Investigations into Consumers Preferences Concerning Privacy: An Initial Step Towards the Development of Modern and Consistent Privacy Protections Around the Globe

Eric K. Clemons
The Wharton School
clemons@wharton.upenn.edu

Josh Wilson
The Wharton School
wilsonjs@wharton.upenn.edu

Fujie JIN
The Wharton School
jinfujie@wharton.upenn.edu

Abstract

Online privacy is becoming an increasingly important topic, and an increasingly controversial one. The EU is imposing strict limitations on the use of data obtained from its citizens' online activities [9], while Big Data advocates and online advertisers in the United States are concerned that this may represent interference in their basic business models or even in international trade [13].

It is clear that laws and regulations are inconsistent across national borders. They are also inconsistent within nations, depending on the industry classification of companies, or even the designation given to specific technologies. ISPs are prohibited from reading subscribers' email; other information services companies can do so legally. Data stored electronically is offered protection that is denied to data stored in the cloud.

This paper proposes that regulatory confusion be addressed starting with some basic principles of uniformity. More importantly, it suggests that regulation be driven by what consumers actually want, and provides some preliminary research aimed at determining what consumers want from privacy regulation around the world.

1. Introduction

Consumer privacy legislation has received a great deal of attention globally. The EU, Japan, Korea, Malaysia, and many other nations are actively reviewing their policies towards online privacy[1]. The US has preferred to allow the online information services industry to regulate itself, and the FTC has convened a meeting of the W3C to determine if an acceptable policy could be developed without active involvement by the Congress, the FTC, the FCC, or other regulatory bodies [11,14,17]. To some extent, this activity has been prompted by concerns about what companies like Google, Facebook, and AT&T in the US, Softbank and NTT in Japan, Daum and KT in Korea, and other large information services and communications firms might be able to do.

Paradoxically, the recent (June 2013) surge in interest in, and outrage over, potential online privacy abuse in the United States has been prompted by revelations from a former CIA employee and former NSA consultant about widespread abuse of privacy by the National Security Agency. Allegedly the NSA has been monitoring our calls.

We say paradoxically because there has been widespread outrage over what the NSA has been doing [20,28,31], in the name of national security [32], with only limited supervision by the Foreign Intelligence Surveillance Court [30]. And yet, there has been virtually no concern in the US over what Google and Facebook have been doing, sometimes far more extensive than the privacy abuses of the NSA, in the interest purely of corporate gain, and with no supervision whatsoever. Indeed, there has been more concern over the NSA's alleged abuses than there was over Google's admitted violations of its own privacy policies [3,6] and more outrage over the NSA's alleged abuses than there was over Google's admitted violations of its signed consent decree after its initial violations.

Our intent with this paper is to provide a sound basis for public policy concerning privacy, in the US and abroad. In order to provide such a sound basis, we conducted surveys and focus groups, in the US, Japan, Korea, and German, concerning consumers' attitudes towards the privacy policies of large information services firms. We explored both consumers' ***awareness of specific potential activities*** in which these large information services firms might engage and consumers' ***approval of those activities***, whether or not firms actually engaged in them. We also explored consumers' attitudes towards the degree of privacy protection they believe they received from their regulatory systems. Finally, we explored parents' attitudes towards data mining of their children's email accounts, including school email accounts. This work as yet has been completed only in the US and Japan.

Our research is motivated by the following beliefs:

- Technology has rendered traditional relationship between personal identifiers and private information largely irrelevant. We discuss this in sec-

[1] On May 25, 2013 Peking University Law School hosted a workshop on the future of Internet regulation, with a significant focus on the future of privacy regulation. On May 30th and 31st Keio University and Korea University hosted a workshop on Big Data, with a session on the future of privacy regulation in Asia. On June 3, Tokyo University held a workshop on privacy and privacy regulation in Japan.

978-1-4799-2504-9/14 $31.00 © 2014 IEEE
DOI 10.1109/HICSS.2014.504

tion 3 below.

- Regulations are largely inconsistent and have not kept pace with technology driven changes in the marketplace for information services. There are two ways in which regulations have created a gap in which non-traditional information services companies act with limited regulatory oversight. We discuss this in section 4, below.

We believe that regulatory protections for consumers should be the same, regardless of the company or carrier being regulated. Regulations generally place meaningful restrictions on telecommunications companies and other common carriers, because their capabilities and the dangers to privacy that they represented were well understood. Laws, sometimes even constitutional protections, limit the ability of postal services and telecommunications companies to read their customers' messages, eavesdrop on their customers' conversations, or even use the logs of customers' conversations to track their behavior. These protections have been extended to protect forms of communication that were unanticipated by the framers' of these regulations, such as email messages carried by common carriers. In contrast, most information services companies, such as Facebook, Google, and Microsoft, face no similar restrictions on their use of customers' information.

We believe that regulatory protections for consumers should be simple, reasonable, and what the customer would expect. They should be driven by the need to provide meaningful and modern protection, and not driven by regulators' interpretation of and attempt to apply inapplicable legislation or precedent. Once again, regulations are largely inconsistent and have not kept pace with technology driven changes in the marketplace for information services. It is clearly reasonable that conversations overheard because the speaker made no attempt to protect the conversation cannot be considered private or protected. It is clearly ***not*** reasonable that, by extension, email conversations that have been overheard for the same reasons cannot be considered private or protected. There is no plausible rationale for arguing by extension that the text messages on a phone are not private because reading them is no different from overhearing a conversation [18]. There is likewise no plausible rationale for arguing that although email stored on a vendor's electronic storage is private and protected because it is covered by laws regulating the electronic storage of data, but the same email is not private and protected if stored in the cloud, because the cloud is not electronic storage [21,33].

