

**Footprints Near the Surf:
Individual Privacy Decisions in Online Contexts**

Submitted in partial fulfillment of the requirements for
the degree of
Doctor of Philosophy
in
Engineering & Public Policy

Aleecia M. McDonald

B.A., Professional Writing, Carnegie Mellon University
M.S., Public Policy and Management, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

December, 2010

Copyright 2010 Aleecia M. McDonald.

This work is licensed under the Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Dedication

In memory of Ferdinand David Schoeman.

Abstract

As more people seek the benefits of going online, more people are exposed to privacy risks from their time online. With a largely unregulated Internet, self-determination about privacy risks must be feasible for people from all walks of life. Yet in many cases decisions are either not obvious or not accessible. As one example, privacy policies are written beyond most adults reading comprehension level, and few people read policies let alone act based on the information policies contain. In my thesis I examine decisions made about threats from website data collection. In the course of multiple studies I use a variety of tools including lab-based studies, online studies, mental models interviews, economic analysis, and analysis of cookies used for tracking. Privacy literature is full of apparent conflicts between people saying they care very much about their privacy, yet not taking the steps required to protect their privacy. By using multiple approaches and crossing multiple disciplines I am able to contribute to a more coherent picture of whether people are able to make choices about protecting their online privacy.

Contents

Dedication	iii
Abstract	v
1 Introduction	1
1.1 Privacy Policies	2
1.2 Targeting	5
1.3 Overview	8
I Privacy Policies	9
2 Value of Time to Read Privacy Policies	11
2.1 Introduction	12
2.1.1 Economic Theories of Privacy Policies	13
2.2 Inputs to the Model	15
2.2.1 Time to Read or Skim Privacy Policies	16
2.2.2 Monthly Number of Unique Websites Visited	20
2.2.3 Annual Number of Unique Websites Visited	22
2.2.4 Opportunity Cost of Time	23
2.3 Time and Economic Value to Read Privacy Policies	24
2.3.1 Amount of Time to Read Privacy Policies	24
2.3.2 Value of Time to Read Privacy Policies	25
2.4 Discussion	25

3	Contrasting Formats & Lengths in One Privacy Policy	29
3.1	Introduction	30
3.2	Study Design	30
3.2.1	Formats	31
3.2.2	Length	33
3.2.3	Summary of Conditions	35
3.2.4	Methods	36
3.3	Results	38
3.3.1	Overall	38
3.3.2	Privacy Finder	41
3.3.3	The P3P Expandable Grid	42
3.3.4	Layered	43
3.3.5	Length	44
3.4	Study Limitations and Future Work	44
3.4.1	Representativeness of Policies	45
3.4.2	Length	45
3.5	Observations	46
3.6	Conclusions	47
3.6.1	Accuracy	48
3.6.2	Psychological Acceptability	48
4	Comparative Study of Six Privacy Policies	51
4.1	Introduction	52
4.2	Related Work	52
4.2.1	Privacy Finder	53
4.2.2	Layered Notices	53
4.2.3	Natural language	54
4.3	Methods	54
4.3.1	Study Conditions	54
4.3.2	Study Questions	57
4.3.3	Research Questions	58

4.3.4	Analysis	59
4.4	Accuracy and Speed Results	59
4.4.1	Cookies	60
4.4.2	Opt Out Link	61
4.4.3	Share Email	63
4.4.4	Telemarketing	64
4.5	Psychological Acceptability Results	66
4.5.1	Ease of Finding Information	66
4.5.2	Trust	67
4.5.3	Enjoyment	67
4.6	Demographics	67
4.7	Subsequent Work	68
4.8	Discussion	68

II Targeting and Behavioral Advertising 71

5	Online Behavioral Advertising	73
5.1	Introduction	74
5.2	Background and Related Work	76
5.3	Research Methods	78
5.3.1	Demographics	79
5.3.2	Transferability	80
5.4	Perceptions About Cookies	81
5.4.1	Misperceptions of First Party Cookies	83
5.4.2	Knowledge of Cookies	84
5.4.3	Managing Cookies	86
5.4.4	Unclear on Clearing Cookies	87
5.4.5	Cookies and Browser History	89
5.4.6	Lack of Understanding of Cookies and Data Flows	91
5.4.7	Consumers Do Not Understand NAI Opt-Out Cookies	93
5.5	Tailored Content and Privacy Concerns	95

5.5.1	Mixed Identification of Internet Advertising	95
5.5.2	Mixed Understanding of Current Practices	98
5.5.3	Reasons to Accept or Reject Tailored Advertising	100
5.5.4	Privacy and Security Among Top Priorities for Buying Online	102
5.6	Payment for Privacy	104
5.6.1	Gap Between Willingness to Pay and Willingness to Accept	104
5.6.2	Reasons to Pay or Refuse to Pay for Privacy	105
5.7	Conclusions and Discussion	106
6	Beyond Behavioral	109
6.1	Introduction	110
6.2	Related Work	112
6.3	Research Questions	113
6.4	Methods	114
6.4.1	Study Questions	114
6.4.2	Analysis	115
6.5	Results	115
6.5.1	Definitions	115
6.5.2	Advertising Sections	117
6.5.3	Preferences for Random Advertisements	118
6.5.4	Tradeoffs with Targeted Advertisement	120
6.5.5	Rejection of Payment for Privacy	121
6.5.6	Demographics	127
6.6	Discussion	128
7	Flash Cookies	131
7.1	Introduction	132
7.2	Background and Related Work	132
7.3	Research Methods	135
7.4	Results	138
7.4.1	Use of HTTP Cookies	138

7.4.2	Use of LSOs	140
7.4.3	Mismatched Sites	141
7.4.4	Prevalence of Unique Identifiers and Respawning in LSOs	141
7.5	Discussion	144
7.6	Recommendations	146
8	Policy Options	149
8.1	Status Quo	150
8.1.1	Proponents	150
8.1.2	Opponents	151
8.1.3	Research	151
8.2	Modify Self-Regulation	156
8.2.1	Proponents	156
8.2.2	Opponents	157
8.2.3	Research	157
8.3	Replace Self-Regulation	158
8.3.1	Proponents	158
8.3.2	Opponents	159
8.3.3	Research	159
8.4	The Road Forward	160
A	Statistics for Privacy Policies	161
A.1	Statistical Significance	162
A.1.1	Accuracy	162
A.1.2	Time	162
A.1.3	Psychological Acceptability	163
B	Behavioral Advertising Study	165
C	Targeted Advertising Comparison	191
D	Flash Cookie Sites	201

List of Figures

2.1	Probability Density Function (“PDF”) and Cumulative Distribution Function (“CDF”) of Word Counts in Popular Website Privacy Policies	17
2.2	Median times and inter-quartile ranges to skim one privacy policy	19
2.3	Median time to answer a basic question in one of six policies of different lengths, bracketed by interquartile range	20
2.4	Locations where people read websites	21
3.1	Privacy finder policy	32
3.2	The P3P Expandable Grid policy	33
3.3	Layered policy	34
3.4	Natural language policy	35
3.5	Overall mean Likert scores for psychological acceptability across all conditions. Vertical bars denote 95% confidence intervals around the point estimates.	39
3.6	Overall mean percentage of participants who answered the question correctly, across all conditions. Vertical bars denote 95% confidence intervals around the point estimates.	40
3.7	Overall mean time to answer questions in seconds. Vertical bars denote 95% confidence intervals around the point estimates.	41
5.1	Percentage of respondents interested in targeted ads, discounts, and news by age groups in the Turow study, as contrasted with our results	82
5.2	Four browsers’ interfaces for deleting cookies: Firefox, Internet Explorer, Safari, and Opera.	90

5.3	Four possible mental models of how advertising cookies work.	92
5.4	Screenshot of the NAI Opt Out page	93
6.1	Level of agreement with positive and negative adjectives describing advertising practices, ranging from strongly disagree (1) to strongly agree (7.)	119
6.2	Percentage of respondents who prefer random ads to different types of targeted ads. Scuba and cancer conditions had no significant differences so their average is shown. Affiliate, cloud, and DPI are not statistically different from each other, but do significantly differ from contextual and behavioral.	120
6.3	Percent who self-report they would pay \$1 per month to avoid ads	121
6.4	Percentage of respondents who volunteered a given category of answer when asked why they would, or would not, spend \$1 per month to avoid a type of advertising.	127
7.1	Flow chart of website classification based on SharedObjects	137
7.2	Analysis of the 100 most popular websites in 2010. Purple circles contain the number of sites that fall into a given category.	142
7.3	Analysis of the 100 most popular websites in 2009. Purple circles contain the number of sites that fall into a given category.	143
7.4	Analysis of the 500 randomly selected websites. Purple circles contain the number of sites that fall into a given category.	144

List of Tables

2.1	Times to read entire privacy policies for average readers	17
2.2	Time estimates to skim one policy and answer a basic question	20
2.3	Estimates of the monthly number of unique websites visited by U.S. Internet users	22
2.4	Unique monthly and weekly websites visited by U.S. Internet users show repeat visits to many sites week after week	22
2.5	Estimates of the annual number of unique websites visited by U.S. Internet users .	23
2.6	Estimates for the value of time to read online privacy policies	24
2.7	Annual time estimates for reading and skimming online privacy policies	24
3.1	Number of participants per condition, $\sum n = 865$	36
3.2	Statistically significant differences between mean Likert scores on psychological acceptability, by policy format. Higher Likert scores are better.	42
3.3	Statistically significant differences between mean completion times by policy format.	43
3.4	Statistically significant differences between mean completion times by policy length.	44
4.1	Participants per Condition	55
4.2	Attributes of six companies' privacy policies	57
4.3	Percentage correct and minutes to answer, cookies question.	60
4.4	Percentage correct and minutes to answer for the opt out question.	61
4.5	Percentage correct and minutes to answer for the email sharing question.	63
4.6	Percentage correct and minutes to answer for the telemarketing question.	64
4.7	Examples of skew for demographics	68
5.1	Demographics for online study	80

5.2	Percentage of respondents who want tailored content	81
5.3	Responses to factual questions about cookies — correct answers in bold	85
5.4	Perceived likelihood of practices occurring	98
5.5	Attitudes toward current practices	99
5.6	Mean Likert scores to accept or reject behavioral advertising (Strongly Agree = 7, Strongly Disagree = 1.)	100
5.7	Respondents who buy online	102
5.8	How sellers can entice more online purchases (Matters a lot = 4, Does not matter = 1)	103
5.9	Reasons to pay for privacy or accept a discount	105
7.1	Technical differences between HTTP cookies and Flash cookies	132
7.2	HTTP Cookies	140
A.1	Statistical Significance Tests for Accuracy Questions by Company	162
A.2	Statistical Significance Tests for Accuracy Questions by Format	163
A.3	Statistical Significance Tests for Time to Answer by Company	163
A.4	Statistical Significance Tests for Time to Answer by Format	163
A.5	Statistical Significance Tests for Psychological Acceptability by Company	164
A.6	Statistical Significance Tests for Psychological Acceptability by Format	164
D.1	Quantcast's Top 100 most visited websites as of July 8, 2010	202
D.2	Quantcast's Top 100 most visited websites as of July, 2009	202
D.3	Random selection of 500 sites	203

Chapter 1

Introduction

Internet privacy is a multi-faceted domain drawing upon law, policy, ethics, social norms, economics, and technology. Many practices have evolved with little thought for privacy. Internet users' understanding of online privacy threats is usually incomplete at best. Yet every day, Internet users make decisions with privacy implications: which websites to visit, which technologies to use to protect their privacy, what information to release about themselves, when to lie in response to questions that seem overly invasive. This thesis focuses on privacy in practice and the decisions people make about online privacy. We focus on two areas. First, privacy policies are the way in which users are expected to educate themselves about websites' data practices. Second, targeted advertising is the realm in which most commercial data is collected, where users are called upon to make decisions to protect or forgo their online privacy.

1.1 Privacy Policies

In the late 1990s, the Federal Trade Commission (FTC) decided that the Internet was evolving very quickly and new legislation could stifle growth. In particular, there were concerns that it was premature to legislate to protect privacy before other mechanisms evolved, especially when business was expected to offer more effective and efficient responses than FTC staff could enforce. The Internet was still young, commerce on the Internet was very new, and legislators and regulators adopted a hands-off approach rather than risk stifling innovation. However, concerns remained about data privacy in general and on the Internet in particular. For example, the FTC recommended legislation to protect childrens' privacy, which led to the Childrens Online Privacy Protection Act (COPPA) in 1998 [113].

Prior to COPPA, the FTC adopted Fair Information Principles (FIPs), a set of ideals around data use. The notion of FIPs predates the Internet; several nations adopted differing FIPs in response to concerns about credit databases on mainframes in the 1970s [85]. While FIPs do not themselves carry the force of law, they provide a set of principles for legislation and government oversight. In this way they are similar to Article 12 of the Universal Declaration of Human Rights, which states the principle that "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks," yet leaves the specific legal implementations of those ideals in the hands of individual signatories, including the

United States [149].

The five FIPs the FTC adopted in 1973 — notice/awareness, choice/consent, access/ participation, integrity/security, and enforcement/redress — are a subset of the eight protections enshrined in the Organization for Economic Co-operation and Development (OECD) Guidelines on the Protection of Privacy and Transborder Data Flows of Personal Data [108]. The FIP of notice underlies the notion of privacy policies, which are mechanisms for companies to disclose their practices. In 1998, the FTC commissioned a report that found while 92% of U.S. commercial websites collected some type of data, only 14% provided comprehensive notice of their practices [114]. The FTC was concerned that the FIP of notice/awareness was not faring well on the new Internet: consumers did not know where their data went or what it might be used for [114].

The FTC initiated a series of studies of hundreds of commercial websites to determine how well industry self-regulation worked, in what became known as “Internet sweeps”. Year after year, the number of companies offering privacy policies increased. By that metric it appeared the FTC was successful. However, studies also showed people were reluctant to shop online because they had privacy concerns [114]. The FTC turned to two different innovative approaches, rather than legislation or regulatory action. First, they expressed great hope for online privacy seals [113]. Two seal providers, TRUSTe and the Better Business Bureau (through BBBOnline), began certifying website privacy policies. The BBBOnline seal program is no longer offered. TRUSTe requires companies to follow some basic privacy standards and document their own practices. TRUSTe also advertises that they investigate consumer allegations that licensees are not abiding by their policies [142]. However, TRUSTe faced criticism for not requiring more rigorous privacy standards [93]. One study showed that companies with TRUSTe seals typically offer less privacy-protective policies than those without TRUSTe seals [72].

In addition to the threat of new regulatory action to spur voluntary disclosure, the FTC used fraud and deceptive practices actions to hold companies to whatever content they did publish. In essence, while a company was not strictly required to post a policy, once published, the policy became enforceable. In one case the FTC brought action even without a privacy policy. When Cartmanager surreptitiously rented their customer lists the FTC advanced a legal theory of unfairness rather than fraud [130]. Cartmanager provided online shopping cart software and worked with clients who promised not to sell customer data. The FTC argued that even though Cartmanager

did not have a privacy policy of their own to violate, they still violated the policies of their clients [51]. The FTC also brought action against Sears. Although Sears disclosed their data collection practices, the FTC decided the disclosure was inadequate [54]. This establishes disclosure is necessary but not sufficient, and suggests a reason beyond usability as to why businesses would do well to pay attention to the format in which they present privacy information.

Voluntary disclosure through privacy policies became the primary component of the notice and choice paradigm that underlies the theory of industry self-regulation. How would long would it take for users to read privacy policy, and what is the cost of that time? We address this question in Chapter 2. Because privacy policies were voluntary,¹ there were no requirements for the existence of a policy let alone any restrictions as to the format, length, readability, or content of a given privacy policy. Simultaneously, the FTC encouraged privacy enhancing technologies (PETs) with the hope that PETs would put greater control directly into the hands of consumers [113]. PETs include encryption, anonymity tools, and other software-based approaches. One particularly intriguing approach came from the Platform for Privacy Preferences (P3P) standard, which used privacy policies coded in standardized machine-readable formats. P3P user agents can determine for users if a given website has an acceptable privacy policy [41]. One possible way to improve privacy policies' readability is to create standardized formats. One approach uses P3P to generate a standardized policy. Another effort, called layered or short policies, presents highlights in standardized boxes. Do these presentations actually help users understand privacy policies in order to make decisions based upon them? We analyzed three formats for a single company's privacy policy in Chapter 3, and extend that work in Chapter 4, where we analyze six companies' privacy policies.

Privacy policies remain the most prevalent way for companies to communicate their practices, and for users to gain information. Users visit first-party sites with links to privacy policies, and may choose to click those links, and then may or may not understand the content of the privacy policy. However, as the Internet has evolved, first-party data collection is usually less concerning than the digital dossiers created by third-party advertisers.

¹There are some legal requirements for privacy policies now, including state law in California.

1.2 Targeting

As mass media gave rise to mass advertising, advertisers' reach became national. However, typically only a subset of citizens are interested in any given products or services advertised. As the old advertisers' lament has it, "We know we're wasting half our ad dollars, we just don't know which half" [44]. One approach to better match ads to people who might act upon them is *contextual advertising* where ads are related to the context in which they are shown. For example, ads for golf tees are more successful in golf magazines than quilting magazines. On the Internet, advertisers can buy ad space on websites related to their products. Advertisers can also choose search keywords and display ads contemporaneously with search results on sites like Google. For contextual advertising to work, advertisers need to know about the site they are advertising on, or to match to ephemeral keywords that do not need to be stored.

In addition to contextual advertising, the Internet also enables several novel forms of advertising designed to target likely customers in ways that are not available offline. Ideally, targeted ads help advertisers eliminate the "wasted half" of their advertising budget by showing ads only to people who are more likely to be customers. By targeting individual customers advertisers do not need to rely on high-priced premium sites to reach their audience. It is no longer the type of site that determines ad placement but rather the type of customer, regardless of which site they are visiting. Customers may benefit from ads targeted to their personal interests, reducing irrelevant ads and the time it takes to find products.

However, the data needed to drive targeted advertising is data about potential customers, rather than data about websites hosting ads. These data raise privacy concerns. Advertising networks collect data including search terms, websites visited, articles read online, IP addresses, web browser user agent strings, GPS and other location data. They use that information to infer demographics, interests, relationships, and physical movement. Based on inferences and common characteristics, they classify people into profiles. They then sell access to people deemed more likely to buy a given product or service, for example, perhaps advertising for an assisted living facility shown to affluent senior citizens in a particular state. The advertising networks sit in the middle between the users who see the ads in their web browsers, and the advertisers buying the ad embedded on a site the user visits. In this model, advertising networks do not provide information about the users (and have some economic incentives not to) but rather just display ads

to targeted un-named users. Some of the data that drives targeted advertising is inherently non-anonymized in raw form. As a few examples, people “ego surf” and search for their own names, leading to one path of identification of AOL’s users when AOL released what they believed to be anonymous search logs [17]. IP addresses are legally classified as Personally Identifiable Information in some countries, and laws require ISPs to retain logs of IP address use to assist law enforcement. Approximately 80% of web browsers have a unique “fingerprint” based on their configuration data [45]. One clever advertiser analyzed typing patterns to uniquely identify individuals sharing the same computer, based on the old telegraph operators’ observation that they could tell who they were talking to by the other operator’s “fist,” or the particular pauses and patterns present even in Morse code [26]. In addition to these examples of concerns around raw data being identifiable, privacy concerns also include issues with inferences and profiles. Finally, startling work on re-identification in other areas suggests that even theoretically anonymous profiles may contain a combination of enough data to be uniquely identifiable: raw data may not be the only identifiable data [107, 5, 135].

The FTC has devoted significant resources to understanding the contours of behavioral advertising, including multiple hearings, workshops, and roundtables culminating in guidelines for behavioral advertising in 2009 [53]. We study perceptions and knowledge of behavioral advertising in depth in Chapter 5. And yet, for all of the FTC’s time on behavioral advertising, they pay comparatively little attention to other forms of functionally similar targeted advertising. The FTC’s guidelines for behavioral advertising get into the technology of today’s behavioral advertising, contrasting first party cookies and third party cookies, rather than looking at data flows and outcomes in a generalizable way [53]. Since the FTC’s reasoning for distinguishing first party cookies is user expectations, it is reasonable to ask if users see behavioral advertising differently from other types of advertising, as we do in Chapter 6.

Most targeted advertising relies upon HTTP cookies. Advertisers use persistent identifiers in cookies to help them understand a given customer’s browsing history. This data is used to build interest profiles to command premiums for ads matched to given interests or demographics. Advertisers also use cookies to contribute to analytics data about which customers have viewed ads, clicked on ads, and purchased from ads. Analytics data helps advertisers determine if a given ad is effective with a particular audience. More importantly, without at least basic analytics, ad-

vertising networks would not know how much to charge. Meanwhile, many users prefer not to be tracked and express that preference by deleting their HTTP cookies. This causes tremendous problems for analytics data, where even a small error rate can result in incorrectly billing thousands of dollars in a single advertising campaign.

In response to users deleting HTTP cookies, advertisers have turned to more creative approaches to gather reliable data. One of several is to simply use Flash cookies as a direct replacement for HTTP cookies, or to re-create HTTP cookies users deleted. Flash cookies come from the Flash browser plugin. Flash is used to create multimedia applications including interactive content and animations embedded into web pages. An estimated 99% of desktop web browsers have the free Flash Player plugin enabled [8]. Flash programs cannot read and write HTTP cookies directly, but can use JavaScript to do so [30]. However, while JavaScript is built into all major browsers, users can choose to disable JavaScript and thereby disable Flash's ability to interact with HTTP cookies. Flash writes its own version of cookies.

Because most users have not heard of Flash cookies, and tools did not delete Flash cookies until recently, advertisers discovered Flash cookies solved their data quality problems. Even better for advertisers, Flash cookies do not expire. Under Windows, Flash cookies write to hidden system folders, away from most users' notice or technical ability to delete. Flash cookies are cross-browser, eliminating advertisers' problem with HTTP cookies that a user using Internet Explorer and Firefox is miscounted as two different users. Rather than write a lot of new code to work with Flash cookies, in some cases advertisers simply used Flash cookies to identify a user and then re-create ("respawn") that user's previously deleted cookies, enabling advertisers to continue to use their existing code base. Users did not have to bother re-entering data or to be pestered with information about behind-the-scenes internal processes.

As a technical response to the technical problem of poor quality analytics data, Flash cookies are a good engineering solution. However, problems collecting analytics data are not just a technical glitch: users *intentionally* delete HTTP cookies as an expression of their desire for privacy. We rely upon an industry self-regulation approach to privacy, built on a notice-and-choice theory. Using Flash cookies to respawn or to track users who have deleted HTTP cookies completely undermines user choice, and violates the underlying principles of self-regulation. Users had no visible indication that Flash cookies existed or that HTTP cookies respawned. In order to

understand how much of a problem this poses, we analyze Flash cookie prevalence in Chapter 7.

1.3 Overview

The rest of this dissertation is organized as follows: in Part I, we study using privacy policies to make decisions. We estimate the time to read privacy policies and the cost to do so in Chapter 2. We contrast three different formats for a single privacy policy in Chapter 3, and contrast six companies' policies in Chapter 4. In Part II, we study targeting techniques and behavioral advertising. We quantify Flash cookies use in Chapter 7. We performed user studies to understand views of behavioral advertising, described in Chapter 5. We contrast user perceptions of several types of targeted advertising in Chapter 6. Finally, we conclude with a discussion of policy options in 8.

Part I

Privacy Policies

Chapter 2

Value of Time to Read Privacy Policies

This chapter is largely a reproduction of a paper co-authored with Lorrie Faith Cranor and published in *I/S: A Journal of Law and Policy for the Information Society*, 2008 [96].

2.1 Introduction

Companies collect personally identifiable information that website visitors are not always comfortable sharing. One proposed remedy is to use economics rather than legislation to address privacy risks by creating a market place for privacy where website visitors would choose to accept or reject offers for small payments in exchange for loss of privacy. The notion of micropayments for privacy has not been realized in practice, perhaps because advertisers might be willing to pay a penny per name and IP address, yet few people would sell their contact information for only a penny [56]. In this chapter we contend that the time to read privacy policies is, in and of itself, a form of payment. Instead of receiving payments to reveal information, website visitors must pay with their time to research policies in order to retain their privacy. We pose the question: if website users were to read the privacy policy for each site they visit just once a year, what would their time be worth?

Studies show privacy policies are hard to read, read infrequently, and do not support rational decision making. We calculated the average time to read privacy policies in two ways. First, we used a list of the 75 most popular websites and assumed an average reading rate of 250 words per minute to find an average reading time of 10 minutes per policy. Second, we conducted an online study of 212 participants to measure time to skim online privacy policies and respond to simple comprehension questions. We used data from Nielsen/Net Ratings to estimate the number of unique websites the average Internet user visits annually with a lower bound of 119 sites. We estimated the total number of Americans online based on Pew Internet & American Life data and Census data. Finally, we estimated the value of time as 25% of average hourly salary for leisure and twice wages for time at work. We present a range of values, and found the national opportunity cost for just the time to read policies is on the order of \$781 billion. Additional time for comparing policies between multiple sites in order to make informed decisions about privacy brings the social cost well above the market for online advertising. Given that web users also have some value for their privacy on top of the time it takes to read policies, this suggests that under the current self-regulation framework, targeted online advertising may have negative social utility.

2.1.1 Economic Theories of Privacy Policies

The FTC started with a set of principles, almost akin to a framework of rights, and encouraged companies to protect these rights by adopting privacy policies. Economists also see utility in privacy policies but from an entirely different basis.

Advertising economics looks at ways to turn a commodity (e.g., water) into a bundle of marketable attributes (e.g., from mountain springs). There are three types of attributes. *Search goods* are things readily evaluated in advance, for example color. *Experience goods* are only evaluated after purchase or use, for example the claims of a hair care product. *Credence attributes* cannot be determined even after use, for example nutrition content of a food. One argument for mandatory nutrition labels on food is that it converts nutrition information from a credence attribute to a search attribute: consumers can read the label prior to purchase [43]. This argument applies equally well to online privacy. Without a privacy policy, consumers do not know if a company will send spam until after they have made the decision to provide their email address. With a privacy policy, consumers can check privacy protections prior to engaging in business with the site.

Another economic perspective that leads to supporting privacy policies is that since privacy is not readily observable, it cannot be properly valued by the market place. Without privacy policies, companies have all of the information about their own practices and consumers have none, leading to an information asymmetry [153]. Information asymmetries are one potential cause of market failure. The canonical example is of a market for used cars: sellers know if their cars are in mint condition or are lemons, but buyers may not be able to tell [9]. Consequently, buyers need to take into account the risk of getting a bad car, and will not pay top dollar for a great car just in case they are being taken for a ride.

Privacy policies should help reduce information asymmetries because companies share information with their customers. However, researchers also note that if the cost for reading privacy policies is too high, people are unlikely to read policies. Time is one potential cost, and the time it takes to read policies may be a serious barrier [41]. This approach assumes rational actors performing personal benefit-cost analysis, at least on an implicit level, to make individual decisions to read or skip privacy policies [6]. If people feel less benefit reading policies than they perceive cost of reading them, it stands to reason people will choose not to read privacy policies.

One question then is what value to place on the time it takes to read privacy policies. There is a growing literature addressing the monetary value of time, starting in the mid-1960s [19]. For example, urban planners estimate the value lost to traffic jams when deciding if it makes sense to invest in new roads or other infrastructure improvements [89]. As benefit cost analysis increased in popularity, government agencies found they had a hard time calculating economic value for “free” services like parks. One way to address their value is to estimate the time people spend traveling to parks and the value of the time they spend enjoying the parks, which again requires estimates of the value of time [18]. We draw upon this body of work.

In this chapter we look at societal and personal opportunity costs to read privacy policies. Under the notion of industry self-regulation, consumers should visit websites, read privacy policies, and choose which websites offer the best privacy protections. In this way a market place for online privacy can evolve, and through competition and consumer pressure, companies have incentives to improve their privacy protections to a socially optimal level. In practice, industry self-regulation has fallen short of the FTC vision. First, the Internet is far more than commercial sites or a place to buy goods. While it may make sense to contrast the privacy policies of Amazon, Barnes and Noble, and OReilly to purchase the same book, there is no direct substitute for popular non-commercial sites like Wikipedia. Second, studies show privacy policies are hard to read [73], read infrequently [74], and do not support rational decision making [6].

Several scholars extended the FTCs vision of an implicit marketplace for privacy by examining ways to explicitly buy and sell personal information. Laudon proposed “[m]arket-based mechanisms based on individual ownership of personal information and a National Information Market (NIM) in which individuals can receive fair compensation for the use of information about themselves.” Under this plan, corporations could buy “baskets of information” containing the financial, health, demographic or other data that individuals were willing to sell about themselves [85]. Varian sees privacy as the “right not to be annoyed” and suggests web-based contracts to sell specific information for specific uses during a fixed time frame [152]. Yet no such market of micropayments for personal information exists. Garfinkel notes that in the current market place, where corporations re-sell information to other corporations, payments are already low. He estimates that payments to individuals for their information would be worth about a penny per name, which is far lower than most people would be willing to accept [56]. Since Garfinkel’s

analysis, the market for personal information has been flooded with readily available information. Even stolen information is worth only about a tenth of what it used to fetch on the black market [141]. Full clickstream data sells for only 40 cents per user per month [24], yet from the outrage when AOL released search term data to researchers [75], it is a good guess that most people value their data at a substantially higher rate than it currently sells for on the open market. With sellers demanding more than buyers will pay, there is no zone of possible agreement, and thus it is likely that no transactions would take place.

In this chapter we explore a different way of looking at privacy transactions. What if online users actually followed the self regulation vision? What would the cost be if all American Internet users took the time to read all of the privacy policies for every site they visit each year? We model this with calculations of the time to read or skim policies, the average number of unique websites that Internet users visit each year, and the average value of time, as we present in section 2.2. In section 2.3, we combine these elements to estimate the total annual time to read policies as well as the cost to do so, both for individuals and nationwide. We discuss our findings and present our conclusions in section 2.4.

2.2 Inputs to the Model

In this section we develop a model to estimate the cost to all United States Internet users if they read the privacy policy once on each site they visit annually. We model cost both in terms of time and the economic value of that time. We estimate the annual time to read ("TR") online privacy policies as

$$TR = p * R * n$$

p is the population of Internet users

R is the average national reading rate

n is the average number of unique sites an Internet user visits each year

Similarly, we estimate the time to skim ("TS") online privacy policies as

$$TS = p * S * n$$

S is the average time to skim a policy

We contrast reading to skimming because while some Internet users might read privacy policies all the way through, studies in our lab show that in practice, people may scan privacy policies for specific information they are interested in learning rather than reading policies word-for-word [118].

