variance (incremental R^2). The regression results also supported Hypothesis 1c, which indicated knowledge in terms of policy understanding. Strong support existed for the social dimension ($\beta = .29, p < .001$) with consistent support for the tech dimension ($\beta = .19, p < .001$).

Table 5. Predictors of Information Control: Social Skill.

	В	t value
Sociodemographics		
Income	06	-1.21
Education	.09	1.91
Age	16	-3.37**
Gender (high: female)	03	-0.74
R ² (%)		.035
Internet experience		
Years of use	.19	3.90***
Daily use (logged)	.17	3.56***
Autonomy	.09	1.93
Incremental R ² (%)		.128
Digital privacy literacy		
Technical familiarity	.26	4.98***
Incremental R ² (%)		.178
Awareness of institutional	.32	6.88***
surveillance		
Incremental R ² (%)		.219
Policy understanding	.29	6.38***
Incremental R ² (%)		.208

Note: Entries are standardized regression coefficients after controlling for the control variables. The coefficients in Block 3 were the results of separate hierarchical regression models while the variables in prior Blocks remained constant.

To examine Hypothesis 2 (positive associations between Internet experiences and the level of information control), four items were analyzed. The second blocks in Tables 5 and 6 show that the supports from year of use and daily use were consistent in both dimensions ($\beta = .19$, p < .001; $\beta = .17$, p < .001, social; $\beta = .19$, p < .05; $\beta = .14$, p < .01, tech). Autonomy was significant only in the tech dimension ($\beta = .16$, p < .01). For Hypothesis 3, the impact of age remained consistent and significant as the hierarchical model provided support for both social and tech dimensions ($\beta = -.16$, p < .01; $\beta = -.16$, p < .01). Regarding gender, there was no difference in the social dimension, but a significant difference in the tech dimension ($\beta = -.14$, p < .01), indicating male users tended to exercise more information control in that aspect. The influences of income and education did not reach significance levels.

To examine RQ3, the interactions between the first- and the second-level predictors were analyzed after controlling for all prior blocks (see Table 7). In the social dimension, a significant interaction between technical familiarity and daily use was present ($\beta = -.40$, p < .01). In the tech dimension, technical familiarity interacted with year of use, daily use, and autonomy ($\beta = -.50$, p < .001; $\beta = -.24$, p < .05; $\beta = -.36$, p < .01). Figure 1 displays

 $^{0.00. &}gt; q^{**} \cdot 10. > q^{**}$

Table 6. Predictors of Information Control: Tech S	Table 6.	Control: Tech Skill.	f Information
---	----------	----------------------	---------------

	В	t value
Sociodemographics		
Income	04	-0.89
Education	.05	1.08
Age	16	-3.42**
Gender (high: female)	14	−3.07**
R^{2} (%)		.052
Internet experience		
Years of use	.19	3.99***
Daily use (logged)	.14	2.98**
Autonomy	.16	3.43**
Incremental R ² (%)		.157
Digital privacy literacy		
Technical familiarity	.45	9.23***
Incremental R ² (%)		.304
Awareness of institutional surveillance	.27	5.81***
Incremental R ² (%)		.223
Policy understanding	.19	4.06***
Incremental R ² (%)		.190

Note: Entries are standardized regression coefficients after controlling for the control variables. The coefficients in Block 3 were the results of separate hierarchical regression models while the variables in prior blocks remained constant.

Table 7. Interactions Between the First- and Second-Level Digital Divide Predictors

	Social skill		Tech skill	
	В	t value	В	t value
Prior blocks R ² (%)		.178		.304
Technical familiarity × Years of use	16	-1.38	50	-4.45***
Technical familiarity × Daily use	40	-3.04**	24	-1.99*
Technical familiarity × Autonomy	02	-0.23	36	−3.39***
Prior blocks R ² (%)		.219		.223
Awareness of institutional surveillance × Years of use	09	-1.02	00	-0.68
Awareness of institutional surveillance × Daily use	25	-2.53*	11	-1.16
Awareness of institutional surveillance \times Autonomy	10	-0.87	16	−I.40
Prior blocks R ² (%)		.208		.190
Policy understanding × Years of use	08	-1.27	07	-1.10
Policy understanding × Daily use	16	-2.44 *	09	-1.41
Policy understanding × Autonomy	.04	0.56	11	−I.67

Note: Entries are standardized regression coefficients after controlling for the control variables. Prior blocks include all the predictor variables analyzed in Tables 5 and 6.