Most importantly for the contribution of this paper, we believe research into the privacy protections that consumers want and expect online if fundamental to privacy policy. If privacy policy is to be simple, reasonable, and consistent with consumers' preferences, it must be informed by an understanding of those preferences. The principal contribution of our paper is an analysis of what consumers want and expect from privacy protection, based on our ongoing research around the world.

Our research findings can be summarized as follows:

(1) Consumers mostly do not know what search engines and other data intermediaries are doing with their data, and mostly they know that they do not know.

(2) Consumers mostly do not approve of the majority of the practices of online service vendors when they use consumers' data. This is captured through questions such as, "If it were true that Google did track your searches, would you approve or disapprove?" or "If it were true that your service provider did read your texts, would you approve or disapprove?" The vast majority of consumers disapproved or strongly disapproved of most forms of the use of their online data, no matter what activities resulted in the capture of that data.

(3) Consumers' silence regarding privacy practices and the use of their data by information services firms does not represent consent, and most certainly does not represent informed consent. We use the term "informed consent" to refer to those consumers who are aware of a practice online and who also approve of that practice. Our data suggests that informed consent is very limited, on the order of 0-1%, for all forms of online privacy abuse. The data are largely consistent across populations surveyed in the US, Japan, Korea, and Germany

(4) Consumers by overwhelming majorities support the position that actually protecting consumers' privacy online should be the default setting on browsers and email services, and on linking online activities with text or GPS information obtained from the user's phone. There was even stronger support for the position that any privacy settings a consumer had chosen to protect privacy, whether set explicitly by the consumer or set implicitly by accepting default settings, should be honored by online service providers. To be clear, consumers believed that the default settings should be do not track individual services and do not integrate across multiple services, and that these default settings should be honored by all online service providers.

(5) Consumers do not believe that regulators are doing enough to inform them about online risks to their privacy and consumers do not believe that regulators are doing enough to protect them from online risks to their privacy.

(6) Consumers have equally strong views about protecting the privacy of their children from data mining activities of information services vendors.

(7) Consumers' feel strongly about protecting the privacy of their children from data mining. They continue to feel strongly even if the email service provider offers email without charge in exchange to the right to perform data mining..

We do believe that public policy is important for two reasons:

(1) Many consumers, including the youngest consumers, do not appear to be fully aware of the privacy risks they create for themselves.

(2) Until the recent incident with the NSA, consumers did not appear to be greatly concerned with privacy risks, and even after the NSA disclosures, most consumers appear to be more concerned with the potential of governmental abuses stemming from governmental privacy intrusions than the potential for corporate abuses stemming from far more pervasive corporate privacy intrusions.

We suspect that dealing with this public policy explicitly is important, and that the results of this survey are important, because it contradicts the young, hip, blogosphere. When Microsoft tried to create public awareness of and public interest in the problems of privacy violation [26], TechCrunch and its readers found the subject just silly [29]. To the best of our knowledge, ours is the first study that attempts to gather opinion from a range of subjects.

This paper is structured as follows: Section 2 provides alternative definitions of privacy and of invasions of privacy. Section 3 provides a brief review of the regulatory gap in which most information services companies, other than common carriers and telecommunications providers, currently operate. Section 4 describes the "myth of anonymization"; despite popular belief to the contrary, anonymous ads are not anonymous, better targeted ads are not better for you, and clicking on targeted ads can hurt you. Section 5 describes our experimental design for our first survey, which assessed consumers' attitudes towards protecting their own privacy. Section 6 summarizes our findings from our first experiment, summarizing consumers' attitudes towards protecting their own privacy. Section 7 describes our experimental design for our second survey, which assessed consumers' attitudes towards protecting their children's privacy. Section 8 summarizes our findings from our first experiment, summarizing consumers' attitudes towards protecting their own privacy. Section 9 concludes the paper.

2. Defining Privacy

There are at least three different ways to characterize privacy and online invasions of privacy [2,8,24,38]:

(1) Perhaps the least threatening form of privacy violation is represented by an uninvited intrusion into a user's personal space. Online marketing, spam advertising, pop-ups, and sponsored sites around the edges of a web-page can all be seen as invasions of the user's personal space. Our focus groups indicated that this was the form of privacy violation most salient in users' minds, across age groups and across nations. When users thought of privacy violations from their email, phone, search, or social network providers, they thought almost exclusively about unwanted ads and unwanted interruptions. That is, they thought of privacy violations in terms of unwanted knocking on a hotel room door when a "Privacy Please" sign was visibly hanging from the doorknob.

(2) Surely the most extreme and most threatening form of privacy violation is represented by fraudulent ecommerce transactions, or even by identity theft. There is no indication that Google, Daum, Navor, Yahoo, or Bing have been associated with such threats to privacy, and there is no indication that consumers were concerned about this.

(3) And surely the form of privacy violations most important to Google, Daum, Navor, and Yahoo are based on personal profiling for some form of commercial advantage. This is definitely an intermediate form of privacy violation. It is far more than simply an invasion of personal space, and far less than identity theft. It involves uniquely identifying an individual and associating him with one or more characteristics of interest to an advertiser. The advertiser's intent may be benign; the firm simply wants to know everyone interested in visiting Osaka, or everyone interested in buying a food processor; this simply results in being sent ads for flights of interest, or products of interest. The advertiser's intent may also be less benign; the advertiser wants to know who engages in risky hobbies, so that he can avoid offering life insurance at rates that are too low, or who desperately needs to get to Chicago, so he can offer higher airfares.