Estimating the economic value of time is more complex. As we discuss further, based on literature in the value of time domain, leisure time is valued at a lower hourly rate than value of loss of productivity during work hours. We estimate time at home as $1/4 W$ and time at work as $2W$ where W represents average wages. Consequently we estimate not just the annual number of unique websites, but also the proportion of sites that Internet users visit at home and at work.

2.2.1 Time to Read or Skim Privacy Policies

We used two different methods to estimate the average time to read online privacy policies. First, we took the average word length of the most popular sites privacy policies and multiplied that by typical words per minute (WPM) reading speeds. Second, we performed an online study and measured the time it took participants to answer comprehension questions about an online privacy policy. This allows us to estimate time and costs both for people who read the full policy word for word, and people who skim policies to find answers to privacy questions they have. In each case, we use a range of values for our estimates with median values as a point estimate and high and low values from the first and third quartiles.¹

Calculated Estimate to Read Popular Website Privacy Policies

We measured the word count of the 75 most popular websites based on a list of 30,000 most frequently clicked-on websites from AOL search data in October, 2005 [47]. Because these are the most popular sites, they encompass the sorts of policies Internet users would be most likely to encounter.

As seen in Figure 2.1, we found a wide range of policy lengths from a low of only 144 words to a high of 7,669 words— about 15 pages of text. We used a range of word count values from the first quartile to the third quartile, with the mean value as a point estimate.

¹In this chapter, the first quartile is the average of all data points below the median; the third quartile is the average of all data points above the median. These are single values and not a range of values. Point estimates are our single best guess in the face of uncertainty.

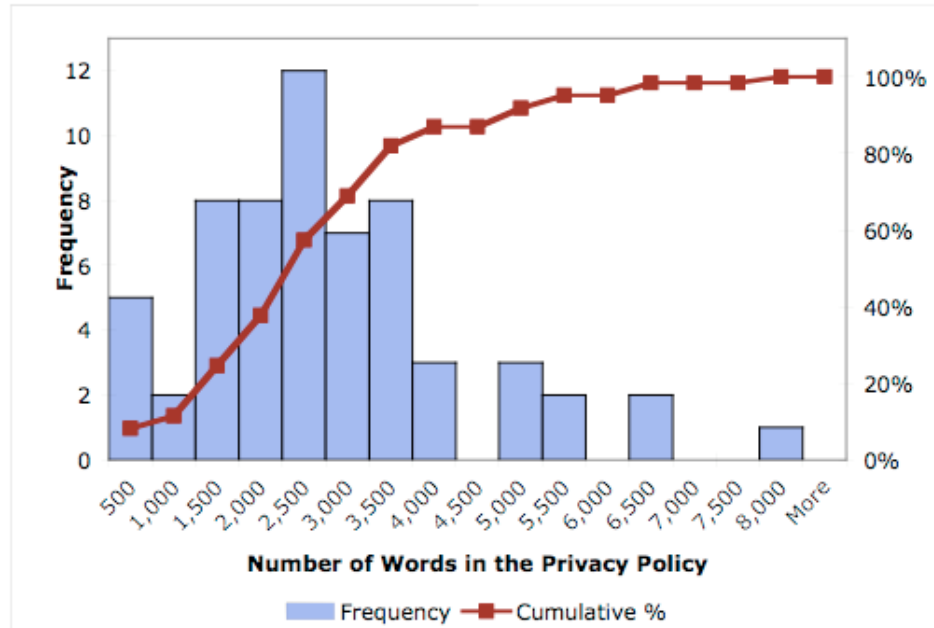


Figure 2.1: Probability Density Function (“PDF”) and Cumulative Distribution Function (“CDF”) of Word Counts in Popular Website Privacy Policies

We calculated the time to read policies as the word length of common privacy policies times 250 WPM, which is a typical reading rate for people with a high school education [31].

Table 2.1: Times to read entire privacy policies for average readers

	Word Count	Reading Rate	Time to Read One Policy
First Quartile	2,071	/ 250 WPM	= 8 minutes
Median	2,514	/ 250 WPM	= 10 minutes
Third Quartile	3,112	/ 250 WPM	= 12 minutes

As seen in Table 2.1, we find that it takes about eight to twelve minutes to read privacy policies on the most popular sites, with a point estimate of ten minutes per policy. These estimates may be slightly low due to the jargon and advanced vocabulary in privacy policies. In addition, some people read more slowly online than on paper, which may also make these time estimates slightly low.

Measured Time to Skim Policies

Internet users might be more likely to skim privacy policies to find answers to their questions, or to contrast between two policies, rather than to read the policies word-for-word as envisioned in the prior section. We performed an online-study that asked participants to find the answers to questions posed about privacy protections based on the text of a privacy policy. We based our questions on concerns people have about online privacy, as studied by Cranor et al [41]. We asked five questions including “Does this policy allow Acme to put you on an email marketing list?” and “Does the website use cookies?” All answers were multiple choice, rather than short answer, so the act of answering should not have substantially increased the time to address these questions.

To ensure our results were not overly swayed by one unique policy, participants were presented with one of six different policies of varying lengths. In all, we had 212 participants from which we removed 44 outliers.² We found that the time required to skim policies does not vary linearly with length, as seen in Figure 2.2. We selected one very short policy (928 words), one very long policy (6,329 words) and four policies close to the typical 2,500 word length. The median times to skim one policy ranged from 18 to 26 minutes. The lowest first quartile was 12 minutes; the highest third quartile was 37 minutes. The three policies clustered near 2,500 words ranged in median times from 23 to 24 minutes and did not show statistically significant differences in mean values.³

In a prior study, we asked 93 participants to read an online privacy policy from a publishing site—the same very short 928 word policy. We asked very similar questions but included two additional questions and omitted the time to answer the first question as a training task. We found a far lower time: a point estimate of six minutes to scan a privacy policy and find relevant information. This reflects an artificially low time because, as we have since discovered, the majority of time spent answering questions is devoted to the very first question. Even though our follow up

²During online studies, participants are sometimes distracted by other tasks. We eliminated data points that were clearly implausible, for instance, taking 5 hours to complete a set of tasks that typically takes 20 minutes. In similar studies we have also seen responses indicative of clicking through the answers without reading the text. While we did have a few very speedy respondents that could mathematically be identified as outliers, we chose to retain them. For example, 3 minute response time is possibly the product of someone unusually good at the task, rather than someone who did not attempt to understand the material. In short, we favored removing and retaining outliers in ways that could slightly underestimate the times we measured.

³We contrasted the 2,550 word policy to the three similar length policies using two-sided t-tests assuming unequal variance; 95% confidence interval; $p=.518, .690, .891$.

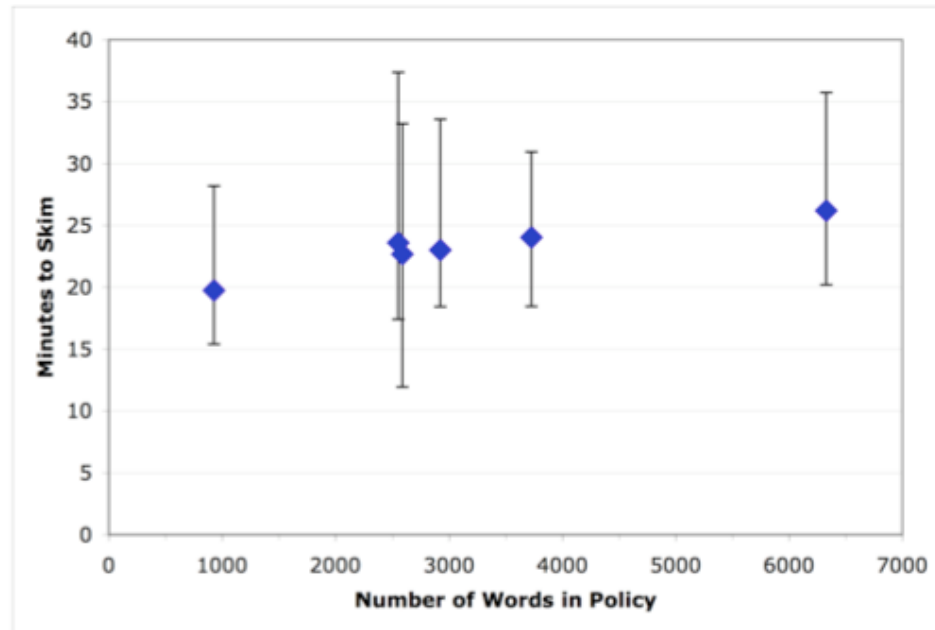


Figure 2.2: Median times and inter-quartile ranges to skim one privacy policy

study started with a basic question, participants typically spent a third to half of their time on the very first question.

Arguably a good lower estimate of the time it takes to skim one policy is to look at the inverse of our first study: just look at the time for the first question, provided it is a question that encourages exploring the full policy. In our second study we always started with a warm up question that asked participants to identify the street address for the company and that information was always in the last few lines of the policy. Participants had to skim the full policy to answer the question. As shown in Figure 2.3, median times ranged from four minutes to eight minutes. The lowest first quartile of all six policies was 4 minutes; the highest third quartile was 12 minutes.

One disadvantage to using just the time for the first question is that it underestimates because we only look at one question, and a very basic question at that. When asked to identify why they read privacy policies, our participants volunteered multiple interests ranging from data security, to information sales, to spam, to opt-out policies. These are captured better in the range of times reported in Figure 2.2. However, one advantage to using just the time for the first question is we eliminate the unsatisfying situation that we can generate longer or shorter overall time estimates just by varying the number of questions we ask.

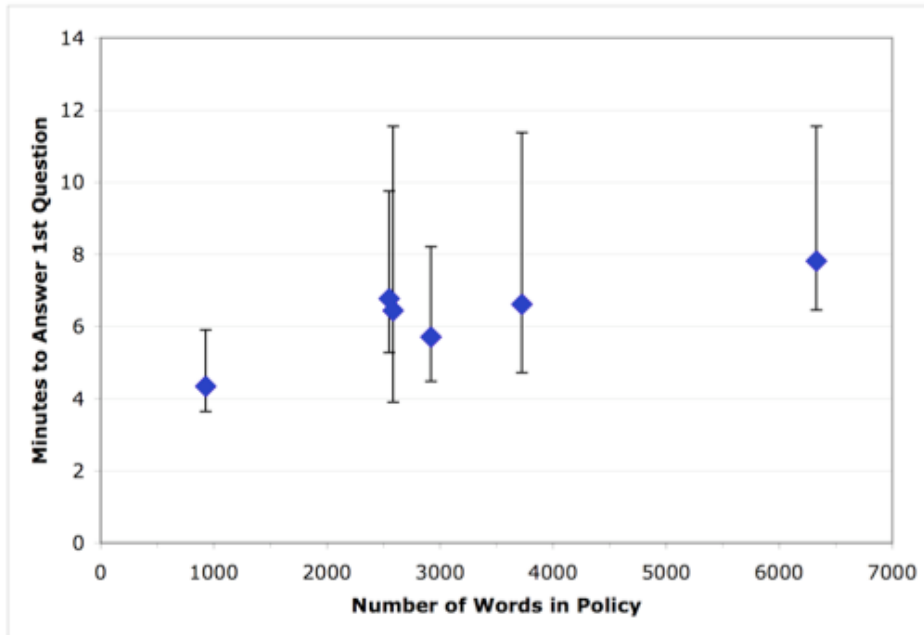


Figure 2.3: Median time to answer a basic question in one of six policies of different lengths, bracketed by interquartile range

We elected to report the more conservative estimates from just looking at the times to answer the first question, with the caveat that these numbers are lower estimates. If people were to read policies regularly, presumably they would get faster at finding information, which is another argument for a more conservative approach. We used the lowest first quartile and highest third quartile for our low and high estimates. We averaged the policies medians as our point estimate; see Table 2.2.

Table 2.2: Time estimates to skim one policy and answer a basic question

	Skim and Answer One Question
Low Estimate	3.6 minutes
Point Estimate	6.3 minutes
High Estimate	11.6 minutes

2.2.2 Monthly Number of Unique Websites Visited

Nielsen Online reported the average number of unique websites that United States Internet users visited at home and at work during March, 2008 as 66 unique sites from work and 119 from home

[106]. The overall average number of unique sites visited per person for the same time period was 105 [105]. The overall figure is lower than the sum of sites visited from work and home because there is duplication. For example, imagine someone who visits Google both at work and at home. Google would appear once in the count of unique sites visited at work, plus once in the count of the unique sites visited at home, yet only be one unique site overall. As depicted in Figure 2.4, on average Internet users visit 52 different sites exclusively at work, 105 different sites exclusively at home, and 14 sites at both work and home.

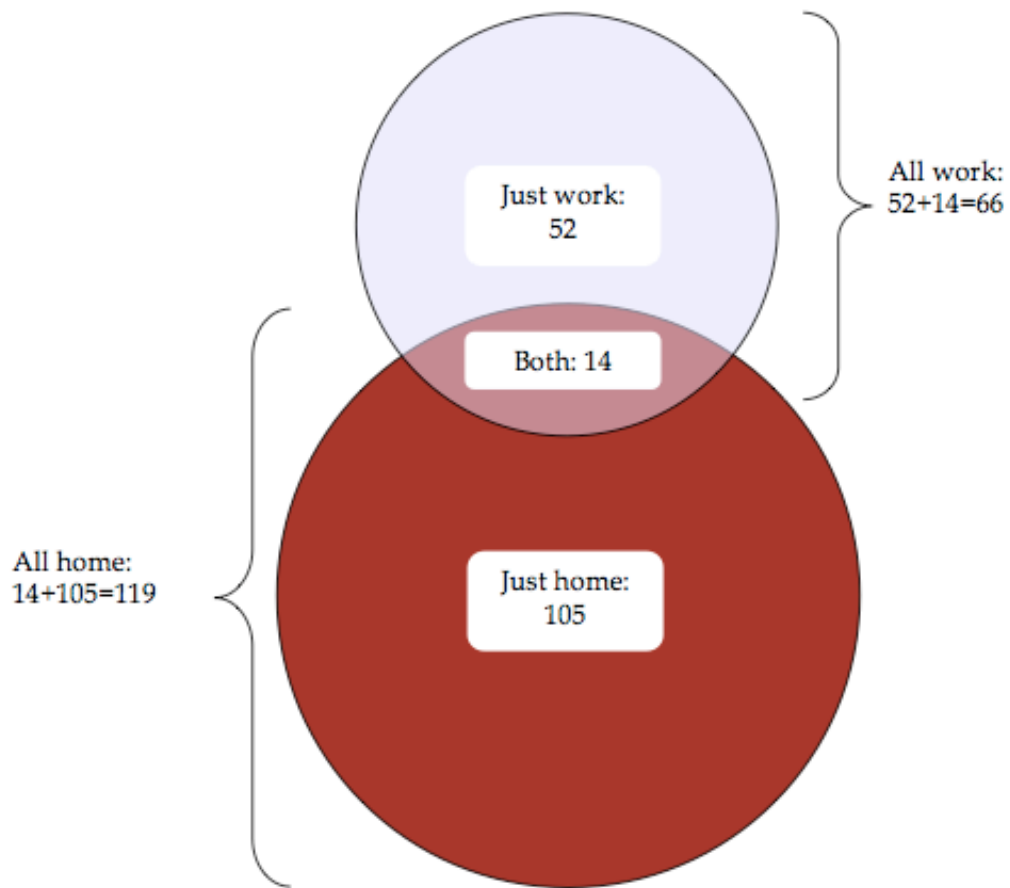


Figure 2.4: Locations where people read websites

We assume that if people read privacy policies, they would read them the first time they encountered a given site. We do not know where people first see the sites they visit both at work and at home. This uncertainty does not affect our time estimates but does affect our estimates

for the value of that time, since time at work has a higher economic value than leisure time. As a lower bound estimate, we assume all of the sites visited at both locations are first encountered at home. As an upper bound estimate, we assume all of the sites visited at both locations are first encountered at work. For our point estimate, we split the difference and assume half are first encountered at work and half at home. These estimates are summarized in Table 2.3.

Table 2.3: Estimates of the monthly number of unique websites visited by U.S. Internet users

Estimate	Policies read at work	Policies read at home
Lower bound	52 / month	119 / month
Point estimate	59 / month	112 / month
Upper bound	66 / month	105 / month

2.2.3 Annual Number of Unique Websites Visited

Unfortunately, Nielsen does not collect data on the average number of websites people visit annually. They do collect weekly statistics, as shown in Table 2.4:

Table 2.4: Unique monthly and weekly websites visited by U.S. Internet users show repeat visits to many sites week after week

Location	Unique sites / month	Unique sites / week	Scale factor
Work	66	25	66%
Home	119	40	74%

People visit some of the same sites each week: if not, we would see 100 unique sites per month at home ($25 * 4$ weeks) rather than 66 (see Table 2.4). Ideally we would only count such sites once. From the Nielsen data we computed a scale factor, which is the percentage of sites that Internet users return to week after week. While our scale factor may not actually scale linearly over a full year it is a reasonable starting point for estimation.

We are unaware of any scholarly work that measures how many websites people visit annually. However, a 2008 study examined 25 subjects over a variable length of time and found an average of 390 unique sites during 52 to 195 days of observation [155]. The mean length of observation was 105 days. Using our point estimate of 112 unique sites per month, 390 unique sites suggests nearly all new sites each month. It seems more likely that these 25 participants,

drawn from the researchers acquaintances, simply visited more sites per month than the Nielsen population. We can draw no firm conclusions. But this study does suggest, even if anecdotally, that our scale factor is not absurdly low. If anything, we may be conservative in our estimates.

For all annual estimates, we first multiplied the monthly estimate by 12 to convert from months to years, and then multiplied by the appropriate scale factor to account for visitors returning to the same sites month after month. Scale factors varied by type of estimate. As a lower bound estimate for the average annual number of websites visited we multiplied by our lower observed scale factor, .66. As an upper bound annual estimate we multiplied our upper monthly estimate by our higher observed scale factor, .74. For our point estimate we used a weighted average of the observed scale factors, multiplying the monthly average work sites by the observed work scale factor of .66 and home by .74. The results are summarized in Table 2.5.

Table 2.5: Estimates of the annual number of unique websites visited by U.S. Internet users

Estimate	Scale factor		Read at work		Read at home		Total
Lower bound	.66	– >	412 / year	+	942 / year	=	1354 / year
Point estimate	Weighted	– >	467 / year	+	995 / year	=	1462 / year
Upper bound	.74	– >	586 / year	+	932 / year	=	1518 / year

2.2.4 Opportunity Cost of Time

Just as the opportunity cost of time in school is a major part of the overall cost of education, Becker argued we should consider the opportunity cost of time as an implicit cost of goods and services [19]. The cost to see a play is not just the price of admission, but also the value that audience members place on their own time [19]. Economics literature suggests that time should be valued as salary plus overhead, which is the value corporations lose [89]. In the United States, overhead is estimated as twice the rate of take home pay [82]. However, that approach may not be an accurate reflection for those who work a fixed number of hours or are not in the workforce [18]. Through revealed-presences and willingness-to-pay studies, studies estimate people value their leisure time at one quarter of their take home pay [89].

Taken together, this suggests that reading privacy policies at work should be valued $2W$ while reading privacy policies at home should be valued as $1/4W$, where W is average wages. The

Bureau of Labor Statistics finds an average hourly wage of \$17.93 for March, 2008 [28]. That gives us estimates of \$35.86/hour for the opportunity cost of reading privacy policies at work and \$4.48/hour for the opportunity cost of reading privacy policies at home as seen below in Table 2.6.

Table 2.6: Estimates for the value of time to read online privacy policies

Location	Average value of time
Home	\$ 4.48 / hour
Work	\$ 35.86 / hour

2.3 Time and Economic Value to Read Privacy Policies

In this section we use the inputs from section 2.2 to estimate how much time it would take for an individual to read the policies of each website she visits annually. We then use those time estimates as the basis for calculating the value of that time. In both cases we look at national figures as well as individuals.

2.3.1 Amount of Time to Read Privacy Policies

We multiplied the estimates for the number of unique sites American Internet users visit annually (section 2.2.2) by the time to read or skim privacy policies (sections 2.2.1 and 2.2.1) and by the estimated 221 million Americans online [105].

Table 2.7: Annual time estimates for reading and skimming online privacy policies

Estimate	Individual time to read	Individual time to skim	National time to read	National time to skim
Lower bound	181 hrs / yr	81 hrs / yr	39.9B hrs / yr	17.9B hrs / yr
Point Estimate	244 hrs / yr	154 hrs / yr	53.8B hrs / yr	33.9B hrs / yr
Upper bound	304 hrs / yr	293 hrs / yr	67.1B hrs / yr	64.8B hrs / yr

We estimate that if all American Internet users were to annually read the online privacy policies word-for-word each time they visited a new site, the nation would spend about 54 billion hours reading privacy policies (see Table 2.7.) To put these figures in perspective, using the point

estimate of 244 hours per year to read privacy policies per person means an average of 40 minutes a day. This is slightly more than half of the estimated 72 minutes a day people spend using the Internet [104]. This exceeds the combined percentage of Internet time devoted to shopping (1.9%) dealing with spam (6.2%) and playing games (13%) in 2005 [104]. The estimated time to read privacy policies exceeds the percentage of time online that people currently spend surfing the web (45.3%) [104]. One study estimates the time lost to delays in booting computers with adware as 60 hours per year per infected user, or about a quarter of the time we estimate to read privacy policies [140]. In 2000, federal income tax payers spent an estimated average of 26.4 hours completing their income taxes and nationwide, U.S. tax payers spent 3.4 billion hours completing federal income taxes [62]—several times less than the amount of time we estimate for reading online privacy policies.

2.3.2 Value of Time to Read Privacy Policies

We multiplied the time to read or skim policies by the number of websites visited at work and the value of time at work, and added that value to the result from the same procedure for policies at home. For national costs, we again estimated 221 million Americans online [105].

We estimate that if all American Internet users were to annually read online privacy policies word-for-word each time they visited a new site, the nation would lose the value of about \$781 billion from the opportunity cost value of the time to read privacy policies. Again, to put this in perspective, in 2005 the average cost to connect to the Internet was \$237/year for dial up and \$508/year for high speed access [125]. This suggests the value of time lost to reading privacy policies would eclipse the cost of high speed Internet access, several times over. In 2007, United States online sales were approximately \$260 billion [37]—more than the cost to businesses if their employees were to read privacy policies on corporate time.

2.4 Discussion

We estimate that reading privacy policies carries costs in time of approximately 201 hours a year, worth about \$3,534 annually per American Internet user. Nationally, if Americans were to read online privacy policies word-for-word, we estimate the value of time lost as about \$781 billion

annually.

These estimates presume that people visit sites, read the policies once a year, and then carry on their business as before. Yet the FTC vision of self-regulation presumes that, at least for consumer sites, Internet users will visit multiple sites to comparison shop for acceptable privacy practices. The true cost of adherence to the self-regulation vision is perhaps on the order of double the costs we estimate, depending on which percentage of sites have ready substitutes and how many sites people are expected to compare. True costs also include Internet connectivity fees, which we did not attempt to quantify.

In the opposite direction, media consolidation means that multiple sites may share one privacy policy. While consolidation itself poses increased threats to online privacy, in some cases it may actually reduce the cost of reading privacy policies because there are fewer unique policies to read. We do note that the resulting privacy policy when companies merge may be more complex and longer than either of the individual policies. Another issue is that people may not care about all possible privacy threats. For instance, if they only care about credit card theft, and they visit a site that does not collect credit card numbers, they may not feel the need to protect any information. Thus, arguably, they do not need to read the policy at every site they visit, but only a subset of sites.

The value of all online advertising in the United States was about \$21 billion in 2007 [71]. Many, though by no means all, online privacy concerns stem from advertisers amassing information about Internet users in order to present ads targeted to specific demographics. The current policy decisions surrounding online privacy suggest that Internet users should give up an estimated \$781 billion of their time to protect themselves from an industry worth substantially less. This is not to say online advertising should be banned. Sales from direct mail are approximately an order of magnitude higher than advertising costs and the cost of online advertisements similarly understates the full market [110]. But it appears the balance between the costs borne by Internet users versus the benefits of targeted ads for industry is out of kilter, at least as envisioned by the FTCs solution that Internet users read privacy policies.

Some Internet users may realize a benefit from targeted advertisements; for example Amazons ability to suggest additional books they might enjoy based on prior purchase history. Yet on the whole, advertisements are usually seen as an economic “bad” rather than a “good” because

participants would pay money to eliminate ads from most types of media [20]. While an analysis of the net social welfare changes created by online advertisement is beyond the scope of this work, we do suggest that any such cost-benefit analysis should include the value of time for reading privacy policies. Preliminary work from a small pilot study in our laboratory revealed that some Internet users believe their only serious risk online is they may lose up to \$50 if their credit card information is stolen. For people who think that is their primary risk, our point estimates show the value of their time to read policies far exceeds this risk. Even for our lower bound estimates of the value of time, it is not worth reading privacy policies. This leads to two implications. First, seeing their only risk as credit card fraud suggests Internet users likely do not understand the risks to their privacy. As an FTC report recently stated, “it is unclear whether consumers even understand that their information is being collected, aggregated, and used to deliver advertising [52].” Second, if the privacy community can find ways to reduce the time cost of reading policies, it may be easier to convince Internet users to do so. For example, if we can help people move from needing to read policies word-for-word and only skim policies by providing useful headings, or if we can offer ways to hide all but relevant information—and thus reduce the effective length of the policies—more people may be willing to read them.

The privacy community and industry groups have responded with several attempts to improve privacy policies. Layered privacy notices specify a few high-level and standardized topics for a one-screen summary of the policy, then link to the full privacy policy for more information [32]. The Platform for Privacy Preferences (“P3P”) is an XML-based specification that enables policy authors to code privacy policies in machine-readable format [154] which fosters comparison between policies in a standardized way, and provides a common format for user agents to help Internet users find acceptable policies. Privacy Bird is a web browser add-on that uses P3P to generate a short privacy report that presents information in bulleted lists with sections that expand and contract to show and hide sections of the privacy policy [41]. The P3P Expandable Grid is also built on P3P and uses icons to convey what information companies collect and how they use it [118]. Icons in the Privacy Finder search engine convey how well a given P3P policy matches users preferences. A Privacy Finder user study demonstrated that Internet users will pay a premium for products from sites rated as more privacy protective [146]. Both education and enhanced privacy policy formats may help Internet users gain the tools they need to protect

themselves online.

Finally, some corporations take the view that their users should read privacy policies and if they fail to do so, it is evidence of lack of concern about privacy. Instead, we counter that websites need to do a better job of conveying their practices in useable ways, which includes reducing the time it takes to read policies. If corporations cannot do so, regulation may be necessary to provide basic privacy protections. Disclosure legislation may be insufficient: adding more text to policies that most consumers do not read does increase transparency, but may otherwise be of limited practical utility.

Chapter 3

A Contrast of Formats and Lengths Using One Company's Privacy Policy

A subset of results in this chapter appear in a paper co-authored with Robert W. Reeder, Patrick Kelley, and Lorrie Faith Cranor presented to the 2008 Workshop on Privacy in the Electronic Society (WPES) [118].

3.1 Introduction

The United States relies on a self-regulation approach to Internet privacy. There are some Internet privacy laws, for example the Children’s Online Privacy Protection Act of 1998 (COPPA), which protects children’s privacy[39], and the Gramm-Leach-Bliley Act (GLB), which applies to financial institutions [60] . But by and large the theory of Internet privacy hinges on two assumptions:

- Consumers will choose companies with acceptable privacy policies.
- Federal Trade Commission (FTC) action for deceptive practices will reign in egregious abuse.

In both cases privacy policies play a vital role in Internet privacy. Free market mechanisms based in consumer choice will fail to protect privacy if consumers do not understand the choices available to them.

Several studies frame willingness to read privacy policies as an economic proposition and conclude that asymmetric information is one reason why people find it not worth their time to read privacy policies [153][6] . Other studies show that privacy policies require a college reading level to understand [65][127]. A study of ambiguities in privacy policies shows they contain “weasel words” and language that downplays privacy issues [115]. These studies all support the notion that increasing ease of readability will improve privacy policies usability and accessibility. In response to these issues, privacy researchers have devised several standardized formats for privacy policies based on the expectation that standardized formats would reduce confusion and improve comprehension for consumers. This study is a comparative analysis to analyze how well standardized policies work in practice.

In section 3.2 we describe the formats we contrasted as well as our study methodology. We present our results in section 3.3. In section 3.4 we discuss implications from these results and conclude in section 3.5.

3.2 Study Design

This section introduces attempts at improving privacy policies as well our methods. In addition to a conventional natural language (NL) policy, we analyzed three standardized formats: layered

policies, the Privacy Finder privacy report (PF), and the P3P Expandable Grid (EG), all of which are described below. We also contrasted three different lengths, from long to short.

3.2.1 Formats

The Platform for Privacy Preferences (P3P) is a standardized format for privacy policies, and is formally recommended by the World Wide Web Consortium (W3C)[154]. P3P provides a taxonomy in which privacy policies can be expressed in XML (eXtended Markup Language), which is computer readable, and thus allows software tools to help people manage their privacy preferences. In our study P3P formed the common basis for different forms of privacy policies we presented in the Privacy Finder and the P3P Expandable Grid formats. By working from a common P3P source we know that participants are responding to differences in presentation, rather than differences in content.

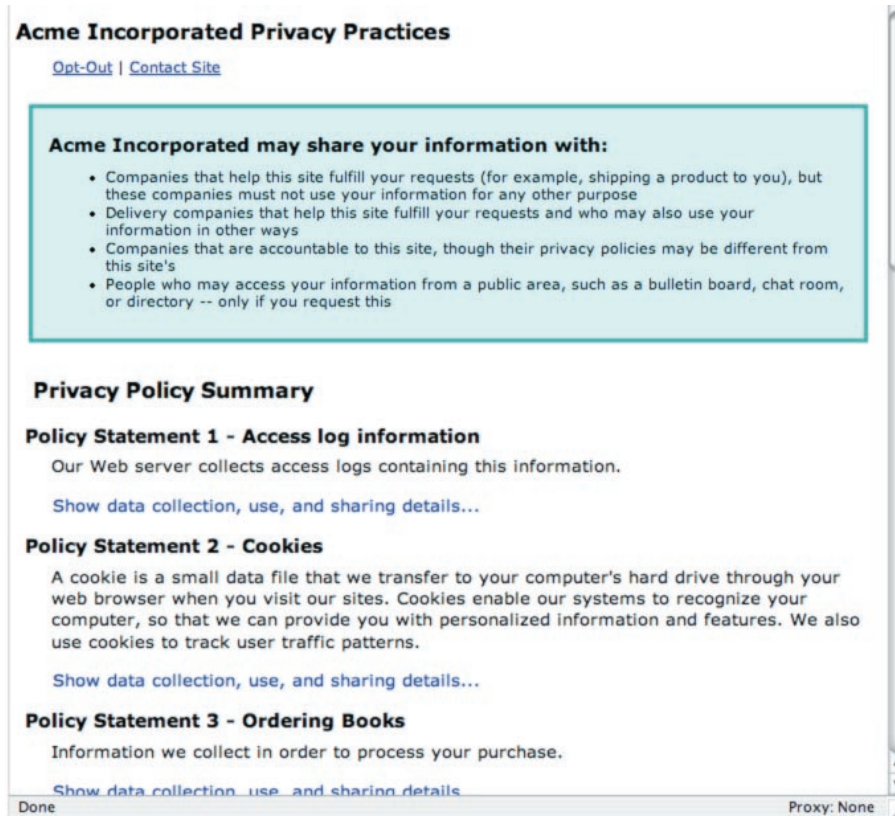
Privacy Finder

Privacy Finder was developed by AT&T and refined at the CMU Usable Privacy and Security (CUPS) laboratory. Privacy Finder has many components including a privacy report which is the section we tested. Privacy Finder's privacy report was designed to avoid many of the problems that stem from free-form natural language policies by generating standardized text from P3P policies. This avoids "weasel words" and ensures uniform presentation. Privacy Finder highlights the most important information at the top of the report and provides links to expand details.