^{.100. &}gt; q^{***} .10. > q^{***}

p < .05. *p < .01. **p < .001.

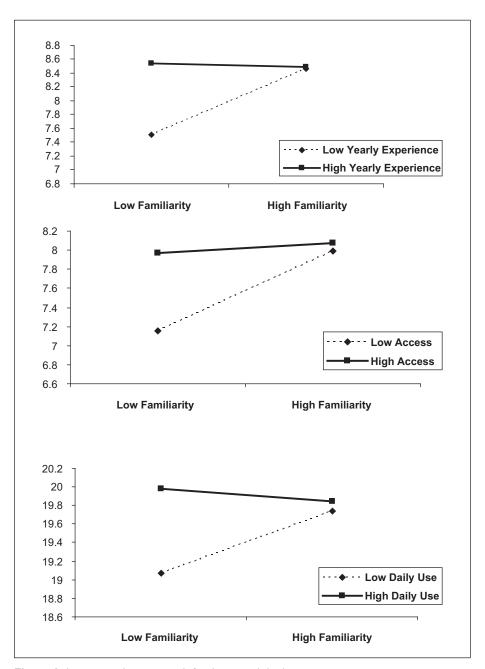


Figure 1. Interaction between tech familiarity and the Internet.

the consistent pattern of interactive effects of technical familiarity in both social and tech dimensions.⁵ Awareness of data surveillance and policy understanding interacted with daily use in the sociodimension ($\beta = -.25$, p < .05; $\beta = -.16$, p < .05). However, the support was far from robust because none of the interaction terms showed significance in the tech dimension

Discussion

In this study, our aim was to examine the impact of digital literacy on new media behaviors in a predictive model. The focus was on the locus of digital divide, with particular concern on user skills and capacities to control their personal information in an increasingly digital world. We included more nuanced measures of knowledge and information control in the discrete dimensions. The extent of knowledge effect was tested, taking into account the multilevel influences in hierarchical models. Interactions between Internet access experiences and knowledge were also observed.

The findings supported the hypotheses that derived from digital divide literature and earlier empirical privacy studies. First, the findings supported the hypothesized functions of technical familiarity, surveillance awareness, and policy understanding on personal information control behavior. However, the extent of knowledge and action remained limited, divided by sociodemographic status. The impact of age was the most consistent in explaining information control behavior (see Freese et al., 2004). In sum, the main thesis was robust, showing knowledge likely supported, encouraged, and empowered users to action.

The role that each knowledge dimension played, however, was subtle (e.g., Hargittai, 2004; Neuman, 1986). The familiarity with p3p, the most direct item that concerned Internet privacy protection, provided no or modest support (albeit, the lack of support is likely due to its low variance), while the familiarity with generic Internet terminologies offered greater significance. Furthermore, the findings were mixed when accounting for the interactions between knowledge and Internet experiences. While technical familiarity in interaction with Internet experiences provided support in both dimensions of information control, there was little support for such interactions with surveillance awareness and policy understanding in the social dimension and no support at all for the tech dimension.

This suggests that the cognitive dimensions may be highly correlated but operate in subtle and slightly different behavioral contexts (Figure 2), far from producing monolithic effects of knowledge. At least from this study, it is clear that generic technical familiarity functions as the most significant predictor of personal information control as its explanatory power is supported in other Internet uses, such as online content creation and sharing (e.g., Hargittai, 2002, 2004; Hargittai & Hinnant, 2008). Further studies are needed to verify this subtle operation of a particular type of knowledge. In the delicate dynamics of the movement from knowledge to concrete actions, experimental studies may extract differentiated behavioral routes unique to privacy-specific risk and protection awareness.