Interestingly, consumers participating in our focus groups in Japan and Korea initially seemed to focus solely on the first form of privacy violation, the uninvited intrusion into personal space through various forms of spam and targeted marketing. A few hypothetical examples of integration and profiling were sufficient to arouse consumers' concerns. What if your phone tracked your position and it was clear that you now had a new geographic center of activity, your home, your office, your favorite restaurants, and the apartment of your new girlfriend? What if the phone read your text and knew whom you were seeing for lunch? What if it started suggesting her favorite restaurants to you? What if it accidentally started suggesting your new girlfriend's favorite restaurants to your friends? What if it accidentally started suggesting your new girlfriend's favorite restaurants to your other girlfriend?

A participant at one of the focus groups got truly

agitated when one of the other participants described targeted ads based on his search history. He suddenly realized that shortly after he was diagnosed with cancer he started to get email ads for cancer treatments, nursing care, and hospices. He realized that this was not simply coincidence and random spam; his search engine provider had made his most personal and sensitive medical history public, based on his searching for information on his specific form of cancer.

Participants at focus groups changed their attitudes towards privacy violations as they considered increasingly intrusive hypothetical examples of data mining and data integration. While these hypotheticals eventually became quite intrusive, none actually violated the published privacy policies of major search engine providers. Subjects came to realize during the course of the focus groups how much their search engine providers know. Moreover, they came to believe that if your search engine provider knows who you are and where you are and what you are going to do next and with whom you are going to do it, this is potentially compromising. This is potentially compromising, no matter who you are, and no matter how unexceptional your life might appear at the moment.

Since participants became increasingly aware of the full range of potential privacy violations only as the focus groups progressed, we believe that the level of awareness among survey subjects was actually lower than the level of awareness of focus group participants. We thus believe that the survey results are actually weaker than they appear and actually understate levels of concern among online users. Although the survey results appear quite strong, they understand concerns because they are based upon subjects' thinking principally about the weakest of the three levels of privacy violation.

3. Understanding the Regulatory Gap in Which Information Services Companies Operate

Information services companies operate in a regulatory vacuum, created by the difference between tight privacy regulations imposed upon traditional postal carriers and telecommunications carriers on one side, and the almost total absence of privacy regulation on many information systems companies like Google and Facebook.

(1) A traditional postal carrier cannot read your mail, and a rogue individual mail carrier cannot intercept and read your US mail, nor can a curious neighbor; doing so is a felony. Even government agencies cannot read your mail in the process of a criminal investigation without a court order. In contrast, American companies like Google and Facebook, and foreign companies like Daum and Navor, have built much of their publicly presented business models on their right to read and data mine any and all of your writings, whether public, like a Facebook post, or private, like a gmail or a text from an Android phone.

(2) A traditional telecommunications provider, like AT&T in the US, cannot eavesdrop on your communications or use your phone logs; indeed, it cannot allow a government agency to tap your phone or observe your phone call log without a court order. In contrast, Google can use your call logs from an Android phone to assess your social network, and Google has applied for patents that would allow it convert voice to text, and thus to text mine even your private voice communications.

(3) Likewise, until recently an ISP provider, like Yahoo in Japan, could not examine your communications, because an ISP provider was treated as a telecommunications company; in contrast, Google Japan could do whatever it wants with the packets that it sees a user generate. Interestingly, regulators responded to this anomaly not by protecting Google's users from invasion of privacy, but rather by subjecting Yahoo's users to the same risks as Google users. Since this so evidently is counter to the preferences expressed by participants in our focus groups and surveys, we believe that this provides further support for the importance of understanding consumers' preferences and of using consumers' preferences in the design of regulatory regimes to protect consumers' privacy.

We believe that "a packet is a packet is a packet" and similar packets should be treated similarly, whether they are carried by a traditional postal carrier, a package delivery service, an email vendor, the users' ISPs, or the backbone network that moves that packet between ISPs. Whether the packet is hand written on a single sheet of paper, typed onto a single sheet of paper, typed and loaded into a single SMS frame, or typed and loaded into numerous packets for delivery over the internet, users' communications should be protected.

4. The Myth of Anonymization

Privacy may not be dead, but anonymization by those who capture personal data surely is. Despite any and all claims that an information vendor protects consumers' privacy through anonymization, this simply is false. There are two factors, completely separate, that independently are sufficient to render anonymization impossible. Historically anonymization involved stripping off what are called personal identifiers, such as name and address and date of birth, or social security number, or other combinations that uniquely identify a single individual. This is how most privacy legislation is written today. Institutional Review Boards follow federal guidelines when they insist that research data be stripped of personal identifiers, and if it is

necessary to be able to identify individuals later the data and the identifying keys be encrypted and stored in separate locations. Both factors undo the effects of anonymization that is based on stripping off unique personal identifiers.

(1) The first factor is the ability to combine dozens, hundreds, even thousands of seemingly inconsequential factors, to identify an individual. Big data eliminates the relationship between identity and personal identifiers. An individual can now be uniquely identified in countless ways, without the use of personal identifying information.

(2) The second factor is the ability to induce the individual to identify himself, under conditions in which he would never willingly consent to do so.

Personal identifying information is no longer necessary for identifying an individual. Work by Acquisti and others shows that city of birth, year of birth, and the last for digits of a social security number is generally sufficient to identify an individual [1,16]. This was precisely the data that Google required to accompany every entry in their competition for children's art, Doodle 4 Google [4]. No one is suggesting that Google systematically engages in identity theft targeted at children. And yet they clearly have captured data that are sufficient to allow them to do so. Work by researchers at Microsoft has demonstrated that an individual's sequence of seemingly irrelevant online "likes" with very high probability can allow the researchers to identify an individual's sex, sexual orientation, marital status, and religion [23]. When this is combined with a modest amount of additional information it is once again possible to identify a unique individual.