The P3P Expandable Grid

The P3P Expandable Grid is a format developed by CUPS and IBM. The P3P Expandable Grid presentation is, like Privacy Finder, based on P3P. However instead of generating standardized text, the P3P Expandable Grid presents a series of icons to denote which information is collected, shared, and so forth, in a table. The P3P Expandable Grid format consolidates information and users can click rows or columns to get more information.

Figure 3.1: Privacy finder policy



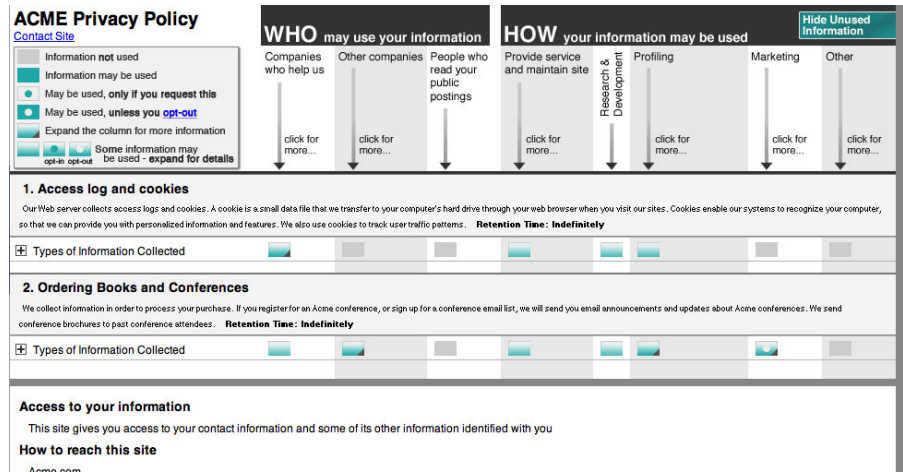
Layered Notices

The law firm Hunton & Williams popularized the notion of layered notices [137] so privacy policies look more like nutrition labels: brief, standardized, and easy to compare directly. The first layer provides a short overview. This layer requires standardized headings to ensure readers can find important information. Although the text within each section is free form, layered policies are typically only about a screen of text. As a result of this brevity the first layer omits many details and links to the second layer, which is a full natural language policy.

Natural language

Most privacy policies today are in natural language format: companies explain their practices in prose. One noted disadvantage to current natural language policies is that companies can choose which information to present, which does not solve the problem of information asymme-

Figure 3.2: The P3P Expandable Grid policy



try between companies and consumers. Further, companies use what have been termed “weasel words” — legalistic, ambiguous, or slanted phrases — to describe their practices. Because this format is the current status quo, natural language policies function similarly to a control condition.

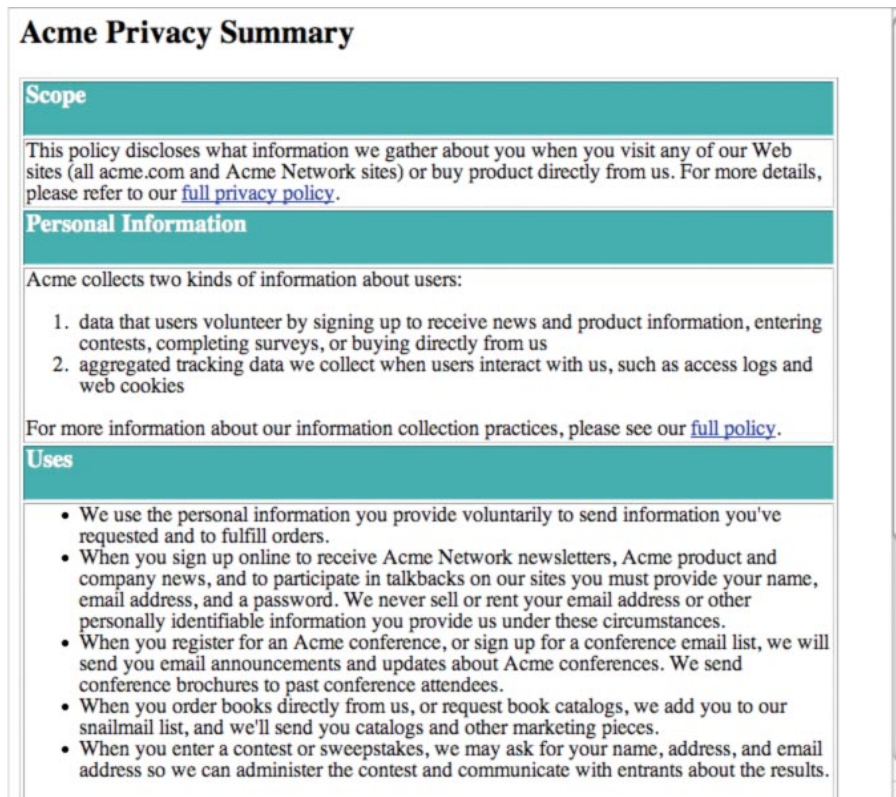
3.2.2 Length

The Gramm-Leach-Bliley Act (GLB) requires financial institutions send annual notices about their privacy policies to their customers [50]. In response to criticisms of GLB, six federal agencies commissioned a report from Kleimann Communication Group [119]. The Kleimann report concludes “...when faced with complex information, they often won’t even bother to read.”[119] Their final recommendation includes a standardized overview on the first page and further details on the second page, keeping the length of the policy limited. Length appears to be a key factor for both comprehension and willingness to read.

In order to test the hypothesis that policy length affects which presentation is most usable, we selected a long policy and pared it down into medium and short versions. We chose to modify one policy to isolate just changes in length. For example, if we had different policies that mentioned cookies in different places within the policy, we might see different timing results due to ordering within the policies.

All of the privacy policies we considered have a P3P policy as well as a natural language

Figure 3.3: Layered policy



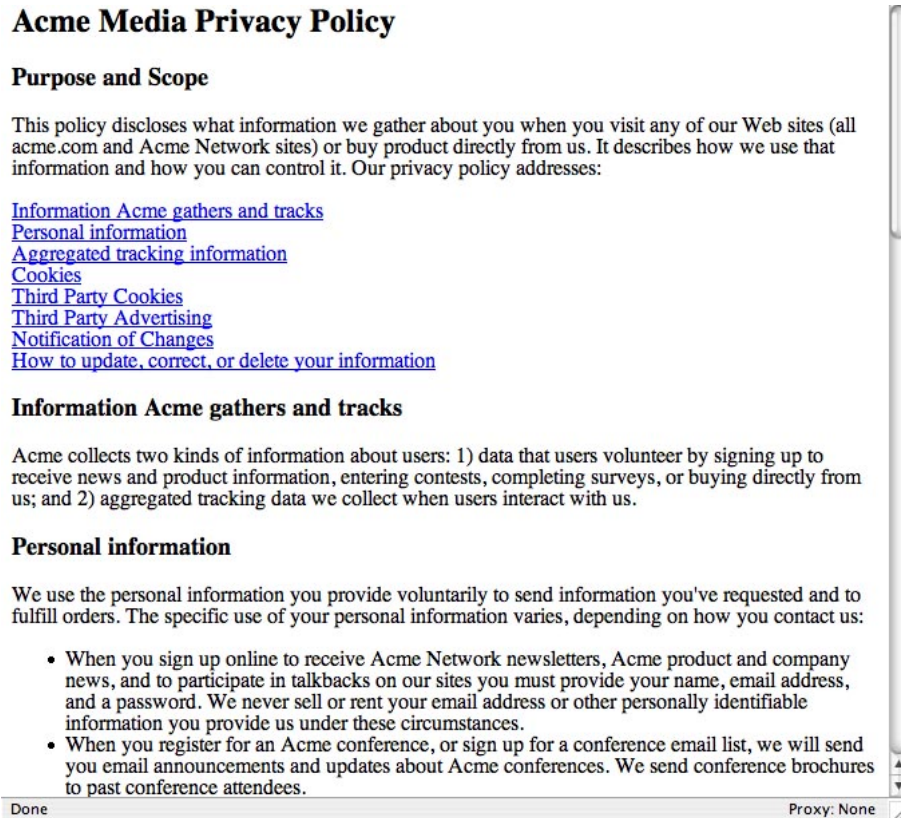
policy. Prior work finds “P3P policies are generally representative of all website privacy policies and therefore provide a useful data source for website privacy policy studies.” [40] We obtained the database of all of Privacy Finder’s cached P3P policies as of 4/15/07.¹

Of 20,476 total initial P3P policies, we eliminated duplicates, policies with unrecoverable syntax errors, those not written in English, sites not based in the United States, and sites from industries subject to privacy policy regulation (for example, medical sites are subject to HIPA). We did this so we could examine non-regulated privacy policies in the United States. Our final database contained 294 P3P policies, of which we selected one.

The policy that best fit our criteria for modular length with few changes needed to the remaining text to make it shorter, and a good test of the P3P Expandable Grid format was the publishing company O’Reilly. The O’Reilly policy has eight data statements, putting it in the top 99%, and we

¹Privacy Finder matches Google and Yahoo! search results to visitors’ privacy preferences. See <http://www.privacyfinder.org/>

Figure 3.4: Natural language policy



used it almost verbatim with no substantive changes in content. We were able to edit it to shorter versions just by reducing the number of statements and cutting to corresponding sections in the natural language policies. We changed all occurrences of O'Reilly to Acme, and also changed the opt-out link to point to a page we controlled so participants would not find it out was actually the O'Reilly site. We also removed logos and used a generic color scheme and standard font choices.

3.2.3 Summary of Conditions

The number of participants for each of the ten conditions ranged from 77 to 95:

Note that while Privacy Finder and the P3P Expandable Grid have three variants, each corresponding to a length from the Natural Language policies, we tested only one Layered policy. This is because we would have identical Layered policies on top of any of the three Natural Language variants, so there was little value in testing the Layered variant more than once.

Formats	Length		
	Long	Medium	Short
Natural Language (NL)	n = 94	n = 95	n = 90
Privacy Finder (PF)	n = 85	n = 84	n = 89
The P3P Expandable Grid (EG)	n = 81	n = 77	n = 84
Layered	n = 86		

Table 3.1: Number of participants per condition, $\sum n = 865$

3.2.4 Methods

We collected the data for this study from August 21 to October 1, 2007. 865 people completed an on-line study. We posted advertisements on craigslist, sweepstakes websites, mailing lists, with Google adwords, and used personal networks to recruit participants. We offered a lottery for a \$250 Amazon gift certificate as incentive for participating in the study.

While we limited advertisements to the United States to capture domestic views of privacy, we did not use nationality or location as exclusion criteria. We were unable to accommodate Internet Explorer on OS X; when we detected that combination we displayed a message asking users to try again with a different browser. Participants who did not have a screen resolution of at least 1024x768 were ineligible for the study, which excluded a small population with very old hardware. The study required Javascript, so for all users with Javascript disabled we provided directions to enable Javascript and at the end of the study provided directions on how to disable Javascript again. While the Javascript requirement did not exclude any participants, some may have self-selected out of the study rather than enable Javascript. We excluded participants under 18 years of age. Beyond these few exclusion criteria the study was open to all.

We ran a between-group design and assigned each participant to one of ten privacy policy representations. We used a between-group design rather than within group design for two reasons. First, in this context it is unrealistic to eliminate learning effects simply by reordering policies. Second, it is unrealistic to expect participants to spend more than 20 minutes unless they are highly compensated, which exceeded our budget. Questions remained constant over all conditions; only the policy changed. The study was online, which is the most natural setting to study online privacy policies.

Study Questions

Study questions comprised several groups:

- *Comprehension.* Participants answered a series of multiple choice questions to determine how well they were able to understand the policy. For example, we asked “Does the Acme website collect your Social Security number?” These questions are realistic information retrieval tasks based on typical privacy concerns, and are similar to questions used in a 2006 study by Cranor et al [41]. We conducted three rounds of pilot tests with over two dozen people to ensure the questions were well-worded and understandable. We randomized the order of these questions to mitigate learning effects and captured both accuracy and time to respond. We also included a warm-up task which we did not score. This training task helped participants gain familiarity with the privacy policy format so we did not unfairly disadvantage unfamiliar approaches.
- *Psychological Acceptability.* Saltzer coined the term psychological acceptability to convey that if people do not like a system they will not use it. He wrote, *It is essential that the human interface be designed for ease of use, so that users routinely and automatically apply the protection mechanisms correctly.* [124] Participants answered a series of questions about what they thought of the privacy policy they saw. These subjective responses were on a seven-point Likert scale to capture reactions to questions like “I believe Acme will protect my personal information more than other companies.”
- *Demographics.* We collected basic information like gender, educational attainment, and income so we could understand how closely our study population resembles Internet users as a whole.

We also measured the time it took for participants to answer each of the comprehension questions.

Analysis

We performed a comparative analysis across all four formats (Natural Language, Privacy Finder, the P3P Expandable Grid, and Layered) and all three lengths (Long, Medium, Short) to see if there

were statistically significant differences in the mean scores for accuracy, time to completion, and psychological acceptability questions.

Accuracy questions were categorical data so we used Chi Squared tests. We used a two sided t-Test between the means of the long natural language policy and the layered policy because of the natural pairing between these two conditions. We performed ANOVA analysis on the main nine variants for time data and psychological acceptability, which we recorded as a seven point Likert scale and treated as continuous variables. We performed all tests of statistical significance at the $\alpha = 95\%$ confidence level.

In addition to significance we also calculated effect size, which is a standardized measure of how important an effect is [38]. While statistical significance establishes that the difference between two mean values is likely not due to random chance, effect size gets closer to answering if that difference is large enough to be concerned with. Additionally, because effect size is dimensionless it allows comparison across multiple domains, in this case allowing us to contrast accuracy to task completion and effect sizes even though these three measures do not use the same scale. Several authors have suggested various cut off values for small, medium, and large effect sizes; Cohen suggests .2, .5, and .8 respectively [38]. These thresholds appear to be reasonable heuristics for our data.

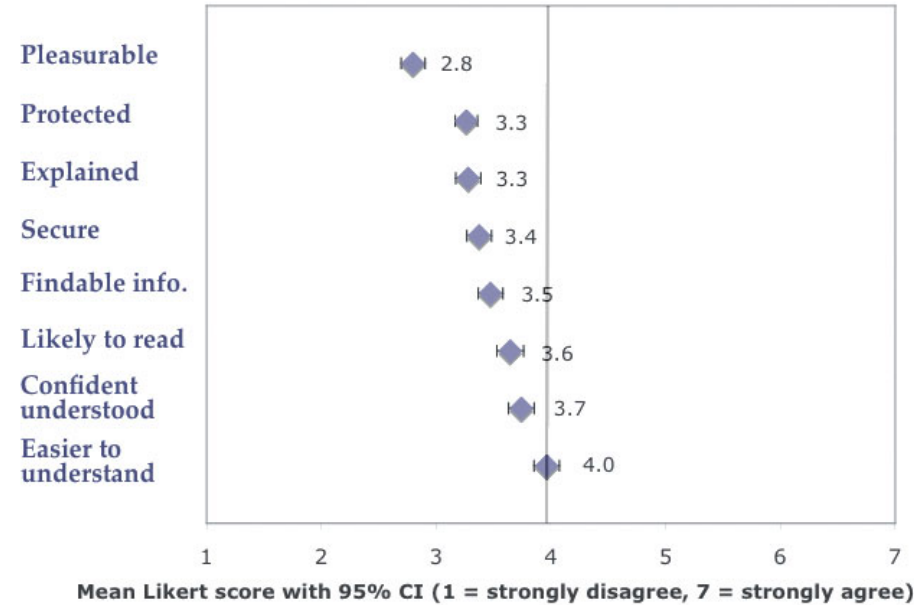
3.3 Results

Of the four formats we tested, we found Natural Language and Layered were nearly indistinguishable. Privacy Finder performed almost as well with small effect sizes for most differences. The P3P Expandable Grid trailed all other formats. While we expected large differences due to policy length, we found few significant differences.

3.3.1 Overall

We found that none of the privacy policy variants scored well for psychological acceptability. Accuracy varied from quite good (95% answering correctly) to just better than chance. In general, participants were able to answer questions fairly quickly.

Figure 3.5: Overall mean Likert scores for psychological acceptability across all conditions. Vertical bars denote 95% confidence intervals around the point estimates.



Psychological Acceptability

Across all nine primary conditions,² the mean scores for psychological acceptability reflect dissatisfaction with privacy policies. As shown below, the best score was for understandability and it scored a neutral 4 on the Likert scale. A subsequent lab study employing a think aloud protocol finds that participants reason that since they do not read many policies and cannot tell if the policy they saw is better, they tend to choose 4 in the absence of a strong opinion. The question that had the lowest score cuts to the heart of psychological acceptability. Finding information was pleasurable had an overall average of only 2.8, indicating disagreement with the statement.

Accuracy

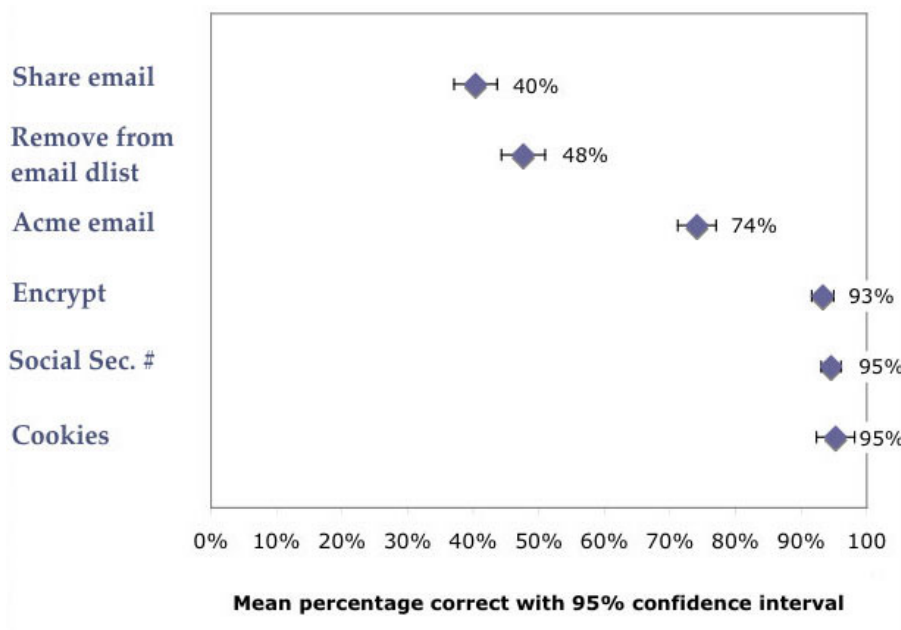
Across all nine primary conditions,³ the mean scores for the accuracy questions show some tasks were significantly more difficult than others.

All but one question was multiple choice with only three possible answers (Yes, No, The policy

²We omitted the layered policy to avoid unfairly weighting the results toward the long natural language policy, which would effectively be tested twice if we included layered.

³We omitted the layered policy to avoid unfairly weighting the results toward the long natural language policy, which would effectively be tested twice if we included layered.

Figure 3.6: Overall mean percentage of participants who answered the question correctly, across all conditions. Vertical bars denote 95% confidence intervals around the point estimates.



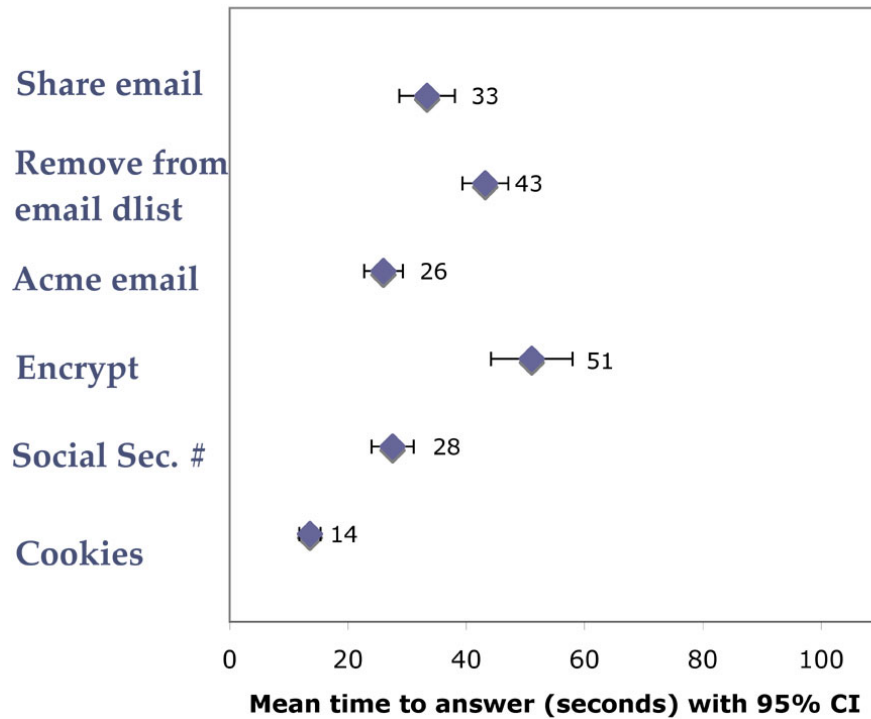
does not say). The worst performance on the question “Does this privacy policy allow Acme to share your email address with a marketing company that might put you on their email marketing list?” (40%) barely outperformed random chance (33%). However, there were five options for “How can you remove yourself from Acme’s email list?” and a correct answer selected three of five.

Task Completion Times

Unlike other sections, task completion times posed issues with outliers. The longest time to answer all of the six questions was nearly three days. Based on pilot studies and personal observation, we found most people take less than 30 minutes to complete the questions, so we used a cut-off threshold of 40 minutes. We eliminated 12 participants out of 865, and removed all of their timing data as outliers.

Across all nine primary conditions, the mean scores for the accuracy questions show most tasks took about a minute. The question about cookies took half as long and the question about social security numbers — which is not answered in the policies — took notably longer.

Figure 3.7: Overall mean time to answer questions in seconds. Vertical bars denote 95% confidence intervals around the point estimates.



3.3.2 Privacy Finder

Privacy Finder underperformed Natural Language on three of the six psychological acceptability questions, with lower mean Likert values on information retrieval questions. Participants responded less favorably when asked about easy of understanding, their confidence they understood, and how easy it was to find information. In all three cases the effect sizes were small, indicating that while there are significant differences they are not highly important. Participants gave higher ratings for Privacy Finder than the P3P Expandable Grid on five of the six psychological acceptability questions, with no significant difference in the mean values for how well protected participants felt by Privacy Finder or the P3P Expandable Grid formats.

Privacy Finder was indistinguishable from Natural Language for accuracy with one exception. Participants answered more accurately with Natural Language than Privacy Finder when asked if the privacy policy allowed the corporation to share their email address, with a medium effect size. Privacy Finder outperformed the P3P Expandable Grid for questions about cookie settings

and Social Security Number collection.

Privacy Finder was indistinguishable from Natural Language for task completion times and outperformed the P3P Expandable Grid on all six times.

3.3.3 The P3P Expandable Grid

We tested an early prototype of the P3P Expandable Grid. All other formats fared better. Subsequent lab studies helped identify interaction issues and the P3P Expandable Grid is being revised. The P3P Expandable Grid underperformed Natural Language for all psychological acceptability, with large effect sizes on all six of the questions indicating this is an important difference. As mentioned above, the P3P Expandable Grid underperformed Privacy Finder for five of six psychological acceptability questions.

Topic	Difference between...	<i>p</i> value
Feel secure	NL $\mu = 3.6$, EG $\mu = 2.8$	$p < .001$
Feel secure	PF $\mu = 3.3$, EG $\mu = 2.8$	$p = .003$
Protect more	NL $\mu = 3.4$, EG $\mu = 2.9$	$p < .001$
Pleasurable	NL $\mu = 3.1$, EG $\mu = 2.2$	$p < .001$
Pleasurable	PF $\mu = 2.9$, EG $\mu = 2.2$	$p < .001$
Explained	NL $\mu = 3.4$, EG $\mu = 2.9$	$p = .003$
Explained	PF $\mu = 3.3$, EG $\mu = 2.9$	$p = .006$
Confident understood	NL $\mu = 4.3$, PF $\mu = 3.8$	$p = .008$
Confident understood	NL $\mu = 4.3$, EG $\mu = 3.1$	$p < .001$
Confident understood	PF $\mu = 3.8$, EG $\mu = 3.1$	$p < .001$
Easier to understand	NL $\mu = 4.5$, PF $\mu = 4.1$	$p = .004$
Easier to understand	NL $\mu = 4.5$, EG $\mu = 3.0$	$p < .001$
Easier to understand	PF $\mu = 4.1$, EG $\mu = 3.0$	$p < .001$
Hard to find	NL $\mu = 4.1$, PF $\mu = 3.6$	$p = .002$
Hard to find	NL $\mu = 4.1$, EG $\mu = 2.5$	$p < .001$
Hard to find	PF $\mu = 3.6$, EG $\mu = 2.5$	$p < .001$

Table 3.2: Statistically significant differences between mean Likert scores on psychological acceptability, by policy format. Higher Likert scores are better.

The P3P Expandable Grid underperformed both Natural Language and Privacy Finder for accuracy questions about cookie settings and Social Security Number collection, and dramatically underperformed Natural Language for accuracy on how to remove oneself from a mailing list, with large effect sizes. We have since confirmed that the opt-out link for the P3P Expandable Grid was hard for participants to find.

We found no significance for the time for participants to answer if the privacy policy allowed

the corporation to share their email address. On the remaining five task completion times, the P3P Expandable Grid underperformed both Natural Language and Privacy Finder, with the exception of the question about cookies where the P3P Expandable Grid underperformed Natural Language and was indistinguishable from Privacy Finder. The question on encryption had a very large effect size; encryption was not mentioned in any of the policies to see if participants could correctly answer that the policy did not provide that information. It took participants longer to determine the absence of information with the P3P Expandable Grid.

Topic	Difference between...	<i>p</i> value
Use cookies	PF $\mu = 10$, EG $\mu = 17$	$p = .002$
Remove (opt out)	NL $\mu = 37$, EG $\mu = 59$	$p < .001$
Remove (opt out)	PF $\mu = 36$, EG $\mu = 59$	$p < .001$
Share email	NL $\mu = 29$, EG $\mu = 39$	$p = .043$
Share email	PF $\mu = 33$, EG $\mu = 39$	$p = .171$
SSN	NL $\mu = 20$, EG $\mu = 42$	$p < .001$
SSN	PF $\mu = 23$, EG $\mu = 42$	$p < .001$
Encryption	NL $\mu = 31$, EG $\mu = 88$	$p < .001$
Encryption	PF $\mu = 39$, EG $\mu = 88$	$p < .001$

Table 3.3: Statistically significant differences between mean completion times by policy format.

3.3.4 Layered

Layered was nearly indistinguishable from the long Natural Language policy. We found no statistically significant differences between the two for any of the seven psychological acceptability questions, and no significance for completion time data.

We found two statistically significant differences in means between the long Natural Language policy and the Layered policy out of the six accuracy scores, with a mixed result. Participants were more likely to answer the cookies question correctly with the long Natural Language policy (99% correct) than the Layered policy (89%; $p = .008$, large effect size of -0.9) but more likely to answer the question about Acme's email practices correctly from the Layered policy (83%) than the long Natural Language policy (65%; $p = .006$, medium effect size of 0.4).

3.3.5 Length

Our hypothesis was that shorter policies would have the highest levels of psychological acceptability. This was not supported. In all but two questions, there were no statistically significant differences between lengths. Furthermore, the longer policies scored slightly more favorably for the questions on protection (Short $\mu = 3.1$, Long $\mu = 3.2$, $p=.039$) and security (Short $\mu = 3.1$, Long $\mu = 3.3$, $p=.013$). A study on EULAs also found that people trust longer texts[58]. However, the effect sizes are small for these results, indicating the differences are not very important.

As expected, longer policies had longer response times.

Topic	Difference between...	p value
Use cookies	Short $\mu = 11$, Medium $\mu = 15$	$p = .028$
Remove (opt out)	Short $\mu = 38$, Long $\mu = 52$	$p = .005$
Remove (opt out)	Medium $\mu = 39$, Long $\mu = 52$	$p = .015$
Share email	Short $\mu = 27$, Long $\mu = 44$	$p = .032$
Encryption	Short $\mu = 39$, Long $\mu = 66$	$p = .009$

Table 3.4: Statistically significant differences between mean completion times by policy length.

3.4 Study Limitations and Future Work

The CUPS lab is currently working on three related studies:

- A follow-up study to contrast different Natural Language policies. This will allow us to confirm the Natural Language policy used in the present study was representative.
- A follow-up study and focus group with students who understand P3P. This will allow us to understand issues with the P3P Expandable Grid format from the perspective of more experienced users.
- A lab study with a think aloud protocol to determine where participants have trouble with the P3P Expandable Grids format. We also tested a new format that is a hybrid of Natural Language and the Privacy Finder format.

If successful, this work will result in ways to improve the readability of privacy policies. We would also be interested in mental model work to understand what currently works, what people

understand from privacy policies, and how they make trust decisions in the absence of reading privacy policies.

3.4.1 Representativeness of Policies

While we deliberately selected a policy with a high number of P3P statements in order to test the affect of differing lengths, in many other regards the original O'Reilly policy is typical. For example, it scores 12 on the Flesch-Kincaid Grade level, indicating a college level text. Note that Flesch-Kincaid is a right-censored score and terminates at 12: how much above college freshman level cannot be determined with this instrument. Many studies have confirmed that privacy policies require college reading level, including recent studies [59]. In this regard, the O'Reilly policy is typical.