The results also shed light on a value of the multilevel model that derives from new media literature. Each level of predictors, (a) knowledge, (b) Internet experiences, and (c) sociodemographics, were significant, advancing analytical understanding of privacy-related behaviors

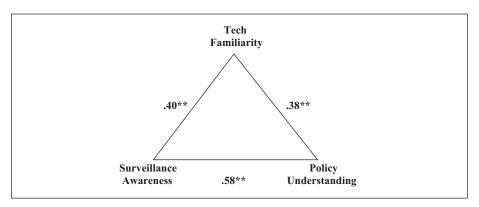


Figure 2. Knowledge structure: Correlations among dimensions. Note: Correlation is significant at the .01 level (two-tailed).

online. This demonstrates the value of merging insights from new media studies with the Internet privacy literature. As Turow (2003) noted, a rather striking ignorance of online data flow practices may lie at the heart of public inaction.

Finally, the presence of the age gap deserves serious attention. The wide divide of age in information behavior is not surprising. However, the fact that older users are less skillful than are younger people in privacy control creates a grim scenario in which they may be the worst victims of identity theft or related online crimes. The social embarrassment of being "foolish" may further dampen older users' enthusiasm in seeking help or learning privacy-related technology. Gender in this regard also raises concern. The results indicate female users score lower than male counterparts in technical knowledge and behaviors. Yet a significant body of literature (e.g., boyd & Hargittai, 2010) reported no gender difference in online activities, while a few studies (e.g., Fogel & Nehmad, 2009) indicated female users exercised even more privacy control than male users did on social networking sites. Thus one should interpret this study's finding with caution until further studies establish the persistence of the gender gap in the tech-related information control behavior.

With regard to the interaction effects, the large effect of Internet access experience among those with low levels of technical familiarity merits further scrutiny. This may mean that users with a low level of technical familiarity still benefit from high-tech experience and access, whereas those with no adequate infrastructural access cannot when they remain ill informed. This is a significant finding that suggests a certain type of knowledge compensates for lack of Internet access experiences in encouraging privacy-related behaviors. Put differently, the respective predictors of technical familiarity and Internet experiences, when combined, seem to magnify existing privacy-related skill gaps.

Another important finding is the resilience of mass inaction and lack of knowledge. In the last few years, new media research (Hargittai, 2007; Hargittai & Hinnant, 2008) has shown that high variations in skill levels persisted despite the surge in online access. The privacy literature has also demonstrated that the typically low levels of knowledge have not changed over time (Turow, 2003; Turow et al., 2005). More so, it confirms the replication of the persistent divide in online activities and cognitive skills among different segments in the privacy domain (DiMaggio et al., 2001). Note that the sample surveyed for the current study included a group of Internet users with high access experiences, a relatively high education level, and broadband Internet connection. Thus the findings regarding income and education are likely conservative. Given such tech-rich experiences, it is also surprising to observe the extent of lack of understanding and action, with the significant connection between two.

Conclusion and Policy Implication

The findings carry significant ramifications for the current FTC policy. First, this study demonstrates the presence of a second-level digital divide in Internet privacy beyond the level of access. It is important to note that the FTC policy assumes that users are typically well informed about data flow and are capable of appropriate responses (Turow, 2003). However, evidence suggests the presence of a digital literacy divide that may function as an impediment to systematic information control, further reinforcing the socioeconomic and demographic divisions in an increasingly digital world (DiMaggio et al., 2001).

In fact, the FTC (2009) consistently prioritized the rollout of free personal information flow over stricter online regulations of data collection, retention and uses, assuming the balance of power between websites and individual users. Yet the users are stratified and far from competent in exercising privacy control, different from such policy premise. Furthermore, while knowledge plays a critical role in privacy behavior, the levels of understanding of surveillance practices common in websites remain miniscule among the majority of users. This shows that the policy assumption regarding most online users may be fundamentally flawed.

Given the unique nature of the data set for a national sample, the findings should serve as a departure point to recognize the function of discrete dimensions of knowledge. It remains unclear why privacy-specific familiarity provided less explanatory power than generic Internet tech familiarity in the multilevel model, although it may be likely that the extremely low variation of p3p (M = 1.47; SD = 0.88) reduces its power. In addition, the causality could hardly be ascertained due to the cross-sectional nature of the survey data. Nevertheless, a novel contribution of this project is that the dimensional knowledge measures as independent variables are dissected to analyze information control skills in discrete levels. In this vein, a longitudinal panel study with the inclusion of more nuanced survey items will establish causal claims. Analytically, a two-stage model with larger cross-sectional sample data will parcel out precise temporal priority between knowledge and action.