More importantly, when an individual clicks on a targeted ad, he uniquely identifies himself and associates himself with the bucket of individuals with the characteristics the advertiser used to define the bucket. We call this the paradox of anonymous ads; these ads are indeed anonymous, until the user clicks on them; at this point the user has been fully identified, by himself. Airlines have long searched for the *"truth-o-matic"* pricing system, which would enable them to charge each individual his maximum willingness to pay for each trip. The best they can do now is to approximate by attributes that they have found are loosely suggestive of willingness to pay. If I buy my ticket in advance, I am willing to accept severe cancellation penalties, and am willing to stay for a weekend; this is probably a leisure trip, and I will be given a deep discount. If I buy later, I am unwilling to pay cancellation fees, and travel mid-week; this is probably a business trip, and I will be offered higher fares. But airlines would like to know far more, and data mining offers this. Imagine that someone texts his best friend in Chicago and says that he is terribly bored and would love to join him for dinner tomorrow night, if he can find a good airfare. Now imagine that instead the same individual texts his daughter in Chicago, says that his best friend is terribly ill, asks her to find a hotel room immediately, and tells her he needs to fly to Chicago as quickly as possible. An information services firm does not need to compromise the privacy of an individual by telling an airline that the *specific individual* is bored, or that the *specific individual* has a friend who is desperately ill. All the information services firm needs to do is send that individual an ad, appropriate to the individual's condition, or descriptive bucket. As soon as the individual clicks on the ad, he identifies himself as bored, or as having a friend who is desperately ill, depending on the ad he received. His airfare can be priced accordingly. The truth-o-matic has been delivered. No major university would be allowed to perform research that induced individual research subjects to identify themselves, especially if this could and would be used in ways that cause the subjects economic harm. And yet there is no regulation of this behavior by the targeted advertising industry. Indeed, many do not understand the problem, and argue that targeted ads are better for the consumer, because they waste less of the consumer's time, or are at worst harmless. They are not harmless, and do not need to be harmless.

We are not arguing that data mining is good, or that it is bad. We are not arguing that targeted advertising is good, or that it is bad. We are arguing that anonymization based on the process of stripping off what are normally called personal identifying information is no longer sufficient, in the presence of data mining. We are not arguing that targeted advertising is good, or that it is bad, or that individuals have or have not consented to receive targeted ads in exchange for free internet services. We are arguing that when an individual clicks on a targeted ad he now uniquely identifies himself, and that he may be doing so in a way that unambiguously links him to a set of actions and attributes with which he would never willingly link himself.

5. Experimental Design — Experiment One

Whether or not consumers actually believe that privacy is dead, it suits the companies that exploit private information to act as if this is so. For example, in 2010, Facebook changed its privacy policy so that its use of subscribers' private information is now the default, requiring users to opt out of publicly displaying information and granting Facebook the right to share information with third party sites [34]. Early last year, Google changed it privacy policy, allowing it to merge all the information it has collected from any and all of its products to develop an integrated profile on all of its users [27].

The study we conducted seeks to understand the following:

- The extent to which consumers do or do not know what information is collected, stored, and analyzed.
- The extent to which consumers do or do not know information is stored, combined, and analyzed.
- The degree to which consumers do or do not approve of practices, whether or not they were aware of them.
- The extent to which they do or do not believe that regulators are doing an adequate job of informing them about potential threats to their privacy.
- The extent to which they do or do not believe that regulators are doing an adequate job of protecting them from potential threat to their privacy.
- Consumers' preferences about whether privacy, do not track, or do not track and integrate, should be the default.

The study[2] can be viewed in some sense as an update of prior studies, including a study conducted by a colleague at the Annenberg School of Communication in 2005 [36]. In the intervening seven years since the Annenberg study was conducted, threats to privacy have become greater. Facebook has grown from millions of users to hundreds of millions. Google has acquired YouTube, and launched Chrome, Android, and Google +. The amount of information captured and the possibility of integration of information from multiple sources is enormously greater than it was seven years ago. Incidents like the Wi-Spy scandal have received press coverage around the world [19,15]. The abuse of privacy through Google analytics has resulted in litigation in the EU [37, 25]. Google recently accepted a 20-year audit because of privacy abuses during the launch of its social network, Google+ [12]. Privacy abuses are both more salient and potentially more damaging than when previous studies were conducted, and new studies are clearly warranted.

As with our prior studies on consumer trust in online shopping [5], we believe that national differences in behavior are interesting. We have therefore conducted this study in the US, Japan, Korea, and Germany.

6. Principal Findings on Consumers' Attitudes towards Protecting Their Own Privacy

Even working with survey subjects who we believe were principally focusing on the weakest of the three forms of privacy intrusion, we found strong disapproval of many common practices of online service providers. Users in general are willing to accept tracking of their searches. They do not always know which activities already occur, and they are less tolerant of tracking or monitoring many of their other online activities. They are less comfortable with the tracking and mining of their texts sent (see table 1), or of texts received (see table 2). Consumer resistance is comparably high for mining the content of emails that they send or receive (see tables 3 and 4). Comparing tables 3 and 4 with table 5, we see that surprisingly, gmail users are both more aware of privacy intrusions and more concerned about these intrusions. Table 6 shows that users' acceptance of having the content of their voice communications mined is even more limited. Finally, table 7 shows that subjects are only slightly more accepting of mining data from social networks. As can be seen from these tables, many subjects are unaware that various intrusions either are contemplated or have already been implemented, or have a misplaced confidence that they have not been. The lack of public objection therefore may be indicative of ignorance rather than indicative of approval.

Tables 8 and 9 show the extent to which consumers believe that their regulators are doing an adequate job ensuring that they are aware of the online threats to their privacy and the extent to which consumers believe that regulators are doing an adequate job of protecting them from online threats to their privacy. Two points can be observed immediately:

(1) In all four populations studied, by a large margin, consumers believe that their regulators are not doing an adequate job of informing them of online threats to their privacy.