Pollach's recent analysis of privacy policies examined five forms of textual ambiguity: downplaying frequency ("From time to time we may send you marketing email,") emphasizing qualities ("If you do not want the benefits of advertising cookies,") hedging claims ("We might use cookies,") obscuring causality ("When we have your permission,") and removing agents (essentially passive voice: "your data is shared.") With the exception of hedging claims, policies typically have one to two occurrences of these types of textual ambiguity [115]. The O'Reilly policy is typical in these four categories.

However, unlike most policies, the O'Reilly policy does not use any hedging claims. In contrast, typical policies have around 20 hedging claims [115]. That suggests the O'Reilly policy may be easier to read and understand than a typical policy in this regard.

The O'Reilly policy does have some areas that appear designed to soothe rather than inform readers. For example, the policy states "We use [analytics] to improve usability on our sites and to help support our customers online needs." Overall, our subjective judgment based on reading hundreds of policies is that the O'Reilly policy is a little easier to read than average, but does reflect a typical policy.

3.4.2 Length

Because P3P describes the structure and content of privacy policies, we were able to analyze each policy's P3P characteristics. At the highest level, P3P policies must have one or more STATE-

MENT elements. For a properly written P3P policy, the more STATEMENTS it has the more modular it is: multiple STATEMENTS indicate data is being collected and / or used in multiple contexts.

To create the medium length policy, we eliminated half of the eight data STATEMENTS in the P3P policy as well as the corresponding sections in the human readable policy. For example, the original policy has a section on entering contests; that section is omitted in the medium length policy. Note that the contest information is not something we asked questions about. We performed similar editing to pare down to the short policy, which had one data STATEMENT.

3.5 Observations

Well-written Natural Language policies supported accurate decisions better than other variants. There is a potential to improve other formats by borrowing from Natural Language. In particular, Natural Language formats may succeed in part because they provide context around unfamiliar concepts and jargon.

Length is by far less important than other factors. If anything, privacy experts' attempts to protect people from policies that appear long may cause more harm than good, since these attempts typically result in more complicated interfaces to the policy and require users to click, expand, scroll, and so forth to see the full details of the policy.

Overall, Layered policies and the Privacy Finder report format were fairly similar to Natural Language. We would neither suggest companies remove those formats from use nor would we suggest an aggressive push to adopt them. The P3P Expandable Grid format is currently undergoing revisions that should help address some of the difficulties participants had with them, and we look forward to seeing that format evolve.

Many researchers start from the observation that privacy policies are not usable in their current format, and suggest ways to fix the problem. These are laudable attempts. As this study shows, the status quo is difficult to improve upon. Privacy policies may need to be seen as having an educational component so readers understand what it means for a site to engage in a given practice. Further, it may help to take a step back and use mental model protocols to more clearly understand what does work in current policies to ensure new approaches are improvements and not merely different.

Financial privacy statements mandated by GLB are evolving. While they are narrowly tailored to the financial industry, lessons learned may be more widely applicable. We have started to test a new format for privacy policies that is similar to the recommendation in the Kleimann Report [119]. While this is preliminary work, about half the study participants indicated they prefer the new format to both Natural Language and the P3P Expandable Grid.

The Natural Language policy we used may be easier to read than other policies, and that may limit or change our conclusions. A follow-up study to contrast different Natural Language policies. This will allow us to confirm the Natural Language policy used in the present study was representative.

This study examined only one privacy policy. One strength of standardized formats is that they facilitate comparison between policies. Natural Language may not perform as well for consumers who contrast two policies side-by-side.

Our population was older, better educated, more affluent, and more experienced with the Internet than the population of United States Internet users[111],[150]. We would expect our population to be better able to read and understand privacy policies, plus have more financial risk from fraud and therefore greater privacy concerns. While this study does not paint an optimistic picture of consumers' ability to read and use privacy policies, a more representative sample could generate even worse results.

3.6 Conclusions

Conveying privacy practices is a hard problem. All formats were unsatisfactory. Participants universally disliked privacy policies of all types, and the highest mean score on the psychological acceptability questions was neutral. While participants were able to answer simple questions very well, with accuracy scores over 90% on several questions, they also struggled to find information and in one case barely beat chance (40% accuracy on a question with three possible answers).

There is a difference — and a tension — between trust and performance. Participants trust long, complicated policies slightly more yet have a little more trouble using and do not like long policies. The goals of public policy makers are presumably to support good decision making, and the goals of privacy policy authors are presumably to encourage customers to trust the site. These two goals may come into conflict.

Legislating standardized formats for online privacy policies, perhaps similar to GLB, appears premature. We did not find a compelling format that improves upon Natural Language. Legislation might be useful if we can devise improved formats. We suggest more research and new or revised standardized formats.

3.6.1 Accuracy

Participants answered a warmup question we did not score, followed by six questions in random order:

- Does the Acme website use cookies?
- Does this privacy policy allow Acme to put you on an email marketing list?
- How can you remove yourself from Acme's email list?
- Does this privacy policy allow Acme to share your email address with a marketing company that might put you on their email marketing list?
- Does the Acme website collect your Social Security number?
- If you send your credit card number to Acme do they keep it encrypted to prevent data theft?

3.6.2 Psychological Acceptability

After completing the initial information search tasks, participants answered a series of questions designed to elicit their reactions to the policy. We sought to understand the psychological acceptability of the varied policy formats.

- I feel secure about sharing my personal information with Acme after viewing their privacy practices
- I believe Acme will protect my personal information more than other companies
- Finding information in Acme's privacy policy was a pleasurable experience
- I feel that Acme's privacy practices are explained thoroughly in the privacy policy I read

- I feel confident in my understanding of what I read of Acme's privacy policy
- This privacy policy was easier to understand than most policies
- It was hard to find information in Acme's policy
- If all privacy policies looked just like this I would be more likely to read them

Participants responded on a 7 point scale to indicate their agreement or disagreement with the eight statements above. Note that question 7 above is reversed in analysis to keep a consistent scale.

Representativeness

Our population was older, better educated, more affluent, and more experienced with the Internet than the population of United States Internet users[111],[150]. We would expect our population to be better able to read and understand privacy policies, plus have more financial risk from fraud and therefore greater privacy concerns.

Chapter 4

A Comparative Study of Six Companies' Privacy Policies

This chapter is largely a reproduction of a paper co-authored with Robert W. Reeder, Patrick Kelley, and Lorrie Faith Cranor presented to the 2009 Privacy Enhancing Technologies Symposium (PETS) [99].

4.1 Introduction

Self-reports show three quarters of Internet users take active measures to protect their privacy, ranging from installing privacy protective technology to providing false information to web sites [6]. Yet only 26% read privacy policies during a recent study and readership outside of laboratory conditions is believed to be far lower [74].

To study the effectiveness of various approaches to improving the readability of privacy policies, we investigated the performance of three different formats for privacy policies and compared policies from six different companies.

4.2 Related Work

Several studies frame willingness to read privacy policies as an economic proposition and conclude that asymmetric information is one reason why people find it not worth their time to read privacy policies [153, 6]. Other studies show that privacy policies and financial disclosures require a college reading level to understand [65, 127, 59, 11]. A study of ambiguities in privacy policies shows they contain language that downplays privacy issues [115]. The 2006 Kleimann report on GLB financial privacy notices found that subheadings and standard formats dramatically improved readability [119]. In response to these issues, privacy researchers and industry groups devised several standardized formats for privacy policies based on the expectation that standardized formats would improve comprehension. Our study is a comparative analysis to analyze how well standardized policies work in practice.

While not in the realm of privacy policies, Kay and Terry's research on open source license agreements includes testing multiple formats. Early work found modest improvements in likelihood to read well designed agreements but no improvement in retention of the material [76]. Tsai found when study participants searched for products to purchase and saw a single icon view that evaluated the privacy practices for each site, they were willing to pay a small premium for more privacy-protective sites [146, 48]. On the other hand, translating an entire privacy policy into a grid that conveyed information by icons and colors did not improve comprehension [118]. Attempts at visualizing privacy are ongoing, including a set of icons modeled after Creative Commons [21]. This study, in contrast, examines three text-based formats as described below.

4.2.1 Privacy Finder

Privacy Finder (PF) is a privacy-enhanced front end to Yahoo! and Google search that was developed by AT&T and refined at the Cylab Usable Privacy and Security (CUPS) Laboratory. Privacy Finder includes a privacy report that displays standardized text generated automatically from Platform for Privacy Preferences (P3P) policies. P3P is a standardized format for privacy policies, and is formally recommended by the World Wide Web Consortium (W3C) [154]. P3P policies are encoded in XML (eXtended Markup Language), which is computer readable and thus allows software tools to help people manage their privacy preferences.

Because Privacy Finder generates text from P3P tags, the Privacy Finder report avoids emotionally charged language and ensures uniform presentation. However, Privacy Finder reports allow a free-form text description of the highest level of policy statements. This can improve readability by providing context for readers, but also means that companies with identical practices may have different Privacy Finder reports.

4.2.2 Layered Notices

The law firm Hunton & Williams popularized the notion of layered notices [137] which include a short one-screen overview with standardized headings which then links to the full natural language policy. Although the headings for the first layer are standardized the text within each section is free form.

By 2005, several large companies deployed layered policies including Microsoft (MSN), Procter & Gamble, IBM, and JP Morgan [86]. European Union Information Commissioner Richard Thomas called for the use of layered policies in response to research showing nearly 75% of participants said they would read privacy policies if they were better designed [109]. Article 29 of European Union Directive created the “Working Party on the Protection of Individuals with regard to the processing of Personal Data,” which issued guidance on how to create layered policies [29]. Privacy commissioners in EU countries supported layered policies. In Australia, the Privacy Commissioner released a layered policy for their own office, intending it “as a model for other agencies and organisations” [138].

4.2.3 Natural language

Most privacy policies are in natural language format: companies explain their practices in prose. One noted disadvantage to current natural language policies is that companies can choose which information to present, which does not necessarily solve the problem of information asymmetry between companies and consumers. Further, companies use what have been termed “weasel words” — legalistic, ambiguous, or slanted phrases — to describe their practices [115]. Natural language policies are often long and require college-level reading skills. Furthermore, there are no standards for which information is disclosed, no standard place to find particular information, and data practices are not described using consistent language.

4.3 Methods

We conducted an online study from August to December 2008 in which we presented a privacy policy to participants and asked them to answer questions about it. We posted advertisements on craigslist and used personal networks to recruit participants. We offered a lottery for a chance to win one of several \$75 Amazon gift certificates as incentive for participating in the study.

We used a between subjects design and assigned each participant to one of 15 privacy policy representations. We used a between subjects design rather than within group design because in this context it is unrealistic to eliminate learning effects simply by reordering policies. Reading the questions could affect how participants read subsequent policies. It is also unrealistic to expect participants to spend more than 20 minutes completing an online survey. Questions remained constant over all conditions; only the policy differed.

4.3.1 Study Conditions

We contrasted six different companies’ conventional natural language (NL) policies and their corresponding Privacy Finder privacy report format (PF) plus three layered policies. We refer to these companies as A through F. We analyzed 749 participants across 15 conditions, for an average of 50 participants per condition. Note that we did not study layered policies for companies A, C, and E. The study conditions are listed in Table 4.1.

We replaced all companies’ names with “Acme” to avoid bias from brand effects. For natural

Table 4.1: Participants per Condition

Company	Designation	NL	PF	Layered
Disney	A	41	50	N/A
Microsoft	B	47	46	52
Nextag	C	46	41	N/A
IBM	D	47	47	49
Walmart	E	52	51	N/A
O'Reilly	F	62	55	63

language policies we used black text on white backgrounds regardless of the original graphic design. We left other formatting that might aide comprehension (for example, bulleted lists) intact.

Note that we did not study layered policies for companies A, C, and E. Of the six companies, only B and D had layered policies. We followed the directions from the Center for Information Policy Leadership [32] to create a third layered policy for company F as part of a prior study [118] and used it here to facilitate comparisons between studies.

As deployed in practice, Privacy Finder highlights the most important information at the top of the report and provides links to expand details. We discovered in earlier testing that people rarely expanded the Privacy Finder report. We were interested in testing how well people are able to use the information in the Privacy Finder report, not how well they are able to navigate the user interface so in our research we presented all information in a single flat file.

We selected privacy policies from six popular websites that engage in e-commerce, and thus must collect a variety of personal information as part of their business. We chose what we believe to be a comparatively easy to read and a comparatively difficult to read policy with several typical policies.

Examples from two policies follow. In the policy from company F, the text is fairly clear and readable. Sentence structure is straight-forward. The bulleted list makes it easier to identify important points:

We share customer information only with affiliated companies and as described below:

- With third parties we retain to perform functions on our behalf, such as fulfilling orders, processing credit card payments, managing mailing lists, and delivering

packages. These parties are restricted from using your information for any other purpose.

- We rent our snailmail list for one-time use to third parties we deem relevant and appropriate. We do not rent or sell our email lists.
- We release personal information when we believe that release is appropriate to comply with the law, or to protect the rights, property, or safety of Acme Incorporated, our users, or others. This may include exchanging information with other companies and organizations for fraud protection and credit risk reduction.

In contrast, the policy from company A is legalistic. Familiar words are redefined to mean something slightly different. Cross references to other sections in the policy require either a good memory or flipping back and forth between sections to understand precisely what is going on. Sentences are long and convoluted:

Subject to your opt-out choices (see A4 below), The Acme Family of Companies may share your personal information with selected third parties so that they can send you promotional materials about goods and services (including special offers and promotions) offered by them. (We call this type of sharing “promotional sharing” or sharing for “promotional purposes.”) When sharing your information for promotional purposes, The Acme Family of Companies attempts to select only reputable companies that offer high quality products and services. Moreover, The Acme Family of Companies will not share your e-mail address with third parties for promotional purposes, except when you consent to such sharing in the course of your participation in a Sponsored Activity as described below. When The Acme Family of Companies shares your personal information with a third party under any circumstance described in this “Third Parties Offering Promotions, Products, or Services” section, your personal information will become permanently subject to the information use and sharing practices of the third party, and the third party will not be restricted by this Privacy Policy with respect to its use and further sharing of your personal information.

We selected privacy policies from six popular websites that engage in e-commerce, and thus must collect a variety of personal information as part of their business. We chose what we believe

to be a comparatively easy to read and a comparatively difficult to read policy with several typical policies. We selected policies guided by several measurements of readability summarized in Table 4.2. For each company, we noted the length of the natural language policy. We calculated the Flesch-Kincaid Reading Ease Score, which ranges from a low of 1 to a high of 100 based on syllable count and line lengths. High Flesch-Kincaid scores are more readable than low scores. In general, experts suggest a score of at least 60—70, which is considered easily understandable by 8th and 9th graders [103]. *Reader's Digest* has a readability index in the mid 60s, *Time* is in the low 50s, and *Harvard Law Review* in the low 30s [69]. Note that while the policies we selected span a range from 32 to 46, even the most readable policy is more challenging than is normally recommended for a general audience.

We calculated the percentage of sentences written in the passive voice, which is both more difficult for readers to understand and an indicator the company may not be comfortable taking full responsibility for their privacy practices. We counted the number of cross references within each policy; the more times readers are asked to refer to other parts of the document the more difficult it is to understand. Finally, we note that the standardized Privacy Finder format also has a range of lengths due to differing numbers of statements, how much information they collect, and how much text the policy authors elected to supply.

Table 4.2: Attributes of six companies' privacy policies

Co.	NL Words	NL Pages	Flesch	% Passive	Cross ref.s	PF Words
A	6329	13	31.8	11%	27	880
B	3725	7	35.5	22%	0	1964
C	2920	6	36.3	17%	7	2011
D	2586	8	42.8	18%	2	554
E	2550	8	44.9	11%	0	1373
F	928	3	46.3	9%	1	1843

4.3.2 Study Questions

Study questions comprised several groups:

- *Comprehension.* Participants answered a series of multiple choice questions to determine how well they were able to understand the policy. These questions are realistic information

retrieval tasks based on typical privacy concerns, and are similar to questions used in an earlier study by Cranor et al [41]. We conducted three rounds of pilot tests with over two dozen people to ensure the questions were well-worded and understandable. We randomized the order of these questions to mitigate learning effects and captured both accuracy and time to respond. We also included a warm-up task which we did not score.

- *Psychological Acceptability.* Saltzer and Schroeder coined the term psychological acceptability to convey that if people do not like a system they will not use it. They wrote, “It is essential that the human interface be designed for ease of use, so that users routinely and automatically apply the protection mechanisms correctly.” [124] Participants answered subjective questions on a seven-point Likert scale.
- *Demographics.* We collected basic information like gender, educational attainment, and income so we could understand how closely our study population resembles Internet users as a whole.

We also measured the time it took for participants to answer each one of the comprehension questions. When not engaged in a research study, few people even skim privacy policies let alone read them to find answers to their concerns [76]. The times we measured do not reflect normal practices, but they do allow us to compare performance between formats, which is our goal.

4.3.3 Research Questions

Standardized formats were designed with care to help readers make sense of online privacy policies. With all of the resources invested in standardized policies we expected they would help people understand privacy policies. We held multiple hypotheses:

- Participants will have (a) higher accuracy scores, (b) shorter times to answer, and (c) greater psychological acceptability with both of the standardized formats than with their natural language counterparts.
- Participants will have (a) higher accuracy scores, (b) shorter times to answer, and (c) greater psychological acceptability with highly readable natural language than they will on natural language policies with low readability metrics.

Understanding these issues contributes to determining the most effective ways to present policies to end users. This is particularly relevant given Gramm-Leach-Bliley regulations on paper-based financial privacy policies; similar legislation could apply to online privacy policies in the future. The FTC’s most recent report on behavioral advertising was described by the FTC Chairman Leibowitz as the last chance to make industry self-regulation work [53]. If we move away from industry self-regulated content, what should we do instead? Do any of the standardized approaches help enough to warrant considering regulation of policy formats?

4.3.4 Analysis

We performed a comparative analysis across all three formats (Natural Language, Privacy Finder, and Layered) and from all six companies to see if there were statistically significant differences in the mean scores for accuracy, time to completion, and psychological acceptability questions.

After we removed outliers¹ we performed ANOVA analysis for both time data and psychological acceptability, which we recorded on a seven point Likert scale and treated as continuous variables. Accuracy questions were categorical data (either accurate or false) so we used Chi Squared tests. We performed all tests of statistical significance at the $\alpha = 95\%$ confidence level. For the sake of readability, all details of statistical significance tests are in Appendix A.

4.4 Accuracy and Speed Results

Accuracy scores are all reported as the percentage of people who answered the question correctly.²

As compared to natural language, we found that layered policies led to lower accuracy scores for

¹We only included results from participants who completed all of the accuracy questions. Because this was an online study to enter a drawing for a gift certificate, a few people just “clicked through” answers without engaging with the material. We picked a fixed lower threshold of 1.5 seconds per question and removed participants entirely if they had two or more questions they answered in under 1.5 seconds (7 participants removed out of an original 756 for a total of 749.) For participants with only one time under 1.5 seconds, it is possible they accidentally double-clicked once but answered other questions properly. We removed the time and accuracy data for just the affected question (3 question/time pairs out of 3000.) At the other extreme, sometimes people were diverted by other tasks while answering questions and we recorded unduly long times to answer. We discarded question times in excess of 2.5 times the mean for their condition along with their corresponding answers. This resulted in $N = 723$ for cookies, 728 for opt out, 726 for share email, and 723 for the telemarketing questions.

²Interpreting results is complicated by potential confusion of how participants answered when answers are inferred. For example, we asked about opt out practices for policies where there is no opt out link. The straight-forward answer we envisioned is “No.” However, participants may also have replied that the policy “Does Not Say,” intending to convey the same information since there is no opt out link within the policy. Arguably, in that case the correct way to score responses is to combine the correct answer with “Does Not Say.” We analyzed the combined percentage for each question and found in all but one case there was no difference in the threshold for statistical significance. Further, the relative ranking of formats and companies remained stable.

topics not in the short layer. Privacy Finder was indistinguishable from natural language until questions became harder, at which point Privacy Finder was slightly superior to natural language.

Accuracy spanned a wide range. An average of 91% of participants answered correctly when asked about cookies, 61% answered correctly about opt out links, 60% understood when their email address would be “shared” with a third party, and only 46% answered correctly regarding telemarketing. With only three possible answers, if participants guessed randomly we would expect 33% accuracy.

All other things being equal, lower times are better because they reflect participants were better able to comprehend the policy. Participants answered more quickly with both layered and Privacy Finder formats. Times to answer increased with question difficulty, with an average of 2.3 minutes to answer the question about cookies, 4.7 minutes to answer about opt out links, 5.3 minutes for email sharing, and 6.7 minutes for telemarketing.

Table 4.3: Percentage correct and minutes to answer, cookies question.

Policy	% correct	Time
A NL	87%	3.6
A PF	96%	1.5
B NL	96%	2.0
B PF	98%	1.6
B Layered	86%	2.3
C NL	93%	2.4
C PF	98%	3.5
D NL	86%	2.6
D PF	91%	1.9
D Layered	69%	2.2
E NL	96%	2.6
E PF	96%	1.8
F NL	100%	2.3
F PF	94%	2.7
F Layered	80%	2.3

4.4.1 Cookies

We asked: Does the Acme website use cookies?

Answer: Yes for all policies.

Most participants got the cookie question right (91%). This was an easy question to answer because our question is phrased with the same term the policies use. All policies, in all formats, call out cookies use explicitly. For example, one policy has a heading of “Cookies and Other Computer Information” with a paragraph that begins: “When you visit Acme.com, you will be assigned a permanent ‘cookie’ (a small text file) to be stored on your computer’s hard drive.” There is no ambiguity. Even someone who has no idea what a cookie is, or what the implications for privacy are, can skim through any of the natural language policies to find the word “cookie” and answer correctly.

We found significant differences in accuracy for company and format. The six companies have a relatively small span between the worst performance (D, 82%) and best performance (E, 96%.) See Table 4.3 for a summary of results.

Layered policies gave participants a little more trouble (78%) than other formats. Cookie information was under the heading “Personal Information” in F Layered (80%,) which may not be where people expected to look. In D Layered (69%,) the policy mentions in passing that “You may also turn off cookies in your browser,” without explicitly saying they use cookies. People must deduce that information or go to the full policy for a direct statement that the site uses cookies. This highlights two results we will see again: first, when participants needed to think about an answer rather than just perform a search for information, accuracy dropped. Second, it appears few people ventured beyond the first page of the layered policies. Kay and Terry found similar issues with layered policies [76].

In another sign that this was an easy question for most participants, times to answer were shorter than the other questions (2.3 minutes.) We found no significance for time based on company but format was significant. Privacy Finder (2.1 minutes) and Layered (2.3 minutes) supported faster responses than Natural Language, but the Layered condition was also had more in incorrect answers.

Table 4.4: Percentage correct and minutes to answer for the opt out question.

Policy	% correct	Time
A NL	33%	5.7
A PF	85%	3.7
B NL	33%	9.3
B PF	91%	4.6
B Layered	18%	4.8
C NL	80%	3.2
C PF	73%	5.1
D NL	29%	6.1
D PF	71%	3.8
D Layered	19%	5.5
E NL	55%	5.4
E PF	51%	4.6
F NL	93%	3.4
F PF	79%	3.7
F Layered	92%	2.2

4.4.2 Opt Out Link

We asked: Does the company provide a link to a webform that allows you to remove yourself from Acme’s email marketing list?

Answer: Yes for all policies except: B NL, D NL, D Layered, E NL, which are No.³

This question is a little more difficult than the question about cookies. Policies refer to this concept as “opting out.” For example, company C’s natural language policy phrases it as “To opt out of receiving all other Acme mailings after you have registered, click here or click the appropriate unsubscribe link contained within the email that you receive.” Participants need to map the concept of removing themselves from an email marketing list to the technical jargon of opting out. However, this question is again fairly straight forward. Either there is an opt out link or there is not. See Table 4.4 for a summary of results.

We found significant differences for company and format. Natural language policy accuracy rates are dissimilar, with averages ranging from 93% (F) to 33% (A). Finding the opt out link in the A NL policy was looking for a needle in a haystack: there is one link halfway through the policy in the middle of a paragraph without any headings or other cues—and the policy runs to 13 pages when printed.

It would seem Privacy Finder should have consistent results across all six policies, since an opt out link is a standard part of Privacy Finder reports. However, companies with an opt out default have additional links for each category of opt out data. As a result, policies with opt out practices fared better, ranging from 85% correct (A PF) with less privacy protective practices and many prominent opt out links, to 51% correct (E PF) which required opt out for all data collection and had only one opt out link. Interestingly, the F PF policy (79%) has identical practices as E PF (51%) yet different accuracy scores. The author of the F PF policy included an additional opt out link in the text at the very end of the policy, which is prime real estate for readers’ attention. Policy authors choices affect outcomes, even within the PF standardized presentation.

Since there is no requirement to discuss opt out choices within the layered format, once again we see dissimilar results across a standardized format. B layered policy (18%) required clicking the opt out link to see what it did, phrased as “For more information about our privacy practices, go to the full Acme Online Privacy Statement. Or use our Web form,” with a link from “Web form” to the opt out page. In contrast, results were quite good with F layered (92%), which contained the same opt out text as at the end of the F PF (79%) policy.

We found significant differences in time to answer for company as well as format. We would

³Answers are not the same across a given company because the companies elected to provide different information in different formats. P3P requires an opt out link, which is then included in Privacy Finder.

expect longer times for longer policies since this is in many ways an information search task. Instead, time appears to be based on the underlying practices: policies without opt out links took longer. Since some of the policies with opt out links mentioned them at the end, it is unlikely the difference in times is based on reading through the entire policy to determine the absence of a link. Instead, participants likely re-read to satisfy themselves that they had not missed anything. Once again participants completed the task more quickly with layered (4.0 minutes) and Privacy Finder (4.2 minutes) than Natural Language (5.4 minutes,) but the wide variance and sometimes poor performance for standardized policies reduces the strength of this result.

We asked: Does this privacy policy allow Acme to share your email address with a company that might put you on their email marketing list (with or without your consent)?

Answer Yes for all policies except: companies E and F (all formats) which are No.

4.4.3 Share Email

We tested the wording of this question in multiple pilot studies to ensure people understood it without asking something pejorative or jargon-laden like “will Acme sell your email address to spammers.” This question requires participants to understand the question, read the policy carefully, and make inferences for most policies. For example, C NL reads: “We may provide your contact information and other personal data to trusted third parties to provide information on products and services that may be of interest to you.” Participants need to understand that “contact information” includes email, that “trusted third parties” are companies other than Acme, and that “provide information on products and services” means marketing messages, in order to correctly answer “Yes.” See Table 4.5 for a

Table 4.5: Percentage correct and minutes to answer for the email sharing question.

Policy	% correct	Time
A NL	76%	3.2
A PF	53%	5.4
B NL	49%	5.9
B PF	64%	5.9
B Layered	52%	4.8
C NL	80%	4.7
C PF	72%	6.9
D NL	67%	4.6
D PF	78%	4.0
D Layered	56%	4.7
E NL	53%	6.9
E PF	44%	6.2
F NL	50%	6.0
F PF	54%	4.4
F Layered	62%	5.0

summary of results.

Overall accuracy was only 60%. We found significant differences for company but not format. Times to answer averaged 5.3 minutes, which indicates people had a harder time completing this task. We found no significant results for time based on company or format.

As the answers to our questions become more nuanced we would expect the more readable policies to shine, yet that is not the case. Company A, with the hardest to read policy, had a higher accuracy score (64%) than F (55%) with the most readable policy and there was no overall discernible pattern based on readability. Similarly, we would expect standardized policies to convey information better, especially the Privacy Finder format which avoids the emotion-rich wording of “trusted third parties” and “valuable offers,” yet we did not find significant differences between formats. Privacy Finder summarizes “With whom this site may share your information” as “Companies that have privacy policies similar to this site’s” which again requires participants to refer to a separate section to determine if the parent company may engage in email marketing.

Table 4.6: Percentage correct and minutes to answer for the telemarketing question.

Policy	% correct	Time
A NL	23%	8.7
A PF	43%	5.9
B NL	41%	6.7
B PF	67%	5.9
B Layered	16%	6.2
C NL	42%	9.2
C PF	68%	5.5
D NL	42%	7.6
D PF	82%	3.2
D Layered	33%	5.5
E NL	65%	10.2
E PF	56%	5.4
F NL	26%	7.1
F PF	55%	7.4
F Layered	34%	5.9

4.4.4 Telemarketing

We asked: Does this privacy policy allow Acme to use your phone number for telemarketing?

Answer Yes for all policies except companies A, E and F (all formats) which are No.

Participants struggled with this question as shown in Table 4.6. Except in the Privacy Finder version where companies are required to provide information about their telemarketing practices, policies typically do not highlight telemarketing practices. The way to answer this question correctly was typically to read through the entire policy for all mentions of when the company

collects phone numbers, then see what policies they have around that data. For example, B NL discloses telemarketing as: “You may also have the option of proactively making choices about the receipt of promotional e-mail, telephone calls, and postal mail from particular Acme sites or services.” Sometimes policies were even more vague, for example D NL, “The information you provide to Acme on certain Acme Web sites may also be used by Acme and selected third parties for marketing purposes. Before we use it, however, we will offer you the opportunity to choose whether or not to have your information used in this way.” Not only is telemarketing swept under the phrase “marketing purposes,” telephone numbers are not mentioned explicitly either. It was necessary to deduce practices from a very careful and nuanced reading, frequently referring to multiple sections of the policy and then putting pieces together like a jigsaw puzzle. One could even make the case that answering “The policy does not say” is correct in cases as above where “information you provide” may be used for “marketing purposes” is by no means an explicit statement about telemarketing. However, we think it is important to note that the company likely does believe they have conveyed their practices: privacy policies are vetted by lawyers and are generally expected to be able to withstand a court or FTC challenge. If necessary, companies can point to the language in their policy and show that they did not violate the text by telemarketing.

We found significant differences in accuracy scores for company and format.⁴ We found no significant results for time based on company but format does have significant differences. Once again layered (5.7 minutes) and Privacy Finder (5.5 minutes) are an improvement over natural language (8.2 minutes) but with the caveat that layered does not do as well for accuracy.

Even though we called out D NL as particularly indirect, it falls solidly in the middle of the accuracy scores (42%.) When participants cannot find information in layered policies, by design they should continue to the full policy for more details. In practice this appears not to happen, with a very low accuracy of 28%.

Privacy Finder does support more accurate answers (61%) even in contrast to natural language (39%.) Privacy Finder is the only format that requires a company to disclose, yes or no, if they telemarket. For example, under the heading “The ways your information may be used” D PF includes “To contact you by telephone to market services or products – unless you opt-out.”

⁴Accuracy scores for telemarketing are the single exception where including “Does Not Say” as a correct answer changes whether we find significance between formats.