A fruitful next line of research in this area may involve the psychological obstacles of Internet users, such as inconvenience and efficacy, which might deter even the most technically knowledgeable users from engaging in robust protective behavior. Respondent status within a household may also link to a level of engagement in information control because

the control of the computer in a household dictates who get engaged in protective behavior and to what extent. This will be critical in further inquiry to capture perhaps the most realistic setting of online privacy and personal information control.⁷

Digital literacy plays a central role in promoting and at times constraining active information control online (e.g., Freese et al., 2004; Hargittai, 2007; see Neuman, 1991). Some users are better positioned to exercise control, whereas others remain incapable of active role. Digital literacy should be promoted across all segments of the user population to encourage active and effective electronic participation in civic and economic life.

Acknowledgments

The author wishes to express his full gratitude to two anonymous reviewers for their helpful comments. Also, the author feels very grateful to Dr. Eszter Hargittai for her inspring talk at Michigan 2008 and to Dr. Joseph Turow for generously sharing his critical insights and survey instruments at the early stage of the development of this study. Finally, the sincerest gratitude goes to Dr. W. Russ Neuman and Dr. Scott Campbell at the University of Michigan for their continuous support.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Notes

- Lewis et al. (2008) base their analyses on data from the early years of Facebook, whereas boyd and Hargittai (2010) rely on data from later years. So the differences in their findings are likely due to the time-contextual gap between the two studies.
- 2. Following Hargittai (2009), we included a bogus knowledge item to check the validity of the response. Thirty-seven people reported a high level of familiarity with this item. Furthermore, these respondents reported extremely inflated scores in other key variables (knowledge, behavior, and attitudes), challenging the validity of the responses by this particular group.
- 3. The social dimension is the combined score of the passive and the active information control items. Passive control behaviors included (a) withdrawal (two items), (b) avoidance, and (c) hiding identities, whereas active behaviors were (a) complaining, (b) rectifying (two items), and (c) using multiple accounts (see Marx, 2003 for the typology of surveillance neutralization moves).
- 4. In the social dimension, $\beta = .19$, t value = 3.62 was found for privacy risk awareness (phishing, M = 3.51, SD = 1.76). However, privacy protection awareness (p3p, M = 1.47, SD = 0.88) generated no support ($\beta = .06$, t value = 1.18). When the social dimension was parceled out, very weak support was found in the active dimension ($\beta = .09$, t value = 1.96, t value = 1.04), with no support for the passive dimension of information control ($\beta = .06$, t value = 1.24).

- 5. We plotted the interactions, using standardized coefficients in the final equations after controlling for other blocks. For the purpose of demonstration, the combination of 0 (low) and 1 (high) was assigned to each of the four groups.
- 6. We ran additional analyses to observe the interaction between sociodemographic status and Internet experiences. A total of 12 interaction terms were created and entered into the hierarchical models. We truncated the models to simplify the computations (e.g., Kwak, 1999) and did not include three discrete knowledge dimensions. In the social dimension, the significant interactions between education and (a) daily use ($\beta = -.09$, t value = -1.95, p < .001) and (b) autonomy ($\beta = .15$, t value = 3.26, p < .05) were found. Age also interacted with year of use ($\beta = -.12$, t value = -2.47, p < .05). In the tech dimension, only gender interacted with year of use, but support was weak ($\beta = .08$, t value = 1.85, p < .10) with none of the other interaction terms found significant.
- 7. The author gratefully incorporated this critical insight suggested by an anonymous reviewer.