(2) In all four populations studied, by a large margin, consumers believe that their regulators are not doing an adequate job of protecting them from online threats to their privacy.

	Strongly Disapprove	Weakly Disapprove	Accept	Sum
Yes	17% | 19% | 27% | 22% 23%	0% | 3% | 2% | 1% 1%	0% | 0% | 1% | 2% 1%	17% | 23% | 29% | 25% 26%
No	21% | 23% | 17% | 28% 22%	1% | 10% | 5% | 1% 4%	1% | 2% | 2% | 2% 2%	24% | 35% | 24% | 31% 28%
IDK	51% | 37% | 40% | 38% 40%	5% | 3% | 5% | 4% 4%	3% | 2% | 1% | 3% 2%	59% | 43% | 46% | 34% 46%
Sum	89% | 79% | 84% | 88% 85%	6% | 17% | 12% | 6% 10%	4% | 4% | 4% | 6% 5%	

Table 1. Awareness of and acceptance of mining texts sent

	Strongly Disapprove	Weakly Disapprove	Accept	Sum
Yes	16% | 23% | 23% | 18% 21%	1% | 2% | 2% | 1% 2%	0% | 0% | 0% | 2% 1%	17% | 24% | 25% | 21% 23%
No	20% | 19% | 17% | 29% 22%	3% | 13% | 6% | 4% 6%	1% | 3% | 3% | 2% 2%	24% | 35% | 27% | 35% 30%
IDK	51% | 35% | 42% | 37% 40%	5% | 3% | 4% | 4% 4%	3% | 2% | 2% | 3% 2%	59% | 40% | 48% | 43% 46%
Sum	87% | 77% | 83% | 84% 83%	9% | 18% | 12% | 9% 12%	4% | 4% | 4% | 7% 5%	

Table 2.Awareness of and acceptance of mining texts received

[2] U.S. (n=310), Japan (n=442), Korea (n=442), and Germany (n=505)

	Strongly Disapprove	Weakly Disapprove	Accept	Sum
Yes	19% | 30% | 26% | 33% 28%	2% | 3% | 3% | 4% 3%	1% | 1% | 1% | 1% 1%	22% | 34% | 30% | 38% 31%
No	15% | 16% | 13% | 15% 14%	2% | 5% | 3% | 1% 3%	2% | 2% | 2% | 0% 2%	19% | 23% | 18% | 16% 19%
IDK	52% | 37% | 49% | 40% 44%	5% | 4% | 3% | 3% 4%	2% | 2% | 1% | 3% 2%	59% | 43% | 53% | 46% 50%
Sum	86% | 83% | 87% | 88% 86%	9% | 12% | 9% | 8% 10%	5% | 5% | 4% | 4% 4%	

Table 3. Awareness of and acceptance of mining emails sent

	Strongly Disapprove	Weakly Disapprove	Accept	Sum
Yes	21% | 24% | 31% | 25% 26%	3% | 3% | 5% | 2% 3%	1% | 1% | 0% | 1% 1%	25% | 27% | 36% | 29% 29%
No	13% | 15% | 18% | 17% 16%	1% | 8% | 3% | 2% 4%	1% | 2% | 2% | 4% 2%	15% | 26% | 24% | 22% 22%
IDK	49% | 38% | 39% | 39% 41%	6% | 6% | 2% | 4% 4%	5% | 2% | 0% | 6% 3%	60% | 47% | 40% | 49% 48%
Sum	82% | 78% | 87% | 81% 82%	10% | 17% | 10% | 8% 11%	7% | 5% | 2% | 11% 7%	

Table 4. Awareness of and acceptance of mining emails received.

	Strongly Disapprove	Weakly Disapprove	Accept	Sum
Yes	15% | 30% | 13% | 25% 24%	1% | 3% | 3% | 3% 2%	1% | 0% | 1% | 1% 1%	16% | 33% | 27% | 29% 27%
No	14% | 14% | 49% | 14% 13%	0% | 6% | 4% | 3% 3%	1% | 2% | 2% | 1% 2%	14% | 21% | 19% | 18% 19%
IDK	59% | 38% | 49% | 46% 47%	9% | 5% | 3% | 4% 5%	2% | 3% | 2% | 3% 2%	69% | 46% | 53% | 53% 54%
Sum	88% | 82% | 85% | 85% 85%	10% | 13% | 10% | 10% 11%	3% | 5% | 5% | 5% 5%	

Table 5. Email users' awareness of and acceptance of mining gmail.

	Strongly Disapprove	Weakly Disapprove	Accept	Sum
Yes	7% | 6% | 10% | 10% 9%	0% | 0% | 0% | 0% 0%	0% | 0% | 0% | 1% 0%	7% | 6% | 11% | 11% 10%
No	32% | 38% | 31% | 40% 35%	3% | 14% | 5% | 6% 7%	0% | 7% | 6% | 3% 5%	35% | 59% | 42% | 50% 46%
IDK	57% | 31% | 41% | 32% 38%	4% | 2% | 3% | 4% 10%	0% | 1% | 3% | 3% 2%	59% | 35% | 47% | 39% 44%
Sum	96% | 76% | 82% | 82% 82%	4% | 16% | 9% | 11% 10%	0% | 8% | 9% | 7% 7%	

Table 6.Awareness of and acceptance of mining voice.

	Strongly Disapprove	Weakly Disapprove	Accept	Sum
Yes	43% | 31% | 41% | 60% 45%	8% | 4% | 4% | 7% 5%	3% | 0% | 1% | 3% 2%	54% | 36% | 46% | 70% 52%
No	5% | 10% | 9% | 5% 7%	0% | 9% | 2% | 1% 3%	0% | 2% | 2% | 2% 1%	5% | 21% | 13% | 8% 12%
IDK	31% | 33% | 35% | 17% 28%	5% | 7% | 3% | 2% 4%	5% | 3% | 3% | 2% 3%	41% | 43% | 41% | 22% 36%
Sum	79% | 74% | 85% | 83% 80%	13% | 21% | 9% | 10% 13%	8% | 5% | 5% | 7% 6%	

Table 7.Awareness of and acceptance of and mining data from social networks.