Again there is a lot of variation between Privacy Finder policies based on the supplemental text they provide. For example B PF, is particularly confusing by stating in free form text “While Acme does not currently support telemarketing, it is possible that in the future Acme properties may contact you by voice telephone,” directly above an automatically generated statement that they may use information for telemarketing.

4.5 Psychological Acceptability Results

After completing the initial accuracy questions, participants answered a series of questions designed to elicit their emotional reactions. Participants responded on a scale from 1 = strongly disagree to 7 = strongly agree. Most answers hovered right around 4, which is a neutral reaction. Higher numbers are always better.

4.5.1 Ease of Finding Information

We asked four questions about how easy it was to find information. We expected responses to these questions to reflect how well participants were able to understand a particular policy, and thus be related to the accuracy questions and times. However, we found few significant results. Participants found layered easier to understand even though they were less accurate with the layered format.

- “I feel that Acme’s privacy practices are explained thoroughly in the privacy policy I read” (M = 4.7, s.d. = 1.5.) We found significant effects for company but not format. A, B, and F (M = 4.8 for all) scored better than C, D, and E (M=4.4 for C and D; M=4.5 for E.)
- “I feel confident in my understanding of what I read of Acme’s privacy policy” (M = 4.7, s.d. = 1.6.) We found no significant differences between companies or formats.
- “This privacy policy was easier to understand than most policies” (M = 4.5, s.d. = 1.5.) We found no significant differences between companies but did find significant results for formats. Layered (M=4.8) scored better than natural language (M=4.4) or Privacy Finder (M=4.4.)

- “It was hard to find information in Acme’s policy” ($M = 3.8$, $s.d. = 1.6$.) We found no significant differences between companies or formats. (Note that based on the wording for this question we had to report the inverse of responses to keep higher numbers as better.)

4.5.2 Trust

If a format conveys information well but results in lack of trust of the company, it is unlikely that corporations will adopt the format. Participants trusted Privacy Finder formats slightly more than other formats.

- “I feel secure about sharing my personal information with Acme after viewing their privacy practices” ($M = 4.0$, $s.d. = 1.7$.) We found significant effects for both company and format.
- “I believe Acme will protect my personal information more than other companies” ($M = 4.0$, $s.d. = 1.6$.) We found significant effects for both company and format.

4.5.3 Enjoyment

We asked two questions to gauge how much participants liked reading the privacy policy. If people are unwilling to read policies then improving them does not provide much benefit. We found no significant differences between formats.

- “Finding information in Acme’s privacy policy was a pleasurable experience” ($M = 3.7$, $s.d. = 1.7$.) We found no significant differences between companies or formats. This was the lowest score of all eight psychological acceptability questions.
- “If all privacy policies looked just like this I would be more likely to read them” ($M = 4.2$, $s.d. = 1.7$.) We found significant effects for format but not company.

4.6 Demographics

We were interested in adult United States Internet users. Women were over-represented and comprised 64% of our study population. Minorities and household income were in keeping with the overall Internet population. See Table 4.7 for contrasts between the study population and

estimates of the Internet population. All estimates are from Pew [111] except for self-reported computer skill level, which is based on the 2005 AOL/NCSA study.

Table 4.7: Examples of skew for demographics

	Our population	Internet population
Age 18-29	49%	25% of over 18
Age over 50	10%	20% of over 18
No education beyond high school	10%	36%
Completed college	55%	39%
Self-reported computer novice	2%	30%
Self-reported computer expert	22%	8%

Overall, our sample was skewed toward slightly younger and better educated participants who are more confident in their computer skills than the overall population. That suggests our participants likely had a better than typical understanding of technical jargon and higher reading comprehension. Our results may be optimistic: a non-biased sample might have even more difficulty understanding policies.

4.7 Subsequent Work

Based in part on these findings, Kelley et. al. developed a new format for privacy policies [80]. They took a “nutrition label approach” to show the types of information a site collects, how that information is used, and with whom they share that information. They use a standardized format based on P3P, similar to the Expandable Grids idea. They present the most relevant information on one page with the human readable policy available for more detail, similar to the layered approach. Kelley et. al. tested this format and found it is an improvement upon prior attempts.

4.8 Discussion

Our hypotheses were not fully supported and in some cases were refuted. Both layered and Privacy Finder formats did improve times to answer, but not by much, and at the expense of accuracy for layered policies. Privacy Finder policies showed modest improvement in accuracy for complex questions but no improvement for easy questions. While the accuracy scores for Privacy

Finder were low in some cases, the format does represent a step forward from the status quo. Readability did not determine outcomes for natural language policies. For natural language, in some cases it appears the practices of the company were greater determinants than the words they used to describe those practices. We found few statistically significant differences in psychological acceptability.

Many researchers start from the observation that privacy policies are not usable in their current format and suggest ways to fix the problem. All of the formats were tested were unsatisfactory with a low rate of comprehension on questions that required synthesis of information. Participants did not like privacy policies of any type, and the highest mean score on the psychological acceptability questions was barely above neutral.

Privacy researchers tend to talk about policies as being uniformly bad. We expected that more readable natural language policies would have higher accuracy scores, lower times, and improved psychological acceptability than less readable policies, but that was not the case. These results could suggest that readability metrics are not a good way to differentiate between policies. This seems unlikely because the Flesch index has proven robust in many contexts and we do not immediately see any reason why privacy policies should be dramatically different from other types of textual analysis. It seems more likely that the range from 32 to 46 on the Flesch index is too similar to see major variations in outcome: even the most readable policies are too difficult for most people to understand and even the best policies are confusing.

Our results are robust across a variety of different policies, but our study does not concretely identify what makes a given policy comprehensible. However, we can offer three observations. First, results from the layered format suggest participants did not continue to the full policy when the information they sought was not available on the short notice. Unless it is possible to identify all of the topics users care about and summarize to one page, the layered notice effectively hides information and reduces transparency. Second, participants struggled to map concepts in the questions to the terms used in policies. It may prove fruitful to research how people internally represent privacy concepts: which terms do they currently use and which industry terms do they understand? As suggested in the Kleimann report for printed financial statements, online privacy policies may need an educational component so readers understand what it means for a site to engage in a given practice [119]. Third, the standardized formats we studied still offer policy

authors quite a bit of leeway. Companies with identical practices conveyed different information, and these differences were reflected in participants' ability to understand the policies. The flexibility of the standardized formats may undermine their expected benefits to consumers.

Our study used a between subjects rather than within subjects structure. We expect that we would see larger differences, particularly in psychological acceptability, if we were to place policies side-by-side. Prior work[41] found that when participants have both the natural language and the Privacy Finder versions available, Privacy Finder fares well. If people are reading multiple companies' policies to compare them, Privacy Finder may be advantageous. However, for just understanding a single policy, we find differences between formats are not as pronounced. By only showing one policy, our study did not capture one of the potential advantages to standardized formats. Standardized formats should be more useful once readers understand where to find information. Learning effects may play a role over time when people can take greater advantage of standardized formats as they become more familiar with their layout.

At this time, we do not recommend regulating the format of online privacy policies. While we did not find substantial benefit from the standardized formats we tested, that is not an indictment of the concept of standardized formats. Early results testing a new format for privacy policies based around a nutrition label concept are encouraging [80]. Ideally, future formats will identify problems with existing approaches and attempt to improve upon what has come before. In the future, we encourage rigorous testing for new formats before their supporters encourage widespread adoption.

Part II

Targeting and Behavioral Advertising

Chapter 5

An Empirical Study of How People Perceive Online Behavioral Advertising

This chapter is largely a reproduction of a paper co-authored with Lorrie Faith Cranor to appear at the 2010 Research Conference on Communication, Information and Internet Policy (TPRC) [98], which substantially expands upon a paper co-authored with Lorrie Faith Cranor to appear at the 2010 Workshop on Privacy in the Electronic Society (WPES) [97].

5.1 Introduction

This chapter presents empirical data on American adult Internet users' knowledge about and perceptions of Internet advertising techniques. We present the results of in-depth interviews and an online survey focusing on participants' views of online advertising and their ability to make decisions about privacy tradeoffs. We find users hold misconceptions about the purpose of cookies and the effects of clearing them, which limits cookie management as a self-help mechanism enabling user choice. Only 11% of respondents understood the text description of NAI opt-out cookies, which are a self-help mechanism that enables user choice. 86% believe ads are tailored to websites they have visited in the past, but only 39% believe there are currently ads based on email content, and only 9% think it is ok to see ads based on email content as long as their email service is free. About 20% of participants want the benefits of targeted advertising, but 64% find the idea invasive, and we see signs of a possible chilling effect with 40% self-reporting they would change their online behavior if advertisers were collecting data. We find a gap between people's willingness to pay to protect their privacy and their willingness to accept discounts in exchange for private information. 69% believe privacy is a right and 61% think it is "extortion" to pay to keep their data private. Only 11% say they would pay to avoid ads. With the exception of contextual advertisements, we find most participants would prefer random ads to tailored ads, but approximately 20% of participants would rather tailored ads. We find participants are comfortable with the idea that advertising supports free online content, but they do not believe their data are part of that exchange. We conclude with observations for public policy, technologists, and education.

Real-time mass media was born with national radio networks in the 1920s. As mass media gave rise to mass advertising, advertisers' campaigns became national. However, typically only a subset of people are interested in any given product or service advertised. As the old advertisers' lament has it, "We know we're wasting half our ad dollars, we just don't know which half" [44]. Online advertising can be targeted to users most likely to be interested in a particular product or service. Customers may benefit from ads targeted to their personal interests, reducing irrelevant ads and the time it takes to find products.

Behavioral advertising, which is one form of targeted advertising, is the practice of collecting data about an individual's online activities for use in selecting which advertisement to display. Behavioral advertising creates profiles for Internet users based on a variety of different data types

and inferences drawn from those data. Third-party cookies are one of several mechanisms used to enable behavioral advertising: a central advertising network with ads across thousands of websites can set and read cookies, noting every time a given user visits any of the sites in the network. By correlating which sites an individual visits, ads clicked, inferences about age range and sex, and approximate physical location based on the computer's IP address, advertisers build profiles of that individual's characteristics and likely interests. Profiles indicate if a given user is a good target for certain ads, with interest categories like "cars" or "Hawaiian travel." Google and Yahoo! both use behavioral advertising and made their interest categories public at the end of 2009.

The Internet is a form of mass media with targeted advertisements dependent on massive data collection on a tremendous scale. The Yahoo! ad server reaches over half a billion unique people each month, with 9.7% of the market [14]. Google's DoubleClick and AdSense ad servers have a combined total of 56% of the market and reach at least 1.5 billion unique users each month [14]. Google web beacons are on 88% of nearly 400,000 sampled websites and 92 of the top 100 most popular sites [57]. Google is reported to track approximately 90% of global Internet users [34]. The collection, storage, and use of the data that drives advertising has tremendous potential for privacy harm, as illustrated in the release of AOL search terms [17] and social networking information exposed by Gmail users trying Google Buzz [102]. There are four public policy domains that can benefit from understanding user perceptions of Internet advertising:

1. **Legislation.** State and Federal legislatures are considering new regulations around Internet privacy, including proposals from Representatives Boucher, Stearns, Rush, and Senator Kerry. Understanding what constituents know can help define legislative priorities: in areas where people are already able to protect their privacy interests, there is reduced justification for new laws.
2. **Industry self-regulation.** The Federal Trade Commission (FTC) and industry groups continue their efforts to improve corporate privacy practices without the burdens of regulation. Self-regulation presumes Internet users can make decisions to enact their privacy preferences, which makes understanding preferences, knowledge, and behavior a valuable contribution to evaluating self-regulation.
3. **Consumer expectations.** The FTC and privacy professionals within companies increasingly

look to issues of surprise to decide which practices are acceptable [16]. With users' subjective responses to privacy loss being used as guidance, rather than formal approaches like privacy rights frameworks, it is crucial to know how users react to current online practices.

4. Education. The first "principle" in the *Self-Regulatory Program for Online Behavioral Advertising* is education [2]. Establishing a baseline of user knowledge before campaigns begin will help establish their successes and any short comings.

In this chapter we review related work in section 5.2 and describe our methods in section . We present our findings regarding using cookie management as a self-help mechanism, participants' views of tailored advertising, and their willingness to pay for privacy in sections , , and respectively. We conclude in section .

5.2 Background and Related Work

Targeted advertising has received a lot of scrutiny in the past few years. There are questions about consumer's online privacy, how easily seemingly anonymous information can be re-identified [107], and the legality of some behavioral advertising business practices. The advertising industry favors continuing an "industry self-regulation" approach. The Federal Trade Commission has held workshops and released guidelines for self-regulation [55, 53], and there are legislative proposals at the Federal [25] and State [13] level, including proposals from Representatives Boucher, Stearns, Rush, and Senator Kerry.

In 2008, TRUSTe commissioned a report on behavioral advertising, finding 57% of respondents are "not comfortable" with browsing history-based behavioral advertising, "even when that information cannot be tied to their names or any other personal information" [143]. In 2009, TRUSTe found that even if it "cannot be tied to my name or other personal information," only 28% of Internet users would feel comfortable with advertisers using web browsing history, and 35% believe their privacy has been invaded in the past year due to information on the Internet [145]. Anton, et. al., performed some of the earliest work on behavioral advertising in 2002, with a follow up study in 2009 [12]. They found the types of privacy concerns remained stable, but the level of concern has increased around information used for behavioral advertising. Gomez et al. estimated that Google Analytics tracks at least 329,330 unique domains, and found confusion

in privacy policies containing “conflicting statements that third-party sharing is not allowed but third-party tracking and affiliate sharing are” [57]. Turow et. al. conducted a nationally representative phone survey in 2009. They found 66% of adults do not want tailored advertising, which increased to as high as 86% when participants were informed of three common techniques used in advertising [148]. In 2003, Turow found that when offered a choice between paying for their favorite website with cash or with their personal information, over half of respondents said they would rather stop using the site all together [147]. Several experiments investigated under which conditions people will pay more to purchase from websites offering better privacy protections when privacy information is presented in search results and in other salient ways [146, 48].

Much of the current self-regulation approach to online privacy is grounded in the Fair Information Principle of notice. Notice, by its nature, requires communication. As Morgan et al. wrote, “An effective communication must focus on the things that people need to know but do not already. This seemingly simple norm is violated remarkably often in risk communication” [101]. We investigated people’s mental models — beliefs about how a system works, interacts, or behaves. Incorrect mental models may form a view of the world that undermines decision making. For example, if people hold the mental model that any company with a privacy policy is bound by law not to release data, the existence of a link to a privacy policy would seem sufficient in and of itself with reduced reason to read the policy. Research shows that people do, in fact, believe the words “privacy policy” mean they are protected by law [67].

Economics literature suggests that the most someone is willing to pay (WTP) to buy something should be equal to the minimum they are willing to accept (WTA) in payment for it: there should be a point of indifference between the good and cash. A difference between WTP and WTA may be indicative of an *endowment effect*, a phrase coined by Richard Thaler to describe when people place more value on an object that they own. The canonical example is that if two groups are asked to put a value on a coffee mug, people answering without owning the mug will generally suggest a lower price than people who first receive the mug as their own property. The endowment effect does not always occur with abstract items. For example, giving people a token that they can redeem for a mug does not have the same effect as giving them the actual mug [46]. Prior work shows a gap between WTP and WTA for revealing private data (for example, number of sexual partners) in an offline experiment [61]. Acquisti et. al. found substantial differences

between WTP and WTA with gift cards and inexpensive tangible goods, including “subjects who started from positions of greater privacy protection were five times more likely than other subjects to forego money to preserve that protection” [7]. We examine the Acquisti hypothesis in an online context. If there is also a gap between WTP and WTA online, then the way privacy choices are framed may affect the decisions people make about online privacy.

5.3 Research Methods

We followed a two-part approach. First we performed a laboratory study to identify a range of views through qualitative interviews. Then we conducted an online survey to test and validate our qualitative results.

In the first study we performed a series of in-depth qualitative interviews with 14 subjects who answered advertisements to participate in a university study about Internet advertising. Subjects were not primed for privacy. We followed a modified mental models protocol of semi-structured interviews, using standard preliminary questions for all participants, then following up to explore participants’ understanding. Our study ran from September 28th through October 1, 2009 in Pittsburgh, PA. We recruited participants with a notice on a website that lists research opportunities. Participants were compensated \$10 for an hour of their time.

In the second study, we recruited 314 participants from the Mechanical Turk¹ website at the end of April, 2010. We paid participants \$2 for what we advertised as a 20-30 minute study. Median completion time was 24 minutes, which is skewed slightly high by participants who likely started the survey, put it aside, and came back to it later. We saw a drop-out rate of 37%. We deliberately started the study with short-answer questions to encourage people not to take the survey unless they were willing to invest some time, and used the reasonableness of responses to short-answer questions to screen participants. We removed two outliers from our dataset; they had unusually short response times and response patterns that suggested they had not read the questions. We coded free-form responses to tabulate categories of responses, or “unclear” when we were unsure what participants had in mind.

¹Mechanical Turk is crowd-source web portal run by Amazon. See www.mturk.com for details. Mechanical Turk users tend to be better educated, less likely to be working, and more likely to be female than our target population of adult US Internet users. However, the Mechanical Turk population may be “more appropriate” for Internet research, as mturk studies can be closer to representative of Internet users than a random sampling of the full US population [122]. There is a growing literature on how best to use Mechanical Turk in research; see [81], [83], [79], [42]

We asked 64 questions split over nine screens. The screen first asked about purchases online (9 questions), the second about willingness to pay for privacy (6 questions), and the third was a mix of information about their computing environment and views on cookies (3 questions). The fourth page showed depictions of how cookies and data flows might work. The next two pages only appeared for participants who answered the questions on the third page correctly, and asked participants questions about ads based on screen shots (we had inconclusive answers to these sections and omit them in this work). The fifth page (or 7th page for those who answered page 3 correctly) showed a screen shot of the NAI opt-out page, which we removed many member companies from in order to fit on one page (4 questions). The sixth page presented hypotheticals about behavioral advertising and advertising based on email (8 questions). The final page asked demographic questions with a "secret code" to paste in to Mechanical Turk to get paid (9 questions). Some questions, especially Likert questions, were multi-part. We randomized the order of options within questions.

5.3.1 Demographics

Of the 14 subjects we interviewed, 8 were male and 6 female. Half were age 21–29 and half were age 30–59. Participants had diverse professional backgrounds including health, architecture, photography, marketing, and information technology.

For the online study, we slightly over-represented women and our population was notably skewed younger than the adult American Internet population, as seen in Table 5.1. To estimate the demographics for US Adult Internet Users, we combined Pew data [112] with Census data [151]. Because Pew and the Census data record race differently, we cannot estimate the portion of Internet users by race. Instead we contrast to national race statistics from the Census. We under-sampled black and hispanic populations. Our respondents were 74% were white (contrast to 81% nationally), 9% American Indian or Alaskan Native (v. 2%), 6% Asian (v. 5%), 4% Black or African American (v. 14%), and 2% Latina/Latino or Hispanic (v. 16%).² We contrast to Quantcast's estimates for operating systems [116] and Axon's for web browsers [15]. Our sample is skewed toward Firefox users at the expense of Internet Explorer, which suggests a more technically sophisticated sample.

²Both our survey and the Census allow more than one selection for race which is why results sum to more than 100%.

Table 5.1: Demographics for online study

Category	Our respondents	US Adult Internet Users
Male	41%	49%
Female	59%	51%
Age 18-29	55%	28%
30-49	33%	40%
50-64	10%	23%
645+	2%	9%
Windows	85%	87%
Macintosh	11%	11%
Other	4%	2%
Firefox	48%	25%
Internet Explorer	34%	60%
Chrome	10%	6%

Our online survey participants have been using the Internet for an average of 13 years, with 15% online for over 15 years and 2% online less than five years. We asked online survey participants an open-ended question of “If you use more than one web browser on your primary computer, why do you do so?” For those who do, the overwhelmingly most popular reason was that not all websites are fully compatible with all browsers (70%). 18% mentioned switching between browsers when they need more speed. Only one person mentioned security, saying “Safari is safer” than Internet Explorer.

Our participants most commonly check two email accounts (44%). 29% check one email account, 20% check three email accounts, and 7% check four or more email accounts. Most use at least one remotely-hosted and professionally-managed email service: Yahoo Mail (50%), Gmail (50%), Hotmail (23%), or AOL mail (16%). Only 9% of participants reported they do not check at least one email account of this type.

5.3.2 Transferability

Early in the online study, before we asked questions that might affect participants’ views, we asked the same three questions Turow et al. asked in their study designed to be representative of the US population [148]. As our sample is not a statistically representative sample of United States Internet users, we contrasted to the Turow work to understand the transferability of results to other contexts [64]. We found similar results for two of their three questions, as shown in Table

5.2.

Table 5.2: Percentage of respondents who want tailored content

Do you want websites you visit to show you...	Turow et al.'s	Our results
ads that are tailored to your interests?	32%	45%
discounts that are tailored to your interests?	47%	80%
news that is tailored to your interests?	40%	41%

Our respondents’ differ demographically from the Turow population in an important way: our sample is skewed younger. Where the representative Turow sample is comprised of 35% of people aged 18 – 34, our sample is 69% in that age range. However, despite age-linked differences in responses, our younger sample does not explain why we saw a substantially higher percentage interested in tailored discounts. We had approximately 20% more interest in tailored discounts in all of our age categories as compared to the Turow work, as seen in Figure 5.1. One possible explanation: we recruited participants willing to spend 20 minutes to answer our survey for \$2 on the Mechanical Turk website. Our participants may be unusually sensitive to financial incentives. For tailored ads and news, our findings mirrored the Turow paper: most respondents are not interested in tailored advertisements or news.

5.4 Perceptions About Cookies

A variety of technologies help facilitate online behavioral tracking for targeted advertising. Third-party cookies are used by advertising companies to set cookies associated with ads embedded in first-party sites; when browsers load advertisers’ ads, they also get advertisers’ third-party cookies. Beacons are associated with invisible or hidden page elements, and again may be first- or third-party. Flash cookies were designed to store information like volume levels for flash content, but are now used in tracking [129]. Browser fingerprinting is a technique that uses information from web browsers’ user agents (for example, the specific version number and operating system) plus potentially other information available from javascript (for example, the specific order fonts load on the system.) By using small bits of seemingly unidentifiable information in concert, approximately 80% of browsers are uniquely identifiable[45]. We planned to survey participants

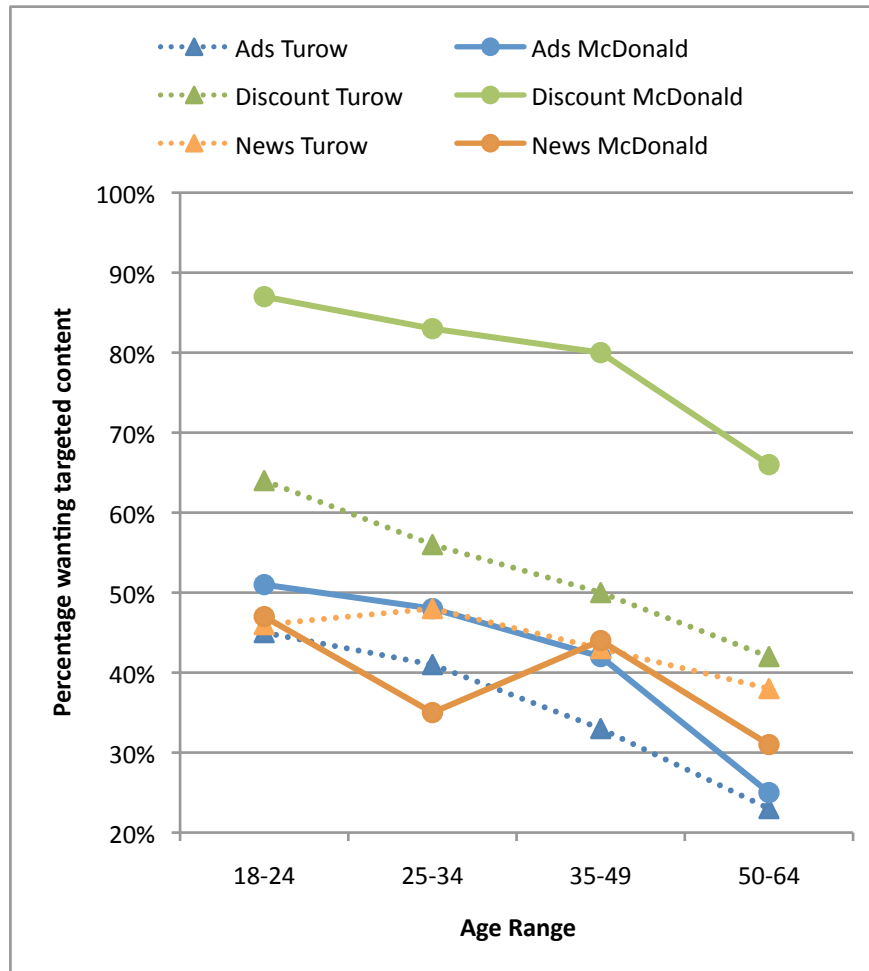


Figure 5.1: Percentage of respondents interested in targeted ads, discounts, and news by age groups in the Turow study, as contrasted with our results

about what they understood about each of these tracking mechanisms.

Our lab study quickly disabused us of any idea of studying user perceptions of beacons, flash cookies, session cookies, or browser fingerprinting: these techniques are invisible to users to the point we would be wasting our time and theirs to ask about them. A few people had heard of session cookies or third party cookies, and those who had were able to give mostly accurate answers. No one had heard of flash cookies, with participants guessing things like they are cookies that “appear in a flash and are gone.” We focused on first- and third-party cookies in part because they are such a popular mechanism in advertising, but also because they are the only technology users are even passingly familiar with. Cookies have been around, discussed, and studied across

decades [3, 63]. If users understand behavioral advertising well enough to attempt to enact their privacy preferences, they are most likely to be able to do so via cookie management. We asked questions to study participants' knowledge of cookies, how they manage cookies, and see if they understand the industry self-regulation approach of setting cookies to opt out of viewing behavioral advertising.

All participants in the interviews had heard of cookies before but we observed widespread confusion. When asked, "What is a cookie?" nearly a third of participants replied immediately that they were not sure. Slightly more than a third of participants gave an answer that was at least partially correct without also saying something factually incorrect. Only one person articulated that a cookie can contain a unique identifier.

5.4.1 Misperceptions of First Party Cookies

While interview participants generally did not understand what cookies are, perhaps it is more important that they understand the effects of cookies rather than their mechanism. We asked follow up questions of "are there ways cookies can help you?" and "are there ways cookies do not help you?" Over a third of participants said that cookies can be related to saving passwords. Similarly, three participants answered that cookies allow them to remain logged in to websites without retyping a password, though during follow-up questions they did not actually know if cookies were involved (as opposed to Apple's Keychain Access, etc.). Three participants believed cookies store their preferences for websites, including details like preferred colors and placement of site elements.

Only three participants said that cookies are related to personalized advertisement. They expressed three very different perspectives. One participant said she has no choices about cookies, because if you "say no then you don't get to go to the site. That's not much of an option." She could not think of any way cookies help her. For ways cookies do not help, she said sites use cookies to personalize, and that "could mean more personalized advertising. It makes me feel like they expect me to be gullible." A second said cookies are things "that programs use to gather information about sites [visited], functionality, and demographics for an ad." He said that "if asked for information [people] would say no," and believes he has "no choices" about cookies. He said that cookies are good when "a set pattern of behaviors, sites, topics, or hobbies" can give

“information on products and services that are more interesting,” but “some [cookies] are used negatively to exploit a person’s history,” and “cookies open pools of information one might prefer to stay private.” Drawing an analogy to shopping offline, he said “you may be shopping in a public place but there is a privacy issue” with companies “knowing where you spend money and time.” Even with a computer collecting and storing the data, there still must be a “person manipulating and interpreting that.” A third participant said advertisers use cookies to “find out as much as [advertisers] can without asking for names,” to gain an “idea of what sort of person” you are. He mentioned ISPs trying to “find ways to catalog this wealth of information,” to pair ads to an audience. He described this practice as a “smart thing” and “reasonable.” He then volunteered that he believes ISPs are constrained by law not to share information. When asked what the law entails, he answered he was not sure and perhaps constraints were not from law but that there would be a “public uproar” and a “bad image” for any company sharing even anonymous customer data. He made the analogy to phone service where recording conversations can be illegal, and said there are “certain cultural norms and expectations” to privacy. Notice the analogies to off-line settings as participants form their views of how privacy works online. Legal protection of privacy in telephone conversations and postal mail are often assumed to carry over to Internet communications as well.

5.4.2 Knowledge of Cookies

Based on the responses we heard in interviews, we asked questions in the online study to understand participants’ knowledge of cookies with options of True, False, or Unsure. See Table 5.3 for details.

Participants understand that cookies are stored on their computers, rather than stored remotely (91% correct.) Participants also mostly understood how cookies are used in not needing to re-enter passwords, and personalizing advertising and websites. Three quarters of participants understand advertisers can use cookies across multiple websites to understand which sites they have visited, and half believe cookies can be combined with data that identifies them by name. This suggests a working understanding of cookies and advertising. However, participants held other views that show they are confused, including half of participants who believe if they do not accept cookies their location can not be identified, and half believe cookies contain informa-

Table 5.3: Responses to factual questions about cookies — correct answers in bold

Description	True	False	Unsure
Cookies are small bits of data stored on my computer	91%	1%	8%
Cookies let me stay logged in over time without needing to enter my password every time I visit a site	77%	8%	15%
Cookies enable personalized advertising based on my prior behavior online	76%	5%	19%
Advertisers can use cookies on multiple websites to learn which sites I visit	74%	5%	21%
Cookies may be combined with other data that identifies me by name	53%	11%	37%
If I do not accept cookies, websites cannot tell where I am physically located	12%	51%	37%
Cookies enable personalized content like color schemes or what type of information I want to see on a website	51%	14%	35%
Cookies contain information from when I first purchased my computer, including my name and home address	13%	48%	39%
Cookies let web browsers' forward and backward arrows work correctly	19%	44%	38%
Cookies are a type of spyware	39%	33%	28%
A website I visit can read every cookie I have, no matter which website the cookie is from	19%	34%	47%
Cookies let people send me spam	38%	29%	33%
Cookies change the color of hyperlinks to websites I have already visited	43%	25%	32%
Cookies let websites display more quickly	60%	19%	22%
By law, cookies may not contain credit card information	30%	11%	59%
The PATRIOT ACT allows law enforcement officials to read my cookies if I exchange email with someone on the terrorist watch list	38%	6%	56%

tion from when they purchased their computer, including their name and home address, with more than another third unsure if this is true or not. Neither of these beliefs is true, and shows a lack of understanding of how cookies work and what they are. Similarly, 44% incorrectly believe cookies enable the forward and backward arrows in their browser, a third incorrectly believe all websites can read all cookies, and a quarter confuse cookies and history, incorrectly thinking cookies change visited hyperlinks to a different color. There is nearly an even split in thirds between participants who believe cookies are a type of spyware, are not a type of spyware, or are

unsure — which is not unreasonable, as there is disagreement within the technical community as to whether some cookies, all cookies, or no cookies should be seen as spyware. 29% believe cookies are responsible for spam, which is not the case, with another 33% unsure. The greatest confusion is around legal protections. 30% incorrectly believe cookies may not contain credit card information by law, with 59% uncertain. 38% incorrectly believe the PATRIOT ACT allows law enforcement to read cookies if they exchange email with “someone on the terrorist watch list,” with 56% unsure. This very last question is the only one we wrote ourselves: all of the rest come directly from perceptions from our lab study participants. Several participants expressed concern that the government could read cookies, but used vague language; we tested a specific example.