References

- Acquisti, A., & Grosslags, J. (2005). Privacy and rationality in decision making. *IEEE Security and Privacy*, 3(1), 26-33.
- Acquisti, A., & Gross, R. (2006). Imagined communities: Awareness, information sharing, and privacy on the Facebook. Retrieved from http://www.heinz.cmu.edu/~acquisti/papers/acquisti-gross-facebookprivacy-PET-final.pdf
- Agre, P. (1998). Technology and privacy: The new landscape. Cambridge, MA: MIT Press.
- Ball, K., & Webster, F. (2003). The intensification of surveillance. London, UK: Pluto Press.
- Barnes, S. (2006). A privacy paradox: Social networking in the United States. First Monday, 11(9). Retrieved from http://www.firstmonday.org/issues/issue11_9/barnes/index.html
- Bellotti, V., & Samarajiva, R. (1998). Interactivity as though privacy mattered. In P. Agre & M. Rotenberg (Eds.), *Technology and privacy: The new landscape* (pp. 277-310). Cambridge, MA: MIT Press.
- boyd, D., & Hargittai, E. (2010). Facebook privacy settings: Who cares? *First Monday, 15*(8). Retrieved from http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3086/2589
- Bunz, U. (2004). Growing from computer literacy towards computer-mediated communication competence: Evolution of a field and evaluation of a new measurement instrument. *Information Technology, Education, and Society*, 4(2), 53-84.
- Castells, M. (1996). The rise of network society. Oxford, UK: Oxford University Press.
- Culnan, M. (1995). Consumer awareness of name removal procedures: Implications for direct marketing. Journal of Direct Marketing, 9(2), 10-19.
- Culnan, M., & Armstrong, P. (1999). Information privacy concerns, procedural fairness and impersonal trust: An empirical investigation. *Organization Science*, 10, 104-115.
- Debatin, B., Lovejoy, J., Horn, A., & Hughes, B. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. *Journal of Computer Mediated Communication*, 15(1), 83-108.
- DiMaggio, P., Hargittai, E., Neuman, W. R., & Robinson, J. P. (2001). Social implications of the Internet. Annual Review of Sociology, 27, 307-336.
- Dommeyer, C. J., & Gross, B. (2003). Consumer knowledge, awareness, and use of privacy protection strategies. *Journal of Interactive Marketing*, 17(2), 34-51.
- Dutton, W., & Anderson, R. (1989). Computers and literacy: Differing perspectives in the social sciences. *Social Science Computer Review*, 7(1), 1-5.

Federal Communications Commission. (2010). May 2010 wired and wireless Internet survey. Retrieved from http://blog.broadband.gov/?entryId=479436

- Federal Trade Commission. (2009). Self-regulatory principles for online behavioral advertising: Tracking, targeting, and technology. Retrieved from http://www.ftc.gov/os/2009/02/P085400behavadreport.pdf
- Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concern. Computers in Human Behavior, 25(1), 153-160.
- Fox, S. (2000). Trust and privacy online: Why Americans want to rewrite the rules. Retrieved from http://www.pewinternet.org/~/media//Files/Reports/2000/PIP_Trust_Privacy_Report.pdf.pdf
- Freese, J., Rivas, S., & Hargittai, E. (2006). Cognitive ability and Internet use among older adults. *Poetics*, 34, 236-249.
- Goffman, E. (1959). The presentation of self in everyday life. New York, NY: Anchor Books, University of Edinburgh Social Sciences Research Centre.
- Graeff, T. R., & Harmon, S. (2002). Collecting and using personal data: Consumers' awareness and concerns. Journal of Consumer Marketing, 19, 302-318.
- Hargittai, E. (2002). Second-level digital divide: Differences in people's online skills. First Monday, 7(4).
 Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/942/864
- Hargittai, E. (2004). Internet access and use in context. New Media and Society, 6(1), 137-143.
- Hargittai, E. (2005). Survey measures of Web-oriented digital literacy. Social Science Computer Review, 23, 371-379.
- Hargittai, E. (2007). A framework for studying differences in people's digital media uses. In N. Kutscher & H. Otto (Eds.), *Cyberworld unlimited* (pp. 121-137). Wiesbaden, Germany: VS Verlag f
 ür Sozialwissenschaften GWV/Fachverlage GmbH.
- Hargittai, E. (2009). An update on survey measures of web-oriented digital literacy. Social Science Computer Review, 27(1), 130-137.
- Hargittai, E., & Hinnant, A. (2008). Digital inequality: Differences in young adults' use of the Internet. Communication Research, 35, 602-621.
- Hoffman, D., Novak, T., & Peralta, M. (1999). Information privacy in the marketplace: Implications for the commercial uses of anonymity on the web. *Information Society*, 15(2), 129-139.
- Jenkins, H. (2006, Winter). Game on! The future of literacy education in a participatory media culture. Threshold. Retrieved from http://web.mit.edu/cms/People/henry3/publications.html
- Kwak, N. (1999). Revisiting the knowledge gap hypothesis. Communication Research, 26, 385-413.
- Kwak, N., Skoric, M., Williams, A., & Poor, N. (2004). To broadband or not to broadband: The relationship between high-speed internet and knowledge and participation. *Journal of Broadcasting and Electronic Media*, 48, 420-445.
- LaRose, R., & Rifon, N. (2007). Promoting i-safety: Effects of privacy seals on risk assessment and online privacy behavior. Journal of Consumer Affairs, 41(1), 127-149.
- Lewis, K., Kaufman, J., & Christakis, N. (2008). The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of Computer Mediated Communication*, 14(1), 79-100.
- Loges, B., & Jung, J. (2001). Exploring the digital divide: Internet connectedness and age. Communication Research, 28, 536-562.
- Marx, G. (2003). A tack in the shoe: Neutralizing and resisting the new surveillance. *Journal of Social Issues*, 59, 369-390.