	Yes	No	IDK
U.S.	11%	68%	21%
Japan	4%	77%	19%
Korea	18%	63%	19%
Germany	10%	82%	8%

Table 8.Consumers' confidence that regulators are informing them adequately about threats to privacy online.

	Yes	No	IDK
U.S.	8%	67%	25%
Japan	3%	76%	21%
Korea	5%	80%	15%
Germany	9%	83%	8%

Table 9. Consumers' confidence that regulators are protecting them adequately from online threats to privacy.

Our findings regarding privacy settings are summarized in table 10. Our survey defined *tracking* as recording and analyzing a user's behavior over time at a single website and *integration* as recording, integrating, and analyzing all of a user's online behavior, including search, texting, email, phone, and other activities. Comparing the data in tables 1-7 to that of table 10, consumers appear to be much less tolerant of online tracking and integration than their behaviors and their answers to individual questions would suggest. Most significantly, the data appears to be no support for Google's position, adopted by the W3C, that a browser with DNT set as the default would not be in compliance with the standard, and thus that the privacy settings of their users could be ignored [35]. We find ourselves echoing Microsoft's position at the W3C, *"To say that a standard cannot not support a privacy by default choice for consumers is odd to say the least."* [35]

	U.S	Japan	Korea	Germany
Do users object to tracking by a single website	89%	98%	96%	89%
Do users object to full integration	97%	99%	98%	92%
Acceptable to ignore users' default settings	3%	2%	2%	0%

Table 10. Consumers' attitudes towards online tracking and integration

7. Experimental Design — Experiment Two

The second experiment was similar to the first, except that it was aimed at assessing parents' attitudes towards data mining the email accounts of their public school students[3]. While data mining of students' accounts is prohibited by federal law [10] and by law in many states [7], the practice appears to be widespread, and may indeed be legal when approval is granted by the school district's representatives.

Our second experiment had three parts:

1. A first survey was conducted, assessing parents' attitudes towards data mining of their children's email accounts, including a range of forms of detailed tracking and integration with search, texting, and other online activities.
2. A second survey was conducted, asking public school students to describe activities that they performed online, or were aware of other students

[3] U.S. Parents (n=246), U.S. Teenagers (n=469), Japan Parents (n=300), Japan Teenagers (n=241)

performing online.

3. A third survey was conducted, replicating the questions asked of parents in the first survey. In this survey, parents were first informed about what students actually did online.

The survey was performed in the United States and Japan. Results are summarized in the next section.

8. Principal Findings on Consumers' Attitudes towards Protecting Their Children's Privacy

Table 11 shows that Japanese parents are significantly less aware of the activities that are done by online information service providers than their US counterparts are. Approximately one third are aware that searches are tracked, and approximately one fourth are aware that email can be linked to other online activities. More than 75% of US parents are aware that searches are tracked, and 40% or more are aware that some form of integration is performed across different online services.

*Preferred Not to Answer

	Yes	No	PNTA*
Are your children provided with an email account by their school?	50% 13% **30%**	50% 87% **70%**	**N/A**
Are you aware that search engines track your child's entire search history, including searches for inappropriate content?	77% 35% **54%**	22% 63% **45%**	1% 3% **2%**
Are you aware that some email providers read the contents of all your children's email and can use the content to sell ads?	46% 26% **35%**	53% 71% **63%**	1% 3% **2%**
Are you aware that some email providers can link your children's email to their text and search histories?	40% 23% **31%**	60% 74% **68%**	0% 3% **2%**

Table 11. US and Japanese parents' awareness of data mining their children's online activities. Data values are stacked vertically, US / Japan / Average

Table 12 shows that both US and Japanese parents would strongly prefer that their children's online activities be free from data mining. However, Japanese parents are significantly more tolerant of the possibility that students' email may be data mined, even if the email is provided by the students' school district. This may in part be explained by their belief that their students are less likely to engage in inappropriate behavior online.

Our data (not reported here due to length limitations) show that US parents have reason to be concerned about data mining their students' online behavior. A significant fraction of students have engaged in inappropriate behavior online, or are aware of students who have. As we have discussed in section 4 above, nothing learned about online behavior is truly anonymous. Japanese parents do appear to have less reason for concern, given the apparently better behavior of Japanese students online. However, self-reported data on sensitive subject areas are always imperfect, and differences in cultural norms may partly explain differences in the data on self-reported student activities.

	Strongly Disapprove	Weakly Disapprove	Approve
Would you object if your child's school email provider engaged in data mining, including linking your child's emails to text, phone calls, GPS data, and search history?	92% 82% **86%**	4% 10% **6%**	4% 8% **4%**
Would you still object if your child's school email provider engaged in data mining, including linking emails to text and search history but provided email without charge to the school district?	91% 84% **87%**	7% 12% **10%**	2% 4% **3%**
If you believed that the company that sold information on your children's online behavior did so without protecting your children's privacy, would you care?	96% 89% **92%**	2% 8% **5%**	2% 2% **2%**

Table 12. US and Japanese parents' attitudes towards data mining their children's online activities. Data values are stacked vertically, US / Japan / Average

Additionally, students' have their own attitudes towards being profiled and to having their online activities tracked and integrated, and may consider this an invasion of privacy. Students' strongly disapprove of tracking and integrating their online behavior, as summarized in tables 13 and 14.