5.4.3 Managing Cookies

There are three ways people manage cookies: by not letting them save to their hard drive in the first place, by deleting them automatically, or deleting them “by hand.” We asked about all three methods in our online study.

Several major web browsers offer a “private browsing” feature that allows users to toggle to a private mode that never saves cookies, history, and cache data. When finished, users exit private browsing and have access to their normal set of cookies, history, and cache data. Only 23% reported they ever use private browsing, 50% do not use private browsing, and 27% are not sure if they use private browsing.

17% use software that deletes cookies for them, 23% are not sure, and 60% answered no. Those who answered yes predominately use either anti-malware software or CC Cleaner, though sometimes they had trouble naming the specific product they use (e.g., “malware by anti-malware.”) Some may delete cookies via anti-malware programs without understanding they are doing so. One participant answered “TACO, NoScript, & Firefox,” which is a sophisticated approach.

9% said they never clear cookies, 9% believe they clear cookies themselves annually or less than once a year, 16% a few times a year, 10% monthly, 17% a few times a month, 16% a few times a week, 12% daily, and 8% clear cookies every time they close their browser. This is self-reported data, but about 70% believe they clear cookies at least once a year.

5.4.4 Unclear on Clearing Cookies

Why do people clear cookies? Interestingly, they are not always sure themselves. Nine participants in our lab study self-reported that they clear cookies. Only one of those nine said they clear cookies on their own computer for privacy. Three clear cookies on shared machines out of privacy concerns.

Participants had a vague notion that too many cookies are bad. They are not sure under which conditions they should delete or retain cookies. Though they do not understand about how cookies work, they do understand some of the benefits of cookies, such as not needing to log in again.

For the online study, we asked an open-ended question about why they deleted or saved cookies and coded the responses. Participants wrote answers that reflect an underlying lack of knowledge like “Someone recommended it to me once and I have done it ever since,” or “I’m not very sure what [cookies] are. I have cleared them before because it was suggested to me that I do.” Family is sometimes mentioned as the source of advice, including “Mom told me to,” “My daughter told me to,” and “My husband doesn’t want them.” Similarly for why people do not clear cookies frequently, participants gave answers like “I don’t really know” or “No particular reason.” We coded these vague responses along with a variety of other non-reason or unclear answers as “Other,” which comprised 8% of all responses. In total, our 314 participants gave 390 reasons to delete or not delete cookies. Of 80 reasons not to delete cookies:

- 31% were some form of apathy, either that cookies do not bother participants or they do not care about cookies.
- 27% have software that deletes cookies automatically.
- 20% were not sure what cookies are, or why they would delete them.
- 19% were unsure how to delete cookies.
- 3% (two people) wrote that they do not care about being tracked online.

Of 278 reasons given to delete cookies:

- 33% were based on the idea that “many cookies slow down my computer.” This seems unlikely in practice.³
- 30% had to do with privacy and security. About a fifth of the privacy and security reasons mentioned deleting history; history is commonly confused with cookies. The remaining four-fifths of privacy and security reasons generally reflected some understanding of how cookies work, for example, “I wouldn’t want someone being able to get on my computer and remain logged into my accounts. Also, I don’t want a website tracking me through them.”
- 28% had to do with freeing up hard drive space, reducing clutter, or a notion of hygiene and cleanliness. Answers included “[I] like having a clean slate on the computer all the time,” “[to] clear up clutter,” and “to make space on my computer.” Few modern computers will run into space problems due to cookies.⁴
- 8% mention viruses, spam, or malware. Some tracking cookies are classified as spyware by Norton Anti-virus and other anti-malware programs.

User confusion is high. Some do not know how to delete cookies and might wish to do so, which limits self-help mechanisms in privacy decision making. Some participants reported what seems to be over-clearing of cookies: they delete cookies to avoid issues that cookies do not cause. Cookie deletion creates uncertainty in measuring the number of people — and unique people — who have seen a given online ad, or have visited a given website. Disagreement over ad impressions has slowed the growth of web ads. Counting impressions often depends upon cookie data. Over- and under-counting ad impressions causes economic harms to members of the advertising community, with hundreds of thousands of dollars disputed in large ad campaigns [132]. When users delete their cookies for reasons that do not match their actual preferences, it causes harm without the gains users expect.

³For DSL users, a webpage with a 3000 byte cookie takes approximately 80 milliseconds longer to load [139] so users are not wrong to associate cookies with delay. However, just deleting all cookies without blocking them does not improve time to load the page: websites would simply download new cookies to replace the deleted cookies. Participants may be confusing cookies with cached images.

⁴RFC 2109 suggests browsers implement a maximum size of 4k per cookie and a maximum number of cookies per domain to avoid denial of service attacks from malicious servers filling hard drives [84], and hard drives today are typically measured in gigabytes.

5.4.5 Cookies and Browser History

More than half of our interview participants confused cookies with browser history. Participants did not understand that browser history is stored independently of cookies, which may make it difficult for people to enact their privacy preferences. One participant in our lab study told us cookies contain a “history of websites” visited and when he deletes cookies, “hyperlinks in different colors goes [sic] away, that’s what it does. It clears the navigation history.” When he was a child he lost his computer privileges because his mother could see where he had been based on the color of web links, which he blamed on cookies. Cookies mean “someone else can follow your previous path, and can see what you’ve read before...” In his view, cookies were only an issue on computers where he shared a single account with multiple people. At work, where he signed into his computer account with his own password, he believed cookies could not provide details of his browsing history because he was the only one with access to the account. Notice the confusion around password-protected accounts and privacy protections: several participants had confusion in similar areas and believe they cannot be tracked unless they log in to a website.

Browser user interfaces in which clearing cookies, clearing history, and clearing cache data settings are intermingled may contribute to user confusion. One component of this confusion is temporal: participants reported they delete cookies and clear history at the same time, which leads them to misattribute properties of browser history to cookies. The reason participants clear cookies and history together likely stems from the way they are swirled together in the user interfaces of web browsers. For example, Firefox presents choices about cookies, history, and bookmarks on the same tab, as shown in Figure 5.2a. There is no visual hint that these three topics are distinct. To the contrary, cookies are in the middle of options for history, which serves to convey history and cookies are related. Moreover, Firefox does not expose any cookie options unless users know to change a setting from Remember history to Use custom settings for history. Anyone looking through preference tabs for cookies will not find them in the default configuration. In Internet Explorer, users must select the Tools menu and then choose Delete Browsing History in order to get to the cookie dialog, shown in Figure 5.2b. The easiest way to delete cookies in Safari is to select Reset Safari from the Safari menu, which then presents options to delete cookies and history together as shown in Figure 5.2c. The exception is Opera, shown in Figure 5.2d. Cookies are not mixed in with history. The Opera dialog attempts to define cookies and avoids jargon.

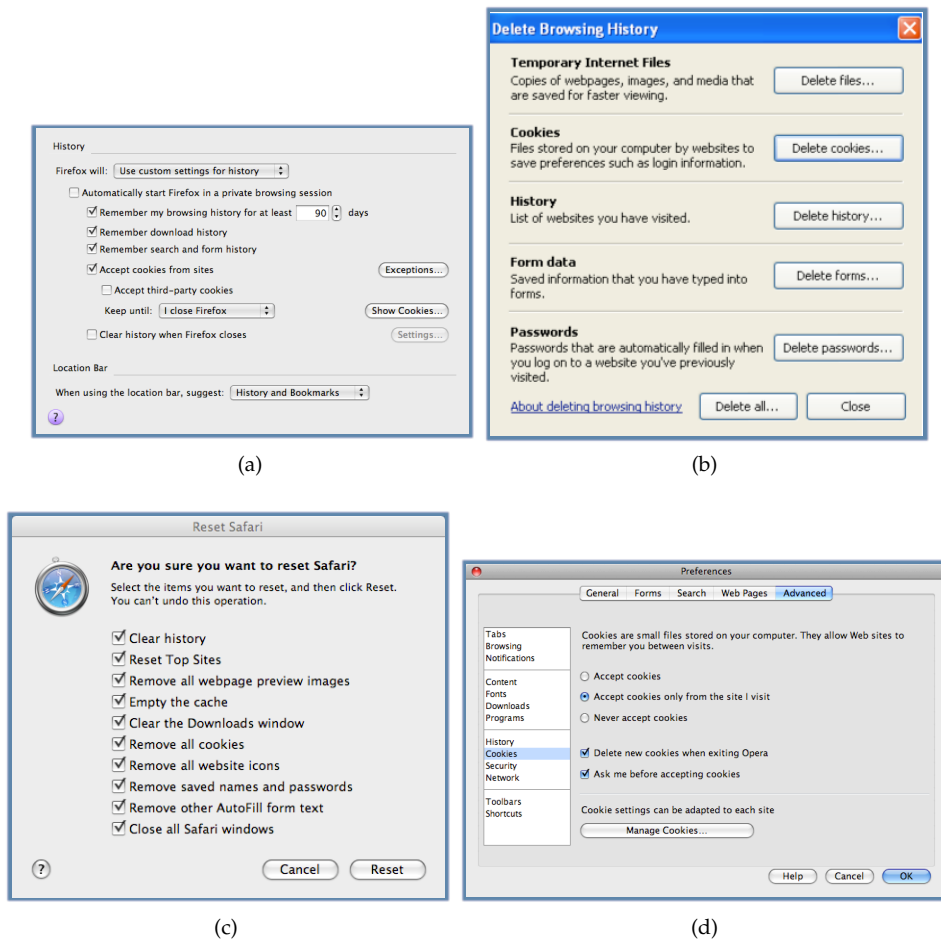


Figure 5.2: Four browsers' interfaces for deleting cookies: Firefox, Internet Explorer, Safari, and Opera.

In the online study, we asked “Sometimes you hear about web browser history. Are cookies and history the same?” 35% of participants incorrectly answered yes. Those who answered no generally had a good working understanding of the difference between cookies and history, with responses like “History is a list of your previous browsing, and cookies are files that registered each site visited.” Of those who correctly answered no, 79% were able to give at least partially correct answers explaining how cookies and history differ, 12% gave clearly incorrect answers, and 8% gave answers that were so unclear we were not able to tell if they understood the difference or not.

5.4.6 Lack of Understanding of Cookies and Data Flows

We knew from our lab study that the phrase “third-party cookie” left participants confused, not because they do not understand what a third-party is but rather because they do not understand what cookies are. We are less interested in definitions of jargon than we are in users’ abilities to make privacy decisions for themselves, so we tried using pictures to elicit participants’ models of how the web works. In particular, behavioral advertising based on third-party cookies works in large part because advertising companies can set and read cookies due to ads hosted on a multitude of websites. If people do not understand these basic mechanics, they will not be able to make informed decisions about accepting, blocking, or deleting third-party cookies.

We asked: “Please refer to the images below to answer questions at the bottom of this page. Imagine you are using a standard web browser to visit The Times website, which has ads as depicted in the diagrams. *There are no other non-visible components to the webpage.*” and gave a choice of four different figures, shown here as Figures 5.3a, 5.3b, 5.3c, and 5.3d, along with a brief text description of each image. We followed up by asking “Which, if any, of the diagrams above could not happen?”

- 22% selected Figure 5.3a, described as: “The Times’ web server sets and reads cookies for all elements on the webpage, including cookies associated with specific ads.” While advertising could work like this, with each host storing and displaying all ads from just the host’s server, modern websites are usually more complicated. 9% answered, incorrectly, that this configuration could never happen.
- 20% selected Figure 5.3b, described as: “Multiple web servers set and read cookies from The Times’ web page.” This graphic introduces the concept of multiple actors with multiple servers, but incorrectly depicts them all being able to read and write cookies from the same section of the website. Servers cannot set and read cross-domain cookies, so this configuration is unlikely, especially in practice. 18% answered that this configuration could never happen, which is a reasonable answer.
- 18% selected Figure 5.3c, described as: “Only the Times’ server can set and read cookies on the Times web page.” This graphic does emphasize the lack of cross-domain cookies, but also shows ads that do not set cookies. While this is possible, it is highly unlikely on modern

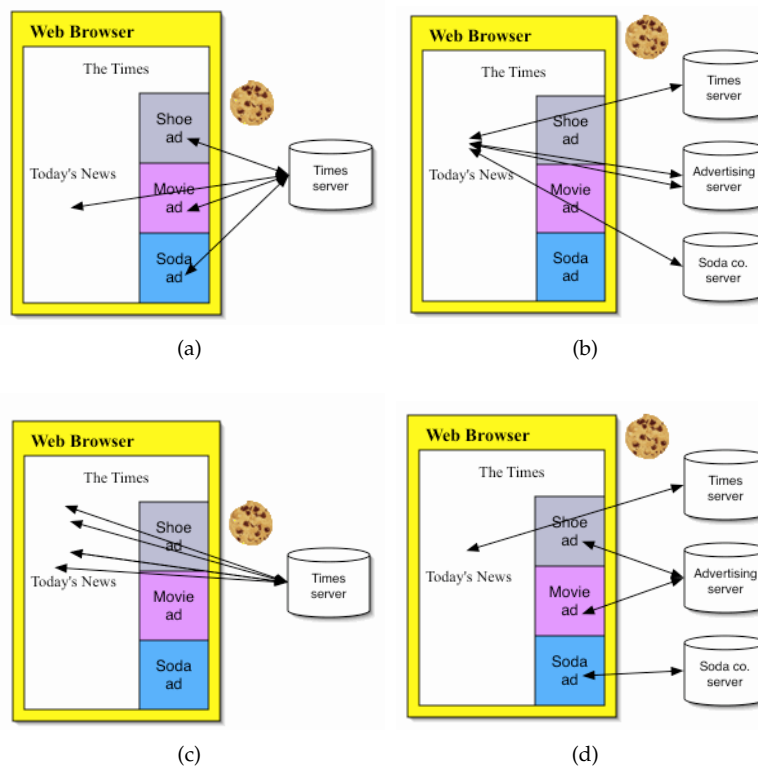


Figure 5.3: Four possible mental models of how advertising cookies work.

sites. 15% answered that this configuration could never happen.

- 40% selected Figure 5.3d, described as: “Different servers set and read cookies from different parts of the Times’ web page.” This is the best choice. It shows common relationships between hosts and advertisers. 10% answered, incorrectly, that this configuration could never happen.
- 48% answered that all four figures are possible.

Participants were most likely to select the graphic that reflects the state of modern third-party cookie use, but not even half gave the best answer. Especially when combined with the majority of respondents confused on what is impossible, it seems people do not understand how cookies work and where data flows. Incorrect mental models of how the web works will make it exceedingly difficult for people to understand what options are available to them, and how to enact their privacy preferences online.



Figure 5.4: Screenshot of the NAI Opt Out page

5.4.7 Consumers Do Not Understand NAI Opt-Out Cookies

None of our interview participants had heard of cookie-based methods to opt-out of tracking cookies, including TACO⁵ and NAI opt-out cookies.⁶ At the end of the protocol, we showed four participants a text description of NAI opt-out cookies from the NAI opt-out website (see Figure 5.4.)⁷

All four participants understood they would continue to see at least some online advertisements. However, there was substantial confusion about what the NAI opt-out does. The text does not disclose that companies may choose to continue all data collection and profiling, and that in some cases the only thing that changes is the type of ads displayed [10]. One participant understood this but the other three did not.

⁵Targeted Advertising Cookie Opt-Out (TACO) is a plugin for the Firefox browser that stores persistent opt out cookies, available from: <https://addons.mozilla.org/en-US/firefox/addon/11073>

⁶The Network Advertising Initiative (NAI) offers non-persistent opt out cookies for all browsers, available from: http://www.networkadvertising.org/managing/opt_out.asp.

⁷Our study used printed materials so we did not test the NAI video, which may communicate more clearly. The degree to which the video's clarity is important hinges on how visitors engage the NAI site.

The first participant believed the NAI opt-out “sets your computer or ethernet so information doesn’t get sent.” She still expected to see ads, but now the ads would be “random.” She said it might “sound old fashioned” but in a choice between “convenience and privacy, I’m going to pick privacy.” She was afraid that opt-out meant “all these people get your information” and therefore “this could be a phishing expedition.” A second participant began his comments by saying “Where do I click? I want this!” He believed the NAI opt-out to be an “opt-out tool so users opt out of being tracked.” He thought “the ads are still there, they just get no data.” A third participant thought it would “reduce the amount of online advertising you receive.” He understood data collection was also involved, but not how, just “some sort of control over what companies use that information.” He would choose to opt-out of companies where “the information they would seek would be too personal to share with a group.” Our final participant understood the NAI text. At first he said if you use Gmail, the opt-out cookie means “stop reading my email and tailoring ads.” He later clarified “What you search is Google property, it’s theirs. They’re going to profile you but not show you that they are.”

During interviews we learned that not only did our participants fail to understand the NAI opt-out page, several of them thought it was a scam. In our online study we learned that is not a widely held view, but neither is the correct explanation for the page’s function. We showed the same screenshot and asked “Based on the image above, if you visited this web site, what would you think it is?”

- 34% answered “A website that lets you tell companies not to collect data about you.” There are some companies for which this is the case. However, some NAI members like Yahoo! continue to collect data exactly as before; they just do not tailor ads to reflect that data.
- 25% answered “A website that lets you tell companies you do not want to see ads from them, but you will still see as many ads overall.” This is incorrect because companies continue to serve ads, just not targeted ads. The ad source is unchanged.
- 18% answered “A website that lets you see fewer online ads.” This is both wrong and prominently disclaimed in the NAI text.
- 11% answered “A website that allows companies to profile you, but not show you ads based on those profiles.” **Correct answer.**

- 6% answered “A scam website to collect your private information.”
- 5% answered “A scam website to find out which websites you have visited.”

These results paint a bleak picture of users’ abilities to make sense of opt-out cookies. Our largest group of respondents misunderstood the NAI text and believed their information would not be collected if they opted out. NAI visitors may think they are selecting which ads they see, rather than targeted v. random ads from the same sources, and make choices that do not reflect their actual preferences. People think the site is a scam at the same rate they understand what it is for. NAI opt-out cookies may not currently be working well as instruments of self-regulation.

5.5 Tailored Content and Privacy Concerns

Advertisers claim consumers are clamoring for more interesting and relevant advertisements, while privacy advocates claim citizens’ rights are being trampled. We found support for both views: there are sizable groups of people with each of those views. In the middle, we found a large group of people who are disinterested in better ads since their goal is to ignore ads in the first place. They see no benefit to targeted advertising, so they do not see reason to share data with advertisers. While they accept the idea that ads support free content, but do not expect data to be part of the exchange.

5.5.1 Mixed Identification of Internet Advertising

Contextual search advertisements are well understood. All participants in our lab study said Google is their search engine of choice. When asked if Google has ads, all participants answered “yes” correctly. Participants knew there are ads down the right hand side of the results page, that “sponsored” links frequently appear at the top of results pages, and that these links are also advertisements. They were all able to recall these details of Google’s advertisements with no prompting beyond asking if there are ads and where they are located. We did not need to show them screen shots of Google search results.

We asked how advertising on Google works. All participants understood that advertisers pay Google to run ads. Participants were less clear on the mechanics of payment. Some thought Google charges for all ads displayed, and some thought Google only charges for ads when people

click on them. No one described beliefs that were technically impossible; everything described has occurred at one time. All told, this is a fairly sophisticated understanding of Google's contextual advertising during search tasks.

In contrast, when we gave participants a printout of a webpage from the *New York Times* and asked them to identify the advertisements, answers varied widely. At one extreme, some participants looked at the graphics only, and discounted anything that came from the *Times* itself (e.g. home delivery and subscriptions) as well as any ads that were text-based. At the other extreme, one participant counted every single item on the page as an advertisement, including hyperlinks in the article to other *Times* articles — and even the article itself. She reasoned the article text was likely a press release and therefore an advertisement. Even while asking specifically about ads, a few people suffered from “ad blindness” and simply did not notice smaller ads that were in unexpected places (e.g. flush against the masthead instead of the right-hand column.) But much of the difference was definitional. While they did not phrase it this way, some participants saw advertisement as strictly a third party endeavor. Anything from the *Times* itself was therefore not an ad.

More interestingly, some participants also discounted all text as a potential source of advertisement. Clearly participants do understand that text can be advertising, or they would not all have been able to answer correctly about Google search ads. Why do some people then discount text as a source of advertisement on the *Times*? We have two hypotheses. First, it could be that Google is uncommonly good at communicating with their users. Ads are always in the same place, the “sponsored” label and yellow background are understood, and the right side is the place people expect to find ads. Second, it could be that people's pre-existing mental models of print media come into play with the *Times*. People have learned with experience that ads in printed newspapers and magazines are usually graphics. To look for text ads on the *Times* people must first unlearn what they already knew, where Google was a blank slate with no direct offline analog. Or it may be a combination of factors that people react to in different ways, which might account for why participants reacted uniformly to Google but with great variance to *Times* advertisements.

Inability to Distinguish Widgets

Regardless of the cause, what the *Times* advertising identification results suggest is that even absent any confusion over technology, participants may have different mental models of advertising. We found participants have a wide range of expectations on the simple question of what is or is not an advertisement on a given web page.

Widgets are another part of a web page, for example an embedded clock or the weather. Widgets are often designed to be customizable to match the look and feel of the web site they are dropped into. Industry guidelines assume people can distinguish third party widgets from first party content and assume that people understand that data flows differently to third party advertisers. Therefore they treat third party widget providers as first party data collectors, subject to fewer guidelines [2]:

In addition, in certain situations where it is clear that the consumer is interacting with a portion of a Web site that is not an advertisement and is being operated by a different entity than the owner of the Web site, the different entity would not be a Third Party for purposes of the Principles, because the consumer would reasonably understand the nature of the direct interaction with that entity. The situation where this occurs most frequently today is where an entity through a “widget” or “video player” enables content on a Web site and it is clear that such content is not an advertisement and that portion of the Web site is provided by the other entity and not the First Party Web site. The other entity (e.g., the “widget” or “video player”) is directly interacting with the consumer and, from the consumer’s perspective, acting as a First Party. Thus, it is unnecessary to apply to these activities the Principles governing data collection and use by Third Parties with which the consumer is not directly interacting.

Instead, we find some people are not even aware of when they are being advertised to, never mind being aware of what data is collected or how it is used by a widget. It appears that self-regulatory guidelines may assume an unrealistic level of media literacy on the part of Internet users.

5.5.2 Mixed Understanding of Current Practices

When we described current advertising practices in our lab study, participants told us they did not believe such things happened. One participant said behavioral advertising sounded like something her “paranoid” friend would dream up, but not something that would ever occur in real life. We asked our online participants about two pervasive current practices described as hypotheticals. First we asked about behavioral ads with the following description:

Imagine you visit the New York Times website. One of the ads is for Continental airlines. That ad does not come to you directly from the airline. Instead, there is an ad company that determines what ad to show to you, personally, based on the history of prior websites you have visited. Your friends might see different ads if they visited the New York Times.

We asked about ads based on content in hosted email, which describes systems in use like Gmail:

Imagine you are online and your email provider displays ads to you. The ads are based on what you write in email you send, as well as email you receive.

Table 5.4: Perceived likelihood of practices occurring

Response	Behavioral Ads	Email Ads
This happens a lot right now	51%	25%
This happens a little right now	35%	14%
This does not happen now but could happen in the future	11%	28%
This will never happen because it is not allowed by law	1%	16%
This will never happen because there would be consumer backlash against companies that engaged in this practice	1%	13%
Other	1%	5%

As shown in Table 5.4, participants seem to have a high degree of understanding that behavioral advertising happens, with only 13% of respondents casting doubt that current practices

occur. Yet only 40% believe advertising based on email content is happening today, and 29% believe this common practice will never occur.

Recall 41% of our participants reported that they check gmail accounts. We found statistically significant differences between gmail users and non-gmail users for the email scenario ($\chi^2=20.1$, d.f.=5, $p<.001$). Gmail users were far more aware that this practice occurs today, with 51% of gmail users saying it happens either a lot or a little now, in contrast to 30% of non-gmail users. It is encouraging to see gmail users are more likely to understand the practices gmail follows, but surprising that half of gmail users do not understand how gmail works. This suggests a lack of informed consent for gmail's business model and a potential for surprise. Gmail users were half as likely to think ads based on email would never happen due to backlash (8% v. 16%) but equally likely to think ads based on email are barred by law (15% v. 16%).

For both scenarios we asked, "How would you feel about this practice?" (Participants were able to select more than one answer.) As shown in Table 5.5, the most popular answer is that 46% of participants find behavioral advertising "creepy," but a small group of 18% welcome targeted advertisements. Responses on how people feel about advertising based on email are markedly more negative, with 62% saying email should be private and that they find ads based on email creepy. Only 4% of respondents saw email-based advertising as a benefit, and only 9% supported the trade off of data and advertising for free services. This matches what we heard in interviews: people understand ads support free content, but do not believe data are part of the deal.

Table 5.5: Attitudes toward current practices

Response	Behavioral Ads	Email Ads
No one should use data from email because it is private like postal mail	N/A	62%
It's creepy to have advertisements based on my emails	N/A	62%
It's creepy to have advertisements based on sites I've visited	46%	N/A
Wouldn't even notice the advertisements, just ignore them	38%	18%
No one should use data from Internet history	30%	28%
Glad to have relevant advertisements about things I am interested in instead of random advertisements	18%	4%
It's ok as long as the email service is free	N/A	9%
Other	3%	5%

We again contrasted our gmail users to non-gmail users for the email scenario. We did not find statistically significant differences between gmail users and non-gmail users for the email scenario ($\chi^2=9.96$, d.f.=5, $p=.076$). This means gmail users are as likely as non-gmail users to find the practices predominately creepy, and believe their email should be private like postal mail. Those who choose to use gmail are not doing so out of lack of concern for privacy in comparison to non-gmail users.

5.5.3 Reasons to Accept or Reject Tailored Advertising

Based on discussions in the laboratory study, we compiled a list of reasons participants gave for being for or against behavioral advertising. We presented online participants with a seven point Likert scale from Strongly Agree (7) to Strongly Disagree (1), summarized in Table 5.6.

Table 5.6: Mean Likert scores to accept or reject behavioral advertising (Strongly Agree = 7, Strongly Disagree = 1.)

Description	Mean	Agree	Disagree
Someone keeping track of my activities online is invasive	5.7	64%	4%
Behavioral targeting works poorly and I get ads that are not relevant to me, even when they are supposed to be	4.8	34%	7%
I would watch what I do online more carefully if I knew advertisers were collecting data	4.7	40%	15%
I ignore ads, so there is no benefit to me if ads are targeted to my interests	4.7	36%	11%
I ignore ads, so I do not care if ads are targeted to my interests or if ads are random	4.4	31%	16%
I ignore ads, so there is no harm to me if ads are targeted to my interests	4.2	24%	17%
I want the benefits of relevant advertising	4.1	21%	21%
I would stop using any site that uses behavioral advertising	3.6	15%	29%
I am protected by law against advertisers collecting data about me	3.6	16%	34%
I do not care if advertisers collect data about my search terms	2.9	10%	51%
I do not care if advertisers collect data about which websites I visit	2.8	12%	53%

Privacy concerns are top priorities. Nearly two-thirds of our participants agreed or strongly agreed that “someone keeping track of my activities online is invasive,” with only 4% disagreeing

or strongly disagreeing. This phrase comes directly from a participant we interviewed in the lab study, and reflects the way she thought about behavioral advertising. It is phrased in a way that would likely garner maximum response by mentioning an unnamed, but presumably human, “someone” and using the possessive “my.” We suggest the way to understand this result is that if behavioral advertising is framed this way in the press, most Americans will respond poorly to it.

Again expressing privacy concerns, 40% agreed or strongly agreed they would be more careful online if they knew advertisers were collecting data. The wording of this question limits data use to advertisers, which may reduce concern. It also explores the notion of a chilling effect. Respondents at least believe they would self-censor if they knew advertisers were collecting data. While self-reported data is not always indicative of actual behavior, it appears people are considering leaving FaceBook in response to publicity about data flows to advertisers [134]. Advertiser’s practices have the potential to reduce Internet adoption and use, and may already be doing so.

Despite claims that users do not care about privacy, half of participants disagreed or strongly disagreed that they do not care if advertisers collect search terms, or if advertisers collect data about websites visited, both of which occur regularly for behavioral advertising and analytics data. Only around a tenth of respondents agreed that they do not care. However, only 15% self-report that they would stop using sites with behavioral advertising.

In our laboratory study we heard two conflicting attitudes from people who ignored ads. Several people told us that because they ignore ads, they get no benefit from targeted advertising and would therefore rather not have any data collected about them. Other people told us that because they ignore ads, they do not care if ads are targeted or random and they do not care if data is collected. We also wondered if there might be people who just do not care at all, and are not particularly cognizant of data collection as an issue. In the online study we found the strongest agreement with the statement “I ignore ads, so there is no benefit to me if ads are targeted to my interests” (36% agree or strongly agree,) the weakest agreement on “no harm to me” for targeted ads (24% agree or strongly agree,) with the most strictly apathetic option of not caring if ads are targeted or random in the middle (31% agree or strongly agree.)⁸ This suggests that of those who ignore ads, they are likely to prefer data not be collected about them, since they do not see any

⁸We found statistically significant differences in means between “no benefit” and “no harm” as well as “do not care” and “no benefit” ($p < .05$, $df=312$, paired two-tailed t-Test, $\alpha = .05$). We did not find significance between “no harm” and “do not care” ($p = .060$).

benefit. However, just because someone claims to ignore ads does not mean that is always the case. Advertisers may still gain benefit from targeting these users. But an argument that targeted ads are a benefit will likely fall flat with the people who are not interested in any ads, let alone better ads. Interestingly, when we put that question to participants directly, we saw an even split. 21% agree or strongly agree that they want the benefits of relevant advertising while 21% disagree or strongly disagree, with a neutral Likert mean of 4.1. What emerges is neither a strong clamoring for nor a backlash against behavioral advertising, but rather several distinct groups with quite different preferences.