- Metzger, M. J. (2004). Exploring the barriers to electronic commerce: Privacy, trust, and disclosure online. *Journal of Computer-Mediated Communication*, 9(4). Retrieved from http://jcmc.indiana.edu/vol9/issue4/metzger.html
- Milne, G., & Rome, A. (2000). Consumer privacy and name removal across direct marketing channels: Exploring opt-in and opt-out alternatives. *Journal of Public Policy and Marketing*, 19, 238-249.
- Neuman, R. (1986). The paradox of mass politics: Knowledge and opinion in the American electorate. Cambridge, MA: Harvard University Press.
- Neuman, R. (1991). The future of the mass audience. Cambridge, MA: Cambridge University Press.
- Nowak, G., & Phelps, J. (1997). Direct marketing and the use of individual-level consumer information: Determining how and when privacy matters. *Journal of Interactive Marketing*, 11(4), 94-108.
- Park, Y. J. (2008). Privacy regime, culture, and user practices in the cyber marketplaces. Info, 10(2), 52-74.
- Pew Internet. (2007). Teens, privacy & online social networks: How teens manage their online identities and personal information in the age of MySpace. Retrieved from http://www.pewinternet.org/~/media//Files/Reports/2007/PIP_Teens_Privacy_SNS_Report_Final.pdf.pdf
- Pool, I. D. (1983). Technologies of freedom. Cambridge, MA: Belknap Press.
- Resnick, R. (2002). Beyond bowling together: Socio technical capital. In J. Carroll (Ed.), *HCI in new millennium* (pp. 242-272). Reading, MA: Addison-Wesley.
- Sheehan, K. B., & Hoy, M. G. (1998). Flaming, complaining, abstaining: How online users respond to privacy concerns. *Journal of Advertising*, 28(3), 37-51.
- Solove, D. (2007). *The future of reputation: Gossip, rumor, and privacy on the Internet.* New Haven, CT: Yale University Press.
- Turow, J. (2003). Americans and online privacy: The system is broken. Report of the Annenberg Public Policy Center, University of Pennsylvania, Philadelphia.
- Turow, J., Feldman, L., & Meltzer, K. (2005). Open to exploitation: American shoppers online and offline.Report of the Annenberg Public Policy Center, University of Pennsylvania, Philadelphia.
- Turow, J., & Hennessy, M. (2007). Internet privacy and institutional trust: Insights from a national survey. New Media & Society, 9, 300-318.
- van Dijk, J. (2005). The deepening divide: Inequality in the information society. London, UK: SAGE.

Author Biography

Yong Jin Park earned his PhD from the University of Michigan. Currently, he is an assistant professor, Radio, Television, Film, School of Communications at Howard University, Washington DC. Dr. Park examines social and policy implications of new technologies, with particular interest in underserved user communities.