	Strongly Disapprove	Weakly Disapprove	Approve
Tracking Searches	82% 81% **82%**	10% 15% **12%**	8% 4% **7%**
Tracking Email	76% 85% **80%**	12% 11% **12%**	12% 4% **9%**
Tracking Text	82% 84% **83%**	9% 12% **10%**	9% 4% **7%**

Table 13. US and Japanese students' attitudes towards tracking their individual online activities. Data values are stacked vertically, US / Japan / Average

	Strongly Disapprove	Weakly Disapprove	Approve
Email Account and Private Email Account	85% 87% **85%**	4% 10% **5%**	11% 3% **9%**
Linking of School Email Account and Texting	82% 90% **84%**	7% 8% **7%**	11% 2% **8%**
Linking of Private Email Account and Texting	71% 86% **76%**	14% 12% **13%**	15% 2% **11%**
Email Acount, Private Email Account, and	85% 87% **85%**	5% 10% **6%**	10% 3% **8%**

Table 14. US and Japanese students' attitudes towards linking their email accounts with their texting. Data values are stacked vertically, US / Japan / Average

9. Conclusions

Consumers' preferences would suggest that the default privacy settings should be no tracking without explicit permission and no integration without explicit permission. Governments should hold firms accountable for violations and should ensure that violations are visible when they occur. Governments should ensure that consumers can know exactly what companies have done, so that they can protect themselves, register displeasure, and change vendors if they deem it necessary to stop violations of their privacy. Our data also show that the public is not widely aware of criminal violations of privacy, such as the WiSpy scandal in the US or iPhone hacking in the US, or Google Analytics violation of EU privacy laws. Given that these violations are strongly counter to consumers' preferences, perhaps they should receive greater attention and be more severely punished when they occur.

Many have argued that consumers' use of systems that compromise their privacy do so fully informed, and accepting the reduction in personal privacy as a free trade and a fair exchange for services rendered. Clearly, we do not believe this to be the case, and clearly we do not believe that consumers are fully informed or have offered informed consent. As Jay Kesan and his colleagues ask [22], *"What data are consumers willing to trade in exchange for convenience and services online? Would they be as willing to engage in this trade if their privacy rights were more protected and if they had the ability to exercise meaningful control over their data?"*

We believe that any policy regulations enacted should have the following characteristics.

- Regulations should treat companies equally, regardless of their industry classification. Telecommunications companies, regulated common carriers, and information services providers should have comparable obligations.
- Regulations should be consistent. If letters are protected, then emails and texts should be protected as well. If information stored in remote electronic data storage facilities is protected, then information stored in remote electronic cloud data storage facilities should be protected as well.
- Companies' policies should be transparent. Users should know what they are giving up, in exchange for "free services." Users should understand the loss of anonymity associated with data mining and targeted ads. Users should understand how the information gathered can be used, both for them and against them.
- The fact that a service is free is not sufficient justification for obscuring the risks associated with that service. No tobacco company could successfully argue that it was not subject to regulation because it provided its tobacco products free of charge to public school students. Similarly, providing information services free of charge does not eliminate the need to provide products that are safe and that observe all relevant regulations.

We are still conducting focus groups and experiments in additional locations. Data analysis is also ongoing.

References:

[1] Acquisti, A., John, L. and Lowenstein, G. 2009. What is privacy worth? Workshop on Information Systems and Economics (WISE).

[2] Altman, I. The Environment and Social Behavior: Privacy, Personal Space, Territory and Crowding. Brooks/Cole Pub. Co., Inc., Monterey, CA, 1975.

[3] Angwin, Julia. (2010, July 10). "Google, FTC Near Settlement on Privacy." The Wall Street Journal. http://online.wsj.com/article/SB10001424052702303567704577517081178553046.html.

[4] Bowdon, Bob. (2011, Feb. 22). "Why has Google been Collecting Kids' Social Security Numbers under the Guise of an Art Contest?" Huffington Post. http://www.huffingtonpost.com/bob-bowdon/why-has-google-been-colle_b_825754.html.

[5] Clemons, E., JIN F., Wilson, J., REN, F., Matt, C., Hess, T., and KOH, N. "The Role of Trust in Successful Ecommerce Websites in China: Field Observations and Experimental Studies." Proceedings, 45th International Conference on System Sciences, Maui, Hawaii, January 2013.

[6] Department of Justice. (2011, Aug. 11). Google Forfeits $500 Million Generated by Online Ads & Prescription Drug Sales by Canadian Online Pharmacies. Press Release. http://www.justice.gov/opa/pr/2011/August/11-dag-1078.html.

[7] Electronic Privacy Information Center. "Student Privacy." http://epic.org/privacy/student/.

[8] Etzioni, A. The Limits of Privacy. Basic Books, New York, 1999.

[9] European Commission. (2013, Feb. 20). MEMO/13/124, "EU Data Protection: European Parliament's Industry committee backs uniform data protection rules." Press Release. http://europa.eu/rapid/press-release_MEMO-13-124_en.htm.

[10] Family Educational Rights and Privacy Act. 20 U.S.C. § 1232g; 34 CFR Part 99.

[11] Federal Trade Commission. 2009. Self-Regulatory Principles for Online Behavioral Advertising. http://www.ftc.gov/os/2009/02/P085400behavadreport.pdf.

[12] Federal Trading Commission. 2011. "FTC Charges Deceptive Privacy Practices in Google's Rollout of Its Buzz Social Network." Press Release. http://www.ftc.gov/opa/2011/03/google.shtm.

[13] Fleischer, P. (2013, Feb. 17). “Don Quixote.” http://peterfleischer.blogspot.com/2013/02/dox-quixote.html.