5.5.4 Privacy and Security Among Top Priorities for Buying Online

98% of our participants indicated they make purchases online. More than half said they never make purchases based on Internet ads or email advertising, as summarized in Table 5.7. This is self-reported data; people may make buying decisions based on ads without being aware they are doing so. Banner ads serve a billboard-like function for those who eventually buy online, even months later [91].

Table 5.7: Respondents who buy online

Frequency	Buy online	Buy based on Internet ads	Buy based on email ads
Never	2%	52%	54%
A few times / month	42%	7%	6%
A few times / year	52%	38%	38%

We asked participants how sellers could entice participants to purchase more products online, and listed 13 possible approaches with responses on a four point Likert scale of “Matters a lot,” “Matters,” “Matters a little,” and “Does not matter.” We created our 13 categories based on responses to a pilot test with an open-ended question. See Table 5.8 for results.

The most popular item was free shipping.⁹ The next three most popular were clustered around privacy and security: not sharing data with advertisers, a policy against spam, and fraud protection. In contrast, the remaining privacy and security item on data retention scored near the very bottom. This may be a function of the specific description, or due to lack of understanding of

⁹The word “free” often gets a strong response. It would be interesting to see if this result is robust when phrased as “discounted shipping.”

Table 5.8: How sellers can entice more online purchases (Matters a lot = 4, Does not matter = 1)

Description	Mean	Matters a lot	Does not matter
Free shipping	3.7	75%	1%
Will not share your data with advertising partners	3.6	70%	3%
No spam policy	3.6	70%	3%
Improved fraud protection for credit card transactions	3.6	68%	3%
No hassle return policy	3.6	67%	2%
Clear information about products	3.6	66%	2%
Web discounts	3.5	57%	1%
Easy-to-use website	3.4	55%	2%
Online coupons	3.2	46%	4%
Local pickup	2.4	18%	26%
Will only retain data about your purchases for three months	2.3	14%	24%
Products recommended based on your past purchases	2.3	10%	23%
Products recommended based on your friends' past purchases	1.8	7%	47%

how data retention limits reduce privacy and security risks, but suggests data retention is not currently a major concern for users.

Return policies and clear information about products scored higher than discounts, all of which scored better than an easy-to-use website or online coupons. No clear story emerges about usability vs. financial incentives. Recommending additional products did not interest our respondents, regardless of whether recommendations came from their own purchasing history or their friends. From the discussions we had during our lab-based study, many people find it “creepy” to get suggestions based on friends’ purchasing history. However, we are surprised to see their own purchasing history score nearly as low, when well-known companies like Amazon have successful services in production. This may suggest users do not think about the mechanics behind such recommendations, or just that they think themselves more immune to advertisements than they are in actual practice.

5.6 Payment for Privacy

We have observed that some people who are highly concerned with privacy are strongly disinclined to spend money to preserve privacy. This can seem counterintuitive, especially since in many domains the amount someone is willing to pay for something indicates how highly it is valued. Instead, some people who believe privacy is a right respond negatively to the idea of paying to protect their privacy.

5.6.1 Gap Between Willingness to Pay and Willingness to Accept

We split our participants into two groups. First we asked them to name their favorite online news source, and answer how frequently they visit it to make our next questions more salient. Then one group answered the question “Would you pay an additional \$1 per month to your Internet service provider (ISP) to avoid having your favorite news site collect your data for behavioral advertisements?” The second group answered a similar question of “Would you accept a discount of \$1 per month off your Internet service provider (ISP) bill to allow your favorite news site to collect your data for behavioral advertisements?” In theory, there should be no difference between the price someone is willing to pay (WTP) to protect privacy and their willingness to accept (WTA) payment for revealing information.

We did find a gap between WTP and WTA. Only 11% of respondents were willing to pay \$1 per month to keep their favorite news site from collecting data, while 31% of respondents were willing to accept a \$1 per month discount to disclose the information. Thus, 11% said they were willing to pay \$1 extra to gain privacy while 69% said they were unwilling to accept a \$1 discount to give up privacy. In the privacy sphere this could have two very interesting effects. First, people who think they have already lost the ability to control private information — that privacy is not something they are endowed with — may value privacy less as a result. Those who believe they have control over information may value privacy more as a result. Second, the difference between opt-in and opt-out rates for online privacy may not just be due to the well-documented tendency for people to keep defaults unchanged. If a service collects data by default and users must opt-out of data collection, that suggests users are not endowed with privacy, and they may respond to that cue by valuing their privacy less.

One limitation in our study is that we did not control for participants’ *ex ante* beliefs. Al-

though we assigned participants randomly to both conditions, it is possible that one group was skewed by more participants who believe that payment still will not protect privacy, or a variety of other views that could affect their willingness to pay.

5.6.2 Reasons to Pay or Refuse to Pay for Privacy

We followed up by asking questions to better understand why people would decide to pay or accept \$1, based on reasons we heard from our lab study participants. We asked “Some websites may offer you a choice of paying for content or receiving content for free in exchange for letting them send you targeted advertising. How strongly do you agree or disagree with the following statements?” with a seven point Likert scale from Strongly Agree (7) to Strongly Disagree (1). See Table 5.9 for details.

Table 5.9: Reasons to pay for privacy or accept a discount

Description	Mean Likert	Agree	Disagree
Privacy is a right and it is wrong to be asked to pay to keep companies from invading my privacy	5.9	69%	3%
Companies asking me to pay for them not to collect data is extortion	5.6	61%	5%
It is not worth paying extra to avoid targeted ads	5.5	59%	5%
Advertisers will collect data whether I pay or not, so there is no point paying	5.4	55%	4%
I hate ads and would pay to avoid them	3.3	11%	36%

Only 3% of respondents either disagreed or strongly disagreed that privacy is a right and it is wrong to be asked to pay for privacy online, even in exchange for free content. The top two ranking replies suggest that one reason people will not pay for privacy is because they feel they should not have to: that privacy should be theirs by right. Yet when phrased as an economic proposition, that it is “not worth paying extra,” participants also predominately agree. One might expect that participants who highly value privacy would disagree, and would think it is worth paying for privacy even if they also believe they should not have to do so, but only 5% did. Distrust of the advertising industry, or perhaps of actors on the Internet as a whole, is another reason people may not be willing to pay for online privacy with just over a majority agreeing or

strongly agreeing that data will be collected even if they pay companies not to collect data. Finally, we can rule out dislike of advertising as a major factor in online privacy decision making, with only 11% willing to pay to avoid ads because they “hate” them. Most participants are accustomed to advertising. Mass media advertising has been part of life since before they were born. It is the data collection that is new, and, to many, a troubling aspect of online advertising.

5.7 Conclusions and Discussion

From what we have observed to date, it appears behavioral advertising violates consumer expectations and is understood as a source of privacy harm. While we do not attempt a full analysis of possible policy responses here, we note several things. First and foremost, consumers cannot protect themselves from risks they do not understand. We find a gap between the knowledge users currently have and the knowledge they would need to possess in order to make effective decisions about their online privacy. This has implications for public policy, commerce, and technologists. One younger participant said in frustration that she did not learn about how to protect her online privacy in school, she was just taught typing. We believe there is a serious need not just for improved notice of practices, but for the education requisite to understand disclosures. Most non-regulatory approaches require consumers to understand tradeoffs and to know enough to take whatever actions will enable their privacy preferences. At the current moment that seems unrealistic, but the outlook could improve in the future.

In general, users do not appear to want targeted advertisement at this time, and do not find value in it. However, a small but vocal subset of users are genuinely eager for relevant ads. They are matched by a subset of users vehemently against the practices that enable targeted ads. In the middle, the majority attempt to ignore ads and see no benefit to giving data to advertisers. Ideally, users could choose for themselves but at present they lack the knowledge to be able to make informed decisions.

Most users understand that cookies store data on their computers, enable tailored ads, and allow tracking across sites. They are unclear on important details like whether cookies may be combined with other data, what data is stored in cookies, if blocking cookies preserves geolocal privacy, and they are particularly unclear about laws and law enforcement. Web browsers may contribute to users’ confusion. Browsers may also be an avenue to help with user under-

standing and decision making in the future. Thus far, browser makers have been largely absent from behavioral advertising issues, and some do not directly profit from behavioral advertising. Microsoft has been involved with behavioral advertising for years, and their adoption of P3P in Internet Explorer changed the third-party cookie landscape. Google's Chrome browser may be another opportunity to welcome browser makers as major stakeholders with tremendous ability to help Internet users make privacy decisions. However, the *Wall Street Journal* reported that Microsoft re-designed Internet Explorer 8 specifically to enable third party tracking for business reasons [156]. It may be naive to expect browser makers to support user privacy at their own expense. The NAI is as a major player in behavioral advertising but their opt-out cookie page is very confusing, with only 11% understanding what it is for. With their leadership role in self-regulation, the NAI may not be supporting Internet users' ability to avail themselves of self-help options.

We found people generally unwilling to pay for privacy, not because they do not value it, but because they believe it is wrong to pay. Paying to keep data private was termed "extortion" by some participants. We also found a gap between willingness to pay to protect data and willingness to accept a discount in exchange for releasing the same data. People may ascribe more value to what they possess. People may value their privacy less when presented with an opt-out for data collection, which suggests data belongs to the company collecting it, rather than an opt-in choice for data collection, which suggests data belongs to the individual.

One of the questions posed by the advertising industry is "where's the harm" in behavioral advertising, with a suggestion that a formal benefit cost analysis should occur before regulation. This question seems to ignore privacy loss as a distinct harm. In contrast, our participants spoke frequently about their privacy concerns. 40% of participants in our online study agree or strongly agree they would watch what they do online more carefully if advertisers were collecting data, which suggests advertising may cause a chilling effect. In our lab study, one technically-savvy participant even described withdrawing from online life as a result of privacy concerns.

With lack of understanding of and a lack of interest in tailored content, unless industry moves rapidly towards an effective self-regulatory solution, regulation may be needed. One possible path for regulation is to require opt-in for all forms of advertising other than contextual. However, opt-in systems are not a panacea: they can be designed so users click them away without

understanding them, and once users opt-in it may be difficult to reverse the choice. If industry elected to, they could use self-regulation mechanisms to improve decision making through education, improved technology and tools, and more privacy-protective policies far more quickly than regulators could act. These tasks will be challenging no matter which parties take the initiative.

Chapter 6

Beyond Behavioral

6.1 Introduction

In this chapter we contrast reactions to contextual advertising to four types of targeted advertisements:

- *Behavioral advertising* creates profiles for Internet users based on a variety of different data types and inferences drawn from those data [92]. For example, Alice visits her local newspaper, The Newtown Bee. She loads the newspaper and also its ads into her browser. When her browser loads the image for an ad, she gets a cookie from the advertising network that placed the ad on the newspaper's site. Alice later visits her favorite hobbyist site, The Quilting Bee, and they use the same advertising network. When her browser loads an ad on The Quilting Bee site, the advertising network can read the cookie it set before at The Newtown Bee and update it with information about her visit to The Quilting Bee. In this way the advertising network can learn about Alice's interests, the ads she clicks, make good guesses about her age range and sex, and combine this with her approximate physical location based on her computer's IP address. This information is summarized into a profile that indicates if Alice is a good target for certain ads, with interest categories like "cars" or "Hawaiian travel." Google and Yahoo! both use behavioral advertising, and made their interest categories public at the end of 2009 [35]. The Yahoo! ad server reaches over half a billion unique people each month, with 9.7% of the market [14]. Google's DoubleClick and AdSense ad servers have a combined total of 56% of the market and reach at least 1.5 billion unique users each month [14]. More recently, Gomez et. al. found Google web beacons on 88% of nearly 400,000 sampled websites and 92 of the top 100 most popular sites [57].
- *Affiliate marketing* stems from the observation that an individual's buying habits are similar to the people he spends time with. If Bob is friends with three people who just bought SUVs, Bob is more likely than average to be in the market for an SUV himself. The most famous example of affiliate marketing was the Facebook Beacon program. Beacon displayed items that people had purchased to their Facebook friends. In Facebook's case, people supply the data themselves on who they are connected to, and Facebook worked with specific partners using javascript and iframes to tie purchases and reviews on external sites back to Facebook users [126]. Other methods might include data mining to determine social connections,

rather than relying on user-supplied data.

- *Cloud computing* allows users to upload content to a remote host, and access it over the Internet. Users do not have to worry about backing up the data or transmitting it between computers. Companies can match keywords in hosted content or its metadata to ad topics [100]. For example, Eve sends email to her sister about an upcoming beach vacation, and based on the words “beach” and “tan,” she sees ads for suntan lotion next to her email messages. Popular examples of end-user cloud computing include email services (hotmail, gmail,) and image hosting (flickr, snapfish.)
- *Deep packet inspection (DPI)* refers to looking beyond just the information in headers required to determine how to route a packets of information across the Internet, and inspecting the data content as well [22]. Based on users’ content, ISPs create profiles, or does keyword matching in real-time, and injects related ads into websites. ISPs are able to intercept all traffic from a given user, putting them in an even more comprehensive position to create profiles than large advertising networks. In the United States, NebuAd partnered with half a dozen ISPs before coming under Congressional scrutiny.¹

These categories are not always confined to separate silos. For example, Google Buzz mixes behavioral, affiliate, and cloud-based advertising strategies. Profits vary in each case, not just by amount but by which type of entity they accrue to. Legality and regulation differs for each, with complicated and sometimes yet unknown status. But the basic idea is the same: by cataloging users’ actions online, advertisers can move closer to their goal of getting the right ad to the right person at the right time, and users may benefit from this too.

We studied how Internet users perceive five different types of advertising — contextual, behavioral, affiliate, cloud-based, and DPI-based — via an online survey. Because most Internet users are not familiar with these types of advertising, we pilot-tested multiple descriptions paired with illustrative examples before settling on two different sets of examples, one centered around cancer and the other around scuba.

In this chapter we begin with a look at related work in section 6.2. In section 6.3 we describe the research questions that informed our approach, which we outline in section 6.4. We summa-

¹Comcast also famously used DPI and received regulatory scrutiny, but that was to reduce peer-to-peer traffic, and not to inject advertisements.

size our results in section 6.5, and conclude in section 6.6.

6.2 Related Work

Anton, et. al., performed some of the earliest work on behavioral advertising in 2002, with a follow up study in 2009 [12]. They found the types of Internet users' privacy concerns remained stable, but the level of concern has increased around the type of information used for behavioral advertising. This includes "websites collecting information about previously visited websites," and data like "purchasing patterns, and targeted marketing and research" [12]. TRUSTe commissioned two studies using stratified samples in 2008 and 2009, and found a 6% decrease in concern about behavioral advertising over one year. TRUSTe found that even if it "cannot be tied to my name or other personal information," only 28% of Internet users would feel comfortable with advertisers using web browsing history, and 35% believe their privacy has been invaded in the past year due to information on the Internet [145].

In prior work, we conducted in-depth studies with a small sample to understand participants' mental models of online advertising, and found participants were unable to describe how cookies work, did not agree on which parts of a website are advertisements, and misunderstood how NAI opt out cookies function [94]. We also found browser's user interfaces may contribute to widespread confusion about cookies [95]. Turow et. al. conducted a nationally representative phone survey in 2009. They found 66% of adults do not want tailored advertising, which increased to as high as 86% when participants were informed of three common techniques used in advertising today [148]. The Progress & Freedom Foundation published a paper critical of the Turow et. al. work, in part because PFF appear to disagree with the concept of survey results informing public policy issues, and in part because PFF believed study participants should have faced an economic tradeoff for declining targeted ads [136].

The last two years have seen tremendous activity from government and industry. AOL conducted a study on behavioral privacy and launched a user education campaign featuring a cartoon penguin [117]. The FTC released and revised behavioral advertising guidelines [53] and held numerous events. The Interactive Advertising Bureau (IAB) and other industry groups released behavioral guidelines to their membership [2]. The Network Advertising Initiative (NAI) went one step further and prohibits their member companies from using flash cookies for targeted

advertising [70]. Many customer-facing changes have happened recently: the Future of Privacy Forum unveiled their “power I” icon to denote targeted ads [36], Google and Yahoo! published their behavioral categories [78], and the IAB rolled out their “Privacy Matters” ad campaign [120], TRUSTe released a widget for advertisers to provide user notice and choice about targeted advertising [144], and Better Advertising, which formed to help advertising companies comply with FTC guidelines, now offers the Ghostery plugin to alert users to web bugs [23]. Press releases around all of these changes lead to increased media coverage of behavioral advertising.

6.3 Research Questions

Behavioral advertising has received a great deal of attention, primarily from the FTC. Do users share the FTC’s concern? We hypothesized that users would be even less accepting of other forms of advertising, particularly cloud- and DPI-based. This hypothesis was supported.

The advertising industry asserts users prefer relevant behavioral ads, even though there is anecdotal evidence that clumsy targeting actually makes ads appear less relevant [131]. Do users prefer behavioral advertising to random ads, and do users self-report willingness to pay for random ads rather than targeted ads? While, as we discuss further, we cannot answer from this study whether people would pay for random ads in practice, we can contrast the differences between how these five forms of advertising are valued. We hypothesized users would strongly prefer contextual advertisements to random ads, which was not fully supported: contextual was most popular, but only by a narrow margin. We further hypothesized that cloud- and DPI-based advertising would be less valued than random ads, which was supported.

Privacy is frequently described as “highly contextual.” Do users have stable privacy preferences even when the contemplated context changes? We used two different sets of illustrative examples, with half of our participants reading about examples related to cancer, and half about scuba. We hypothesized that the cancer condition would evince greater concern about data collection. We are unable to draw a strong conclusion at this time, but surprisingly, we did not see many statistically significant differences between the cancer and scuba conditions.

6.4 Methods

We conducted an online study in January, 2010 with a total of 300 participants. We used a between subjects design with participants randomly assigned to two conditions, one with examples about cancer ($n=148$) and the second with examples about scuba ($n=152$). In the cancer condition we defined terms and used examples centered around a cancer diagnosis. In the scuba condition we used the same definitions, but changed the examples to be about scuba. *A priori* power analysis showed we would need 105 participants in each condition to have a 95% chance of identifying a medium effect size ($d=0.5$), which we exceeded to ensure we would have sufficient statistical power even after eliminating outliers [49].

We recruited participants from Mechanical Turk² and paid them fifty cents to complete the study. Because privacy laws and norms differ between countries, we wanted to eliminate nationality as a possible confound. We used Mechanical Turk to limit participants just to those in the United States, which we later verified in demographic questions.

We had a substantial drop out rate with 42% of respondents who loaded the survey electing not to complete it. This was by design, in accordance with research on how to elicit high-quality responses in Mechanical Turk [81]. We wanted to screen out the portion of Mechanical Turk users who might just “click through” the study without engaging the material, so we placed essay questions on the front page. The first screen signals that this is a study that will take some time and thought to complete. We expected to then eliminate respondents who had implausibly low response times, but did not have any: our fastest participant took five minutes for a study with a median time of eleven minutes. Subjectively, the quality of responses was quite high, with participants writing substantial answers in response to essay questions at the start and end of the study.

6.4.1 Study Questions

Study questions comprised several groups:

- *Open-ended definitions.* We began the study by asking questions asking participants to define common Internet technologies. We also asked how participants read email and which ISP

²Mechanical Turk is crowd-source web portal run by Amazon. Mechanical Turk serves as a market place for tasks that require human judgement or are expensive to automate. See www.mturk.com for details.

they use.

- *Advertising sections.* We randomized the presentation order of the different advertising sections. We described each advertising practice with an example, then asked the same set of questions for each practice. We also asked about trade offs for the practice: would participants prefer random ads, and would they pay to avoid a given type of advertising practice? We followed up with an open-ended question of “Why?”
- *Demographics.* We concluded with basic demographic questions to help us understand our sample population.

6.4.2 Analysis

We used ANOVA to test differences in means between Likert variables. Since we used a seven-point Likert scale, we are able to treat it a continuous variable. We also used ANOVA to test for significant differences between the responses we coded to why participants would or would not pay \$1 per month to avoid targeted advertising, and if there is a difference in the proportion of participants in the cancer and scuba conditions who prefer random ads. ANOVA tests may be used with dichotomous variables provided the proportions measured are under 80% [90], which our data fulfill. We performed all statistical tests at the $\alpha = .05$ significance level.

We coded free-form responses to why participants would or would not pay \$1 per month to avoid targeted advertising into one of twelve categories of response, or “unclear” when we were unsure what participants had in mind. We found no significant differences in “unclear” responses across conditions, with 6% eliminated from further analysis. To verify our coding is repeatable and robust, a second researcher coded 20% of the “why” responses, with a high 95% replication rate.

6.5 Results

6.5.1 Definitions

We asked for definitions of cookies, third party cookies, and behavioral advertising in essay format. Participants were instructed not to look up answers but rather to provide their own, and

that it was ok to say they do not know.

Participants were most familiar with cookies. We asked, “Some websites use ‘cookies’. What is a cookie?” Only 5% of respondents were unable to answer at all, ranging from the simple “I don’t know,” to “I am not sure what a cookie is, I know to clear them out on my computer every now and then but I have no real idea as to what they are.” However, many definitions were vague, confused, or simply wrong. While participants generally were not able to describe what cookies are, they were at least sometimes able to explain what cookies do. For example, participants may understand cookies can be involved in storing login credentials even while also mistakenly believing cookies are commonly used to store IP addresses. As we have seen in prior work, participants confused cookies, history, and occasionally bookmarks [94].

Participants were less familiar with third party cookies and 28% of participants answered they do not know what they are. Those who attempted to answer did fairly well, providing answers like “Information that a site sends you that originates from a different site; for example, when visiting site A, I get cookies from A and from B. The cookies from B are third-party cookies.” Many participants mentioned “tracking” or “advertisers” as part of their answer. While participants understood what a third party is, their lack of knowledge about cookies themselves carried over and diminished their ability to understand third party cookies. For example, one participant defined cookies as “like a bookmark” and third party cookies as “A bookmark from another website.”

Similarly, 23% of participants did not know what behavioral advertising is. Answers given were usually correct, for example, “In behavioral advertising, the ads that you see are selected based on the things that you are clicking on and the pages you are viewing. This ensures that you see adds that would be of more interest to you and you will pay more attention to them.” The few wrong answers were similar to “Using some sort of psychology to try to get consumer to buy a product—maybe like showing beautiful, thin, fun woman playing a sport to sell beer.”

62% of respondents stated they did not know what cloud computing is. Most answers were vague with a few that had the right idea like “Cloud computing is computing based on the internet instead of computing based on your hard drive” and others far afield like “Using pop-ups,” or even speculation about marijuana use.

6.5.2 Advertising Sections

We asked questions about different methods of advertising. We asked about contextual ads and contrast those results to targeted advertising including behavioral, affiliate, cloud computing, and DPI-based. We presented each section on its own page, shown in random order.

We pilot tested our descriptions of advertising and learned it was very difficult to explain these concepts without using examples. Examples might influence what people think of advertising practices. We incorporated this into our study design by using examples about scuba or about cancer. We selected scuba as an innocuous hobby that would have comparatively few privacy concerns. Cancer, however, is something people would want to maintain control over information and tell loved ones and employers carefully. Cancer has the advantage that it is a condition that can happen to anyone regardless of lifestyle choices and has no moral approbation attached to it. For example, the text to describe contextual advertising for the scuba condition was:

Some websites use contextual advertising to change the ads they display in response to your current actions online. For example, if you search for the word "scuba," you might see ads for scuba gear. You would not see ads for scuba gear again while searching for other topics.

And the text to describe contextual advertising for the cancer condition was:

Some websites use contextual advertising to change the ads they display in response to your current actions online. For example, if you search for the word "cancer," you might see ads for medical clinics. You would not see ads related to cancer again while searching for other topics.

To evaluate reactions to a given type of advertising, we presented eight adjectives: Concerning, Creepy, Desirable, Entertaining, Helpful, Invasive, Pushy, and Reasonable. We selected these adjectives by collecting a month of online news reports about behavioral advertising and looking at word frequencies in the news coverage, which included both press releases from industry and pieces critical of industry practice. We selected words with clear antonyms, and balanced the adjectives as half positive and half negative traits. We presented the adjectives in random order and asked a question like, "Please indicate how much you agree or disagree that the words below describe contextual advertising," with a seven point Likert scale from "Strongly Agree" to "Strongly Disagree." We presented the adjectives in a random order each time.

We found only two of the eight adjectives had significant differences in means between the cancer and scuba conditions: pushy ($p < .033$) and reasonable ($p = .022$.) As expected, cancer

shows greater concern with more people agreeing advertising practices are pushy in the cancer condition (cancer $\mu = 5.1$, scuba $\mu = 4.9$.) However, not as expected, fewer people disagree that advertising practices are reasonable in the cancer condition (cancer $\mu = 3.5$, scuba $\mu = 3.3$.) The split result and few significant differences between cancer and scuba could suggest that people have stable privacy preferences independent of the examples selected. However, we do not feel confident offering a strong conclusion at this time. We ran a very similar study eight months prior and we did find significant differences, with participants in the cancer condition expressing greater privacy concerns as we would expect if participants do not have stable privacy preferences. We plan to continue longitudinal studies to understand differences over time.

In contrast, we found significant differences between the mean Likert scores for adjectives describing all five of the advertising types ($p < .001$ for all.) Overall, we found participants were fairly neutral on contextual advertising, with mean scores close to neutral for all adjectives. They were more likely to agree with negative adjectives than positive adjectives for behavioral advertising. They were even more negative toward behavioral, DPI, and cloud-based advertising. This ranking remained stable across both the cancer and scuba conditions.

We found a predominately stable order of adjectives across conditions and advertising types. Participants most strongly agreed a given advertising practice was invasive ($\mu = 5$), followed by pushy ($\mu = 5$), creepy ($\mu = 4.9$), and concerning ($\mu = 4.8$.) Participants most strongly disagreed that a given practice was entertaining ($\mu = 3$), followed by desirable ($\mu = 3.1$), reasonable ($\mu = 3.4$), and helpful (3.5.) We averaged the mean Likert scores for the four positive adjectives and four negative adjectives into an easier-to-visualize combination in Figure 6.1.

6.5.3 Preferences for Random Advertisements

We asked if participants would prefer seeing randomly selected ads, or each of the types of advertising we studied. For example, we asked:

If a website you visited frequently told you they used contextual advertising and offered you a choice of receiving contextual advertising or randomly chosen ads, which would you chose?

This question gets to the heart of the issue of whether Internet users find target ads to be beneficial. If they have to see ads, would they prefer ads that are relevant to their interests? We expected participants would prefer contextual ads to random ads, particularly since contextual

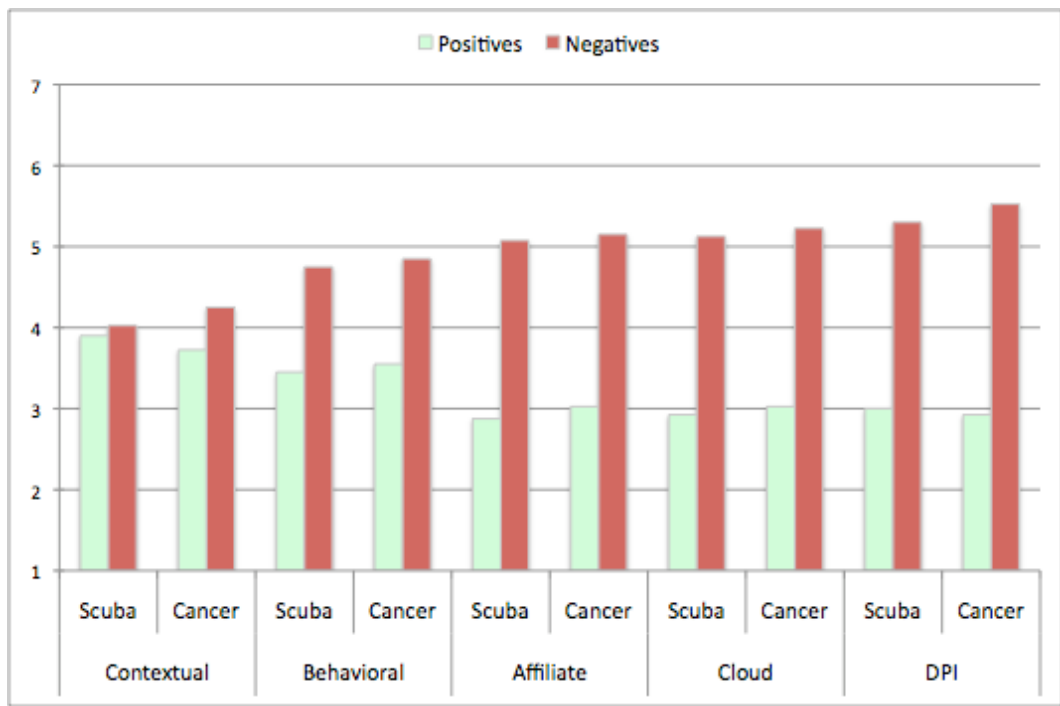


Figure 6.1: Level of agreement with positive and negative adjectives describing advertising practices, ranging from strongly disagree (1) to strongly agree (7.)

search ads are pervasive and not particularly controversial today. And indeed, contextual ads are favored by 53% of respondents. What surprises us is that this is so close to an even split. Behavioral ads were only favored by 35% of respondents. Affiliate, cloud, and DPI were similar, with only 21% to 25% favoring them over random ads. This result strongly suggests the majority of Internet users do not want targeted ads. However, a sizable minority do want targeted ads. Ideally this minority will be able to gain the benefits they see from targeted ads, perhaps by an opt-in system.

We found significant differences between advertising types. As seen in Figure 6.2, affiliate, cloud, and DPI were significantly different from behavioral and contextual, but not distinguishable from each other. We found no significant differences between the cancer and scuba conditions.

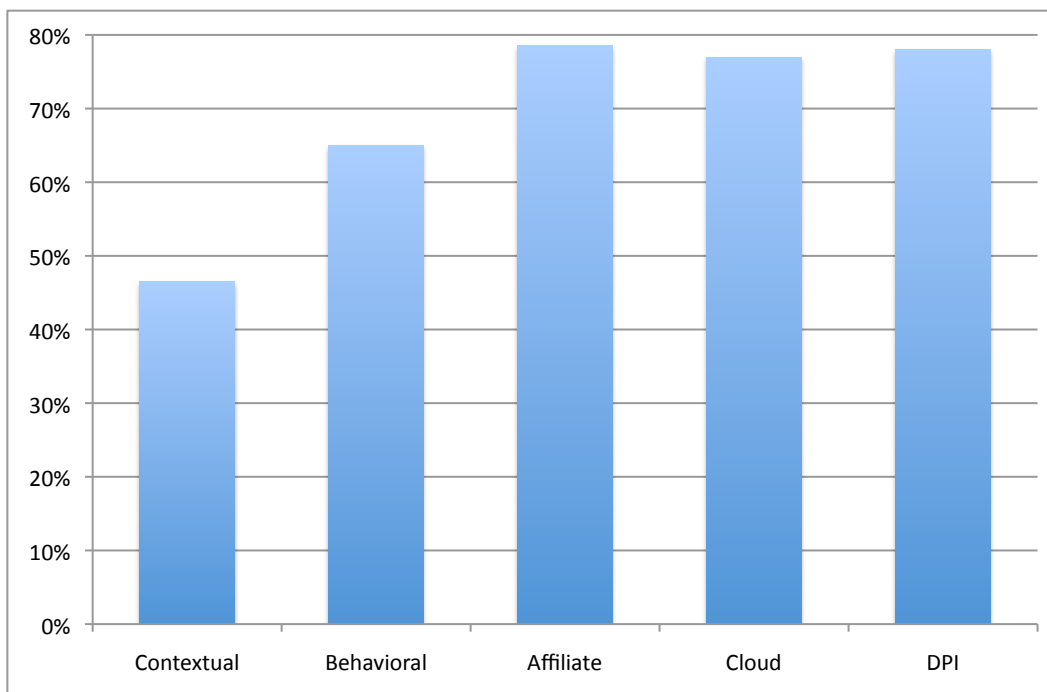


Figure 6.2: Percentage of respondents who prefer random ads to different types of targeted ads. Scuba and cancer conditions had no significant differences so their average is shown. Affiliate, cloud, and DPI are not statistically different from each other, but do significantly differ from contextual and behavioral.

6.5.4 Tradeoffs with Targeted Advertisement

From our pilot tests, we anticipated participants would prefer not to see targeted advertisements. We followed up on the Turow et. al. work in 2003 that found people self-reported they were not willing to participate in targeted advertising, even when it would cost them \$6 to avoid targeted ads [147]. We tried proposing a lower price point of \$1 per month, even though this is likely below what the market will pay for serving targeted ads. Note that our research is not an attempt to determine willingness to pay: we are using self-reported data, and willingness to pay \$1 may be affected by the anchor of paying Mechanical Turk users 50 cents to participate in the study. This should also not be taken as an estimate of absolute numbers of people willing to pay to avoid targeted ads, which would require a different study design and a more experimental approach. We studied the relative differences between conditions. We asked questions like:

Would you pay \$1 per month to avoid having your ISP collect data about you for DPI-based advertisements?

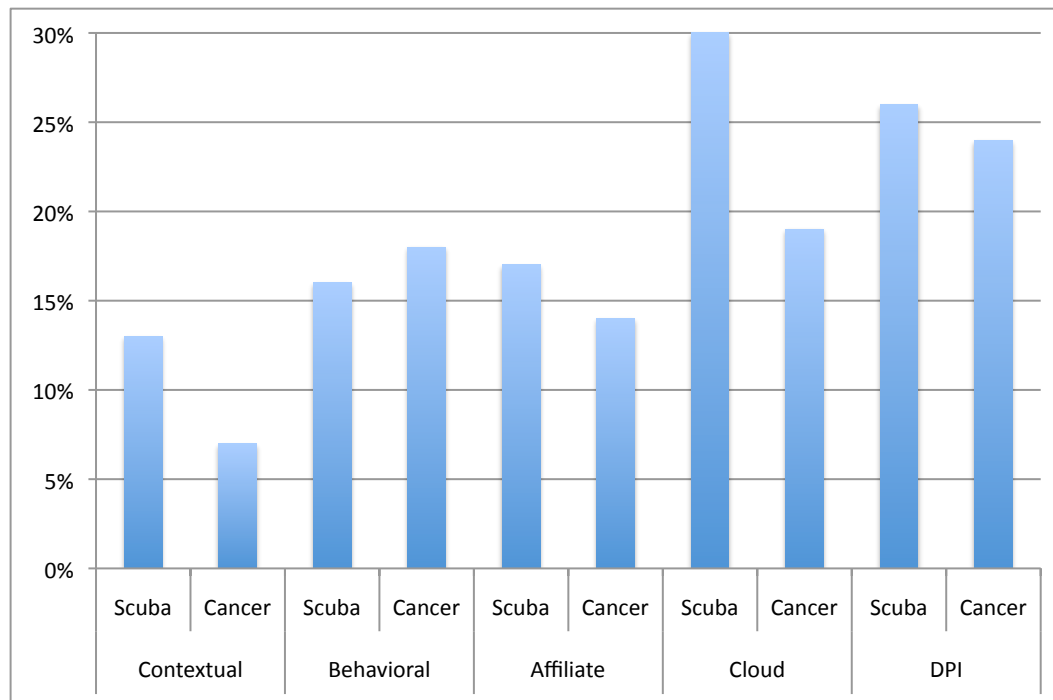


Figure 6.3: Percent who self-report they would pay \$1 per month to avoid ads

We found no significant differences between the cancer and scuba conditions. We did find significance for ad types, nearly exclusively between contextual ads and the other four targeted types, rather than between types of targeted ads, as shown in Figure 6.3.

These results are difficult to interpret because they capture two different competing tendencies. Participants reported they were more willing to pay to avoid data collection they feel is invasive. However, they also refuse to pay to avoid data collection “on principle,” as we discuss further in the next section. People may have greater expectation for privacy with medical data, which would decrease willingness to pay. This is reasonable, as medical data is afforded greater protection by law and industry groups have recommended against creating medical categories for behavioral targeting.

6.5.5 Rejection of Payment for Privacy

After we asked participants if they would pay \$1 to avoid specific types of advertising, we asked them simply “Why?” This gives us interesting qualitative data about privacy decision making. Originally, we asked “why” in an early pilot test, with the expectation we would find common

themes and offer them as multiple choice answers. However, a subsequent pilot test found participants were selecting every answer they believe is true, rather than providing their reasoning for why they would or would not pay a dollar to avoid a type of advertising. The open-ended format of asking “why?” elicited more useful, though less quantifiable, data.

We found a dozen different clusters of common answers, which we describe and follow with illustrative statements from participants in italics:

Want Benefits — Participants want relevant advertisement, find the ads helpful, or like that they get free content. *I prefer to have information presented to me that is relevant to what I care about or It's sometimes helpful or If I'm searching for something, I don't mind having suggestions offered at that particular time. I have found this to be helpful in the past. A person could spend hours searching for a subject. Also, sometimes it mentions good discounts or Ads on pages I'm looking at don't really bother me. I know that having them there keeps most of the services I use free.*

Don't Care — Participants do not mind ads and/or data collection. *It doesn't bother me so much or I am not embarrassed by my searches. If they want to collect my data, go ahead or I just don't care. My expectation of privacy is not that great for most activities online.*

Unaffected — Participants believe the type of advertising does not apply to them, either because they do not use the type of service selected as an illustrative example (social networks, hosted email) or because they use technological measures that they believe protect their privacy (cookie management, anti-virus software.) *Social networking doesn't appeal to me or Major search engine contextual advertising can be easily blocked with Hosts file entries pointing to 127.0.0.1 or It would depend if the virus protection program I have prevents such tracking. I believe it does.*

Ignore Ads — Participants will never buy from online ads, so see no reason to pay money to change the type of ads they ignore. *I will live with it....I am not going to pay any more just to avoid something that I can ignore.*

Financial Concerns — Participants do not have money for extra expenses, feel they pay too much for Internet services and software as it is, or do not believe it is worth the money. *I don't have the money! or Not paying to remove ads...its a stupid concept to get your money. The internet is a free environment, people always think they have to make money and take advantage of others while*

doing it... or It is a nuisance, but not worth paying extra to avoid or Not worth a dollar a month to keep it from occurring.

Targeting Fails — Participants believe targeting results in such poor matches to their interests that they do not see targeted ads as a concern or as a benefit, and will do their own research if they wish to buy something. *Affiliate ads are even more irrelevant. I'm not sure how affiliate ads differ from random ones, in terms of what I'm actually likely to click on or Because IMO it's invasive. It's like having a not very astute person eavesdropping on your conversation and every time he hears you say something food related he pops out with an egg salad sandwich.*

Stop Use — Participants will self-censor, give up using sites with that type of advertising, choose a different ISP, or stop using the Internet all together. *I would start censoring what I say and do online or I don't like it. Think I will go back to writing letters or I would rather stop using the internet then pay to keep my info private. It is taking advantage of their customers.*

Pointless — Participants believe data would be collected regardless of what they pay, there is no point paying for privacy when you cannot trust guarantees given and/or asking not to be tracked could lead to greater threats. *They would still collect my web surfing data, just not putting ads on my screen based on that or Having my activities on the internet tracked is less concerning than giving out my credit card information to not have them track me. or I don't want someone who is planning to do something illegal to be able to hide their intentions by paying the \$1. or Because how would I know they actually stopped for the \$1. I would think they would take the dollar and keep doing whatever they do.*

Ethically Wrong — Participants think it is wrong to be asked to pay money to avoid data collection when they have a right to privacy, or believe the practice described should be prohibited. *I don't think I should have to pay to protect myself from such an invasion of my privacy! It should be illegal! or It should be a basic privacy right to not deal with this issue. My information should be kept private and not used for anything. Monitoring like that leads to no good or I think all people should have the right to privacy, not just the wealthy! or That's not a choice, that's extortion.*

Invasive — Participants find the practice disconcerting or do not want to be part of it, regardless of whether they are willing to pay to avoid it or not. *For this type of advertising online companies/ISPs must keep track of your internet browsing, and I don't like the idea of someone keeping*

a file of my activities. It feel[s] too "Big Brother" to me or STAY OUT OF MY COMPUTER or I share my pc and ads popping up that indicate what I've been doing on my pc would be invasive and a disregard of my privacy or Privacy. The company I use to send personal email should not be looking through my emails. Sounds like communist China.

Pay for Privacy — Participants value their privacy and are willing to pay to protect it. *It's worth \$1 per month to me to not have my browsing data available somewhere out there where I don't have control over it or ISP's already can range wildly in price, if a decent ISP is offering a less invasive service for a small markup it would seem acceptable.*

Hate Ads — Participants are willing to pay to avoid some types of advertising. *I don't like ads. I do anything to avoid them or Would rather not have advertising at all, and would pay to rid myself of it.*

Of these categories, only the last two were strongly associated with willingness to pay \$1: people will pay for privacy or because they hate advertising. The Invasive category spanned people who would and would not pay. The remaining reasons are almost exclusively reasons why people would choose not to pay.

We found no significant differences between the cancer and scuba conditions. For advertising types, we found no significant differences in the proportion of people who reported they have Financial Concerns (21%), believe payment is Ethically Wrong (15%), Ignore Ads (6%), think it is Pointless to pay (4%), Hate Ads (4%), and believe Targeting Fails (3%). Again, our results should be taken in the context of comparisons between conditions, rather than an accurate measure of representative results from the Internet population as a whole. For example, as we discuss in the Demographics section, our sample is heavily comprised of students and the unemployed, which may skew the percentage of those expressing Financial Concerns to be higher. We are also in the middle of an economic downturn, and recruited by paying 50 cents, which may serve to anchor participants, or at the very least, screen for participants willing to see 50 cents as worth ten minutes of their time. Other than Financial Concerns, there are no obvious demographic skews to this data. Interestingly, the relatively high proportion of people who volunteered that paying for privacy is Ethically Wrong did so at statistically indistinguishable rates across all five advertising types. Our respondents did not see privacy rights on a sliding scale: for many, charging money for privacy is simply wrong. "Extortion" was a more popular response than allusions to Big

Brother. Some responded angrily, thinking the study was designed to determine market size for a new product, and chided us for lack of ethics. The remaining four categories were around the same size. A small subset of people volunteered that they hate ads, but this does not seem to drive decision making nearly as much as privacy concerns. A similarly small subset have lost all trust online and believe companies will invade their privacy even if explicitly paid not to.³ In nearly the opposite direction, another small subset believe advertising is so mis-targeted that there is no need to worry about seemingly inept companies. A third small subset expressed the view that there is no reason to pay to avoid specific types of advertisement since they ignore all ads anyway. In some cases this overlapped with people who answered the Don't Care and suggested, or outright stated, data collection is not an issue. In other cases, participants seemed not to think about implications for data collection.

The remaining six categories have at least one significant difference between means for advertising types, as shown in figure 6.4.

- 9% of respondents volunteered they find contextual advertising Invasive, growing to 28% for DPI-based advertising. We found no significant differences between behavioral and affiliate advertising, and cloud-based advertising only had a significant difference with contextual advertising. With an overall average of 22%, respondents were more likely to comment that practices were invasive than any other type of response. However, unlike other categories, Invasive was more of a commentary than a reason: both those willing and those unwilling to pay \$1 gave their view that practices are invasive.
- Overall, 14% self-reported that they would pay to protect their privacy, ranging from 3% for contextual to 21% for DPI-based advertising. This does not mean that 21% of Internet users would pay their ISP an extra dollar every month not to use DPI — self-reports in a survey do not translate well to actions in practice. What this finding does capture is a large difference in concern regarding advertising types. We found statistically significant differences between contextual and all other types of ads, between affiliate/behavioral and cloud/DPI, but not between affiliate and behavioral or between cloud and DPI.
- A small but concerning proportion reported that if faced with advertising technologies,

³This study occurred not long after Facebook changed their policies to publish previously private data. In pilot studies closer to Facebook's action, Facebook was frequently mentioned by name as a reason to mistrust Internet companies. In the final study, Facebook was only explicitly mentioned once.

they would Stop Use and withdraw from online life. Only 1% indicated they would not use sites with contextual advertising, climbing to 11% for behavioral advertising. This does not mean that in practice, a tenth of Internet users would either constrain their behavior or abandon the Internet if they found out they have a behavioral profile. However, it points to a substantial degree of concern. Further, we only surveyed Internet users. If people are choosing not to go online for privacy concerns, they would never appear in our sample. While we did find affiliate marketing significantly different from DPI, cloud and behavioral advertising for those who would stop using part or all of the Internet, this may be due to a higher portion of people not using social networking.

- Affiliate marketing is the highest in the Unaffected category at 7%, dropping to 2% for DPI-based advertising.
- Overall, 8% of respondents said they Want Benefits from advertising technologies such as free content or relevant advertising, ranging from about 4% for DPI, cloud, and affiliate marketing to 11% for behavioral, to a high of 20% for contextual advertising. As advertisers maintain, there are Internet users who want tailored advertising. They are a minority, and our research does not address how well they understand tradeoffs of data privacy, but they can be an enthusiastic market. They find value in advertising as an economic good, rather than an economic bad they would pay to avoid.
- More prevalent than those who want the benefits of advertising, but less prevalent than those who find techniques invasive, is the group of people who report they Don't Care, ranging from a low of 10% for DPI to 29% for contextual advertising. Again, just as we do not expect everyone who said they would stop using the Internet would actually do so in practice, if we were to speak with users right after they learned their ISP used DPI to sell advertising, we would probably find far fewer than 10% saying they do not care about the practice. What we do find is that people care more about DPI, cloud, affiliate, and behavioral technologies than about contextual. We did not find statistically significant differences within the four non-contextual advertising types.

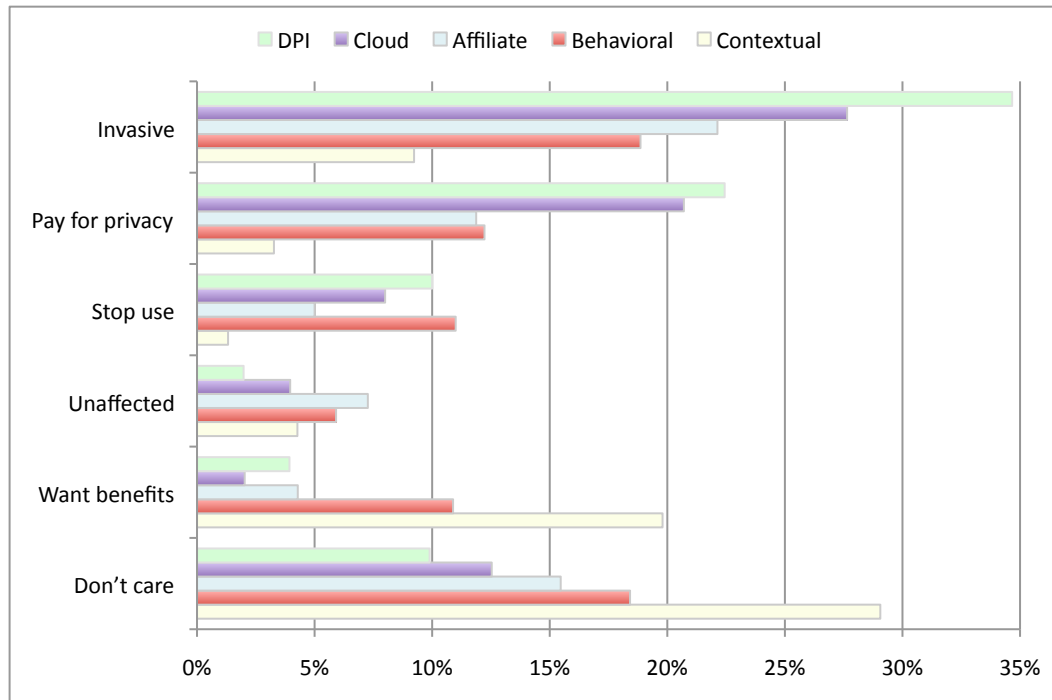


Figure 6.4: Percentage of respondents who volunteered a given category of answer when asked why they would, or would not, spend \$1 per month to avoid a type of advertising.

6.5.6 Demographics

Overall, 58% of our respondents were female and 41% male. 74% of participants were white, 10% Asian, 7% Latino, 6% African American, and 4% Native American, Hawaiian, or Alaskan. All conditions were similar.

We asked how long participants have used the Internet. The mean and median answer was 13 years, with a standard deviation of 3.8 years. The most common occupations were unemployed and student, with 18% each. Most respondents were professionals outside the information technologies fields, which were 7% of the sample.

Income was generally bell-shaped, centered on the \$25,000 - \$37,499 and \$37,500 - \$49,999 categories with 16% each. Larger tails of less than \$12,500 and more than \$125,000, with 13% and 5% respectively, reflect our unemployed and student population on the low end and our primarily professional population on the high end.

6.6 Discussion

As with other studies, we find a small subset of Internet users welcome behavioral advertising, but the majority reject it as a privacy invasion. Given a choice they would rather see random ads than targeted ads. Ideally, Internet policy will allow users who derive benefit from behavioral advertising to do so, but in a way that does not adversely impact the privacy of other users. One possibility is to make all targeted advertising opt-in, rather than opt-out. Advertisers fear they would be subject to low opt-in rates, below users' actual preferences for targeted ads, because many people accept defaults without changing them. That is a valid concern. One potential technological solution is to force a choice. For example, web browsers could query about opt-in or opt-out status and store persistent data that does not go away when users clear cookies. Many other solutions are possible, each with different tradeoffs. A good solution may require architectural changes and technical solutions, rather than exclusively viewing privacy as a policy problem.

People do find behavioral targeting concerning, but less so than other similar advertising approaches. We hope behavioral is the first step, not the end, to the FTC's interest in targeted advertisement. If the FTC heavily regulates behavioral but not other forms of targeted advertising, it is likely to push advertisers to what users consider to be more invasive practices. Our analysis suggests that the FTC might be better served to think in terms of privacy outcomes and data flows, not just specific technologies or approaches, since specific implementation details are not the FTC's goal. However, we agree there are times when concrete guidelines are vastly preferable to overly-broad, open-to-interpretation principles. This is always a difficult balance. Ideally the FTC will devise both principles and operational guidelines that work together, and evolve with technological innovation.

Medical data is particularly sensitive, but we did not see much difference between the cancer condition and the scuba condition. Our results suggest that users' expectations are to have all data protected by an opt-in only system, not just medical or other sensitive data. However, we caution this finding is preliminary, and we plan to do more work in this area.

We welcome multiple online user education campaigns. Once people understand what cookies are, they likely will understand third party cookies too "for free" because they already understand the phrase third party. Our results suggest education should focus on cookies first, which

people still do not understand, rather than focus on third party cookies in particular. Explanations of how targeting works and how decisions about opting in or out work would also help users make privacy decisions. We found in prior work that the NAI opt-out cookie description created user confusion [94]. We recommend continued careful testing of education materials before reaching out to millions of Internet users.

A perennial question is “why won’t people who care about privacy pay for it?” As Tsai et al. found, people will pay a premium for privacy if they have clear information at an actionable time [146]. But our work finds multiple barriers to paying for privacy, including for those who value privacy most. A substantial minority feel privacy is a right and trading money for privacy is wrong. They found even the idea of paying for privacy offensive. These people will not take steps to protect themselves because they already expect to be private and safe online. In contrast, half as many people think they cannot be private online, so they will not pay for privacy either, seeing it as a waste of money. A much larger third group will not pay for anything online beyond what their ISP charges. Some of this is purely financial, perhaps transitory during a tough economic climate, but other answers reflected the belief that the only valid price on the Internet is free. The idea of paying for online activity of any sort baffles them. Finally, a small number of participants erroneously think their anti-virus software protects them from all unexpected data collection. This demonstrates that even those who will pay for privacy and do take action to protect their privacy require more education in order to be effective. Combined, these four groups of people who will not pay money for online privacy are far larger than the group who will. This does not mean people are indifferent to privacy, as the large proportion who labeled targeting practices as invasive establishes. Presenting privacy as an economic tradeoff is in opposition to how people think the world works. This has implications for anyone trying to sell privacy protections, paid for either with cash or in time, since many people expect legal protections. And it suggests privacy is more complex and nuanced than is able to be understood exclusively by watching people’s actions and treating people as black boxes.

Finally, we have concerns about participants who report chilling effects from targeted advertising. People indicated they would use the Internet less, or give up on it all together. This conflicts with the goals of, and public investment in, a national broadband plan. It also conflicts with profits for advertisers. Driving potential customers away is counter-productive. Since we

only studied people who are current Internet users, we do not know to what extent late adopters are avoiding the Internet due to privacy concerns, or how many people use the Internet less than they used to. This seems a fertile area for future work.

Chapter 7

Flash Cookies

7.1 Introduction

Adobe sells several products related to Flash technologies. Some of Adobe’s customers are currently being sued for using Flash to store persistent data on Internet users’ hard drives, allegedly contrary to users’ knowledge and as a way to bypass users’ privacy choices. In this chapter we set out to quantify current use of Flash technology and measure the prevalence of “respawning” deleted HTTP cookies. We review related work in section 7.2 and describe our methods in section 7.3. We present our findings in section 7.4. We conclude in section 7.5.

7.2 Background and Related Work

Flash is used to create multimedia applications including interactive content and animations embedded into web pages. An estimated 99% of desktop web browsers have the free Flash Player plugin enabled [8]. Flash programs cannot read and write HTTP cookies directly, but can use JavaScript to do so [30]. However, while JavaScript is built into all major browsers, users can choose to disable JavaScript and thereby disable Flash’s ability to interact with HTTP cookies.

In 2005, Adobe acquired Macromedia, which had itself acquired FutureWave in 1996, and in doing so purchased Flash. Flash 6 was released in 1996, prior to Adobe’s involvement, and introduced an analog of HTTP cookies called Locally Stored Objects (LSOs) or *Flash cookies*. Flash cookies let Flash programs save things like volume settings or game players’ high scores to their local hard drives. In addition to not depending on JavaScript to access HTTP cookies, Flash cookies are attractive to developers because they hold more data and support more complex data types than HTTP cookies, giving developers more flexibility and control. See Table 7.1 for a summary of some of the differences between HTTP cookies and Flash cookies.

Table 7.1: Technical differences between HTTP cookies and Flash cookies

	HTTP	Flash
Scope	Just the browser that set it	All browsers on the computer
Longevity	Default: until browser closes	Permanent unless deleted
Maximum size	4 KB	100 KB
Data types	Simple Name/Value pairs	Complex data types

Aside from technological differences, users interact with HTTP cookies and Flash cookies in

different ways. Most users do not fully understand what HTTP cookies are but at least they have heard of them; few users have heard of Flash cookies [97]. Users have access to HTTP cookie management through browsers' user interfaces, but until recently the only graphical interface to manage Flash cookies was to set their maximum file size to 0 KB on Adobe's website. Several privacy enhancing technologies (PETs) manage HTTP cookies including anti-spyware packages, Internet Explorer's cookie controls through P3P, browser-specific plugins, CCleaner, and private browsing options built into browsers. Until recently, PETs did not address Flash cookies. So long as Flash Cookies stored innocuous data, these differences were primarily technical details, largely uninteresting to anyone but software engineers. That changed as advertisers started to use Flash cookies to store data that has nothing to do with Flash.

Advertisers use persistent identifiers in cookies to help them understand a given customer's browsing history. This data is used to build interest profiles to command premiums for ads matched to given interests or demographics. Advertisers also use cookies to contribute to analytics data about which customers have viewed ads, clicked on ads, and purchased from ads. Analytics data helps advertisers determine if a given ad is effective with a particular audience. More importantly, without at least basic analytics, advertising networks would not know how much to charge. Meanwhile, many users prefer not to be tracked and express that preference by deleting their HTTP cookies. This caused tremendous problems for analytics data, where even a small error rate can result in incorrectly billing thousands of dollars in a single advertising campaign.

Because users did not know about Flash cookies, and tools did not delete Flash cookies, advertisers discovered Flash cookies solved their data quality problems. Even better for advertisers, Flash cookies do not expire. Under Windows, Flash cookies write to hidden system folders, away from most users' notice or technical ability to delete. Flash cookies are cross-browser, eliminating advertisers' problem with HTTP cookies that a user using Internet Explorer and Firefox is miscounted as two different users. Rather than write a lot of new code to work with Flash cookies, in some cases advertisers simply used Flash cookies to identify a user and then re-create ("respawn") that user's previously deleted cookies, enabling advertisers to continue to use their existing code base. Users did not have to bother re-entering data or to be pestered with information about behind-the-scenes internal processes. Indeed, Flash cookie use sounds like the "best

practices” put forward in a W3C document on mobile web use:

Cookies may play an essential role in application design. However since they may be lost, applications should be prepared to recover the cookie-based information when necessary. If possible, the recovery should use automated means, so the user does not have to re-enter information [133].

As a technical response to the technical problem of poor quality analytics data, Flash cookies are a good engineering solution. However, problems collecting analytics data are not just a technical glitch: users *intentionally* delete HTTP cookies as an expression of their desire for privacy. In the United States we rely upon an industry self-regulation approach to privacy, built on a notice-and-choice theory. Using Flash cookies to respawn or to track users who have deleted HTTP cookies completely undermines user choice, and violates the underlying principles of self-regulation. Users had no visible indication that Flash cookies existed or that HTTP cookies respawned.

That started to change in 2009 with the publication of Soltani et. al.’s paper investigating the use of Flash cookies [129]. They found over half of the sites they studied used Flash cookies to store information about users. Several things changed after this article was published: the press popularized their findings, privacy advocates raised awareness, a few companies using Flash cookies announced new practices, the NAI published guidelines that their member companies must not use Flash to respawn HTTP cookies, some PETs added Flash cookie management, web browsers added Flash cookies to the files they manage during private browsing, and the FTC requested more information from Adobe. In 2010, the Wall Street Journal ran a series of articles about Internet privacy including findings from another Soltani-led study of 50 websites’ use of Flash cookies and other tracking technologies using data collected at the end of 2009 [1]. Several class action lawsuits are currently pending [33].

This chapter provides another data point in the rapidly changing realm of Flash cookies. We investigated more sites than the Soltani studies, though we did not investigate sites as deeply. We also extend knowledge about Flash practices by investigating a random sample in addition to the most popular sites on the web where prior studies focused.

7.3 Research Methods

We investigated three different data sets, all based on Quantcast's list of the million most popular websites:

- 100 most popular sites as of July 8, 2010
- 100 most popular sites from July, 2009 to match the first Soltani study
- 500 randomly selected sites

Looking at the 100 most popular sites captures data about the sites users are most likely to encounter. This is the same method Soltani et. al. used in their study. We revisited the same sites in the initial Soltani paper in order to perform a direct comparison. During the course of the year, 31 sites that had been in the top 100 in 2009 were displaced with different sites in 2010. Because the most popular sites may not follow the same practices as the rest of the web, we also sampled a random population of 500 sites. We list all URLs in Appendix D.

We used two identically configured Windows laptops (XP Pro, version 2002, service pack 3) with Internet Explorer 7 configured to accept all cookies and reject pop ups. We used the most recent version of Flash Player, 10.1. Our two laptops were on different computer networks so they would not have similar IP addresses, eliminating IP tracking as a potential confound. Flash cookies are stored in a binary format. We used custom code from Adobe to save the content of Flash cookies in a text file, which allowed us to automate comparisons of log files rather than open each Flash cookie in a SOL editor. We followed the following automated protocol:

1. Delete all cookies and cached data on both laptops
2. On laptop A, for each site:
 - (a) Launch Internet Explorer
 - (b) Visit the site
 - (c) Wait 60 seconds to allow all cookies to download
 - (d) Move all HTTP cookies, flash cookies (*.sol and *.sor) and log files to another directory
 - (e) Visit the site and move all cookies two more times to get a rotation of ads

- (f) Quit Internet Explorer
 - (g) One final move of all cookies to get anything cached that was saved on exit
3. On laptop B, same procedure as for A above.
 4. Identify sites that have any Flash cookies on either laptop (no need to investigate sites with only HTTP cookies, since by definition they cannot be respawning.)
 5. On laptop A:
 - (a) Copy the final set of Flash cookies only (not HTTP cookies) that had been on laptop B for that site into the `..\Application Data\Macromedia` directory
 - (b) Visit the site
 - (c) Wait 60 seconds to allow all cookies to download
 - (d) Move all HTTP cookies, flash cookies (*.sol and *.sor) and log files to another directory

At the end of this procedure, we compared HTTP cookies on laptops A and B, noting which cookies had not been identical before, but were after using the same Flash cookies. This suggests, but does not establish, that the information in the HTTP cookie propagated from the Flash cookie. See Figure 7.1 for a graphical depiction of how we classified sites.

We collected data during July, 2010, prior to new press coverage and lawsuits. We visited each site three times on both laptops. We did not clear cookies or Flash cookies during these three sweeps. We did copy HTTP cookies and Flash cookies after each sweep so we could determine when they had been set. After we completed the three sweeps per site, then and only then we deleted all HTTP and Flash cookies.

We did not traverse websites; we only visited the top level of any given domain. As an example of where that would affect results, some sites start with login pages and only have flash content after users login. We did not do any logins or deep links, which puts our counts as a minimum bound. We also did not interact with any Flash objects. This is less of a concern for quantifying Flash respawning, as sites using them for tracking would typically not want to require user interaction before setting LSOs. Similarly, if companies are using LSOs to uniquely identify visitors to their sites, we expect they would do so immediately and not require interaction with Flash content. However, do expect that we undercounted the total number of sites