[14] Goldman, David. (2012, June 8). “Microsoft and Google Play Chicken Over Do Not Track.” CNN Money. http://money.cnn.com/2012/06/07/technology/do-not-track/index.htm.

[15] Google Street View Privacy Concerns. In Wikipedia. http://en.wikipedia.org/wiki/Psychology

[16] Gross, R., and Acquisti, A. 2005. Information Revelation and Privacy in Online Social Networks. Proceedings of WPES’05 (pp. 71–80). Alexandria, VA: ACM.

[17] Hearing on Consumer Online Privacy: Hearing Before the Committee on Commerce, Science, and Transportation, Senate, 111th Cong. (S. Hrg. 111-1038). (2010). http://www.gpo.gov/fdsys/pkg/CHRG-111shrg67686/pdf/CHRG-111shrg67686.pdf.

[18] Hill, Kashmir. (2013, April 12). “Should Cops have gotten a Warrant before Reading Customer’s Texts on a Drug Dealer’s iPhone?” Forbes. http://www.forbes.com/sites/kashmirhill/2013/04/12/should-cops-have-gotten-a-warrant-before-reading-customers-texts-on-a-drug-dealers-iphone/.

[19] Hill, Kashmir. (2012, April 30). “Wi-Spy Google Engineer Wanted Snapshots of Where People We and What They Were Doing.” Forbes. http://www.forbes.com/sites/kashmirhill/2012/04/30/wi-spy-google-engineer-wanted-snapshots-of-where-people-were-and-what-they-were-doing/

[20] Jaffer, Jameel. (2013, June 10). “Is the N.S.A. Surveillance Threat Real or Imagined?” Room for Debate. The New York Times. http://www.nytimes.com/roomfordebate/2013/06/09/is-the-nsa-surveillance-threat-real-or-imagined?hp.

[21] Jennings v. Jennings. (S.C. Ct. App. Jul. 14, 2010). http://www.sccourts.org/opinions/HTMLFiles/SC/27177.pdf.

[22] Kesan, J.P., Hayes, C.M., and Bashir, M.N. “Information Privacy and Data Control in Cloud Computing: Consumers, Privacy Preferences, and Market Efficiency,” Washington and Lee Law Review, (70:1), pp. 341-472.

[23] Kosinski, M., Stillwell, D., and Graepel, T. (2013, March 11). “Private traits and attributes are predictable from digital records of human behavior.” Proceedings of the National Academy of Sciences. http://www.pnas.org/content/early/2013/03/06/1218772110.full.pdf+html.

[24] Margulis, S. 1977. "Conceptions of Privacy: Current Status and Next Steps," Journal of Social Issues (33), pp.5–21.

[25] McGee, Matt. (2011, Jan. 12). “German Govt. Says Google Analytics Now Verboted.” Search Engine Land. http://searchengineland.com/german-govt-says-google-analytics-now-verboten-61109.

[26] Microsoft. Advertisement. Scroogled! http://www.scroogled.com/.

[27] Millian, Mark. (2012, Jan. 25). “Google to Merge User Data Across its Services.” CNN.com. http://www.cnn.com/2012/01/24/tech/web/google-privacy-policy.

[28] Nakashima, E., Wilson, S. (2013, June 11). “ACLU Sues over NSA Surveillance Program.” The Washington Post. http://www.washingtonpost.com/politics/aclu-sues-over-nsa-surveillance-program/2013/06/11/fef71e2e-d2ab-11e2-a73e-826d299ff459_story.html.

[29] Perez, Sarah. (2013, Feb. 7). “Seriously, This Again? New, Aggressive Marketing From Microsoft Warns Gmail Users That Google Reads Their Email." Tech Crunch. http://techcrunch.com/2013/02/07/seriously-this-again-new-aggressive-marketing-from-microsoft-warns-gmail-users-that-google-reads-their-email/.

[30] Risen, J., Lichtblau, E. (2009, Jan. 15). “Court Affirms Wiretapping Without Warrants.” The New York Times. http://www.nytimes.com/2009/01/16/washington/16fisa.html?ref=foreignintelligencesurveillanceactfisa.

[31] Roberts, D., Ackerman, S. (2013, June 6). “Anger Swells after NSA Phone Records Court Order Revelations.” The Guardian. http://www.guardian.co.uk/world/2013/jun/06/obama-administration-nsa-verizon-records.

[32] Schmitt, E., Sanger, D., & Savage, C. (2013, June 7). Administration Says Mining of Data is Crucial to Fight Terror. New York Times. http://www.nytimes.com/2013/06/08/us/mining-of-data-is-called-crucial-to-fight-terror.html?hpw.

[33] Sengupta, Somni. (2013, April 24). “Updating an E-Mail Law From Last Century.” The New York Times. http://www.nytimes.com/2013/04/25/technology/updating-an-e-mail-law-from-the-last-century.html?pagewanted=all.

[34] Seward, Zachary. (2013, Jan. 18). “Facebook Has Made It Impossible for You to Opt Out of Search Results.” Business Insider. http://www.businessinsider.com/facebook-has-made-it-impossible-for-you-to-opt-out-of-search-results-2013-1.

[35] Temple, James. (2012, June 6). "Internet Standards Group Rejects Default Do Not Track.” The Technology Chronicles. SFGate. http://blog.sfgate.com/techchron/2012/06/06/internet-standards-group-rejects-default-do-not-track/.

[36] Turow, J. (2005, June). “Open to Exploitation: American Shoppers Online and Offline.” Report of the Annenberg Public Policy Center.

[37] Wauters, Robin. (2009, Nov. 24). “Using Google Analytics is Illegal, German Officials Claim.” TechCrunch. http://techcrunch.com/2009/11/24/google-analytics-illegal-germany/.

[38] Westin, A. Privacy and Freedom. Atheneum Books, New York, 1967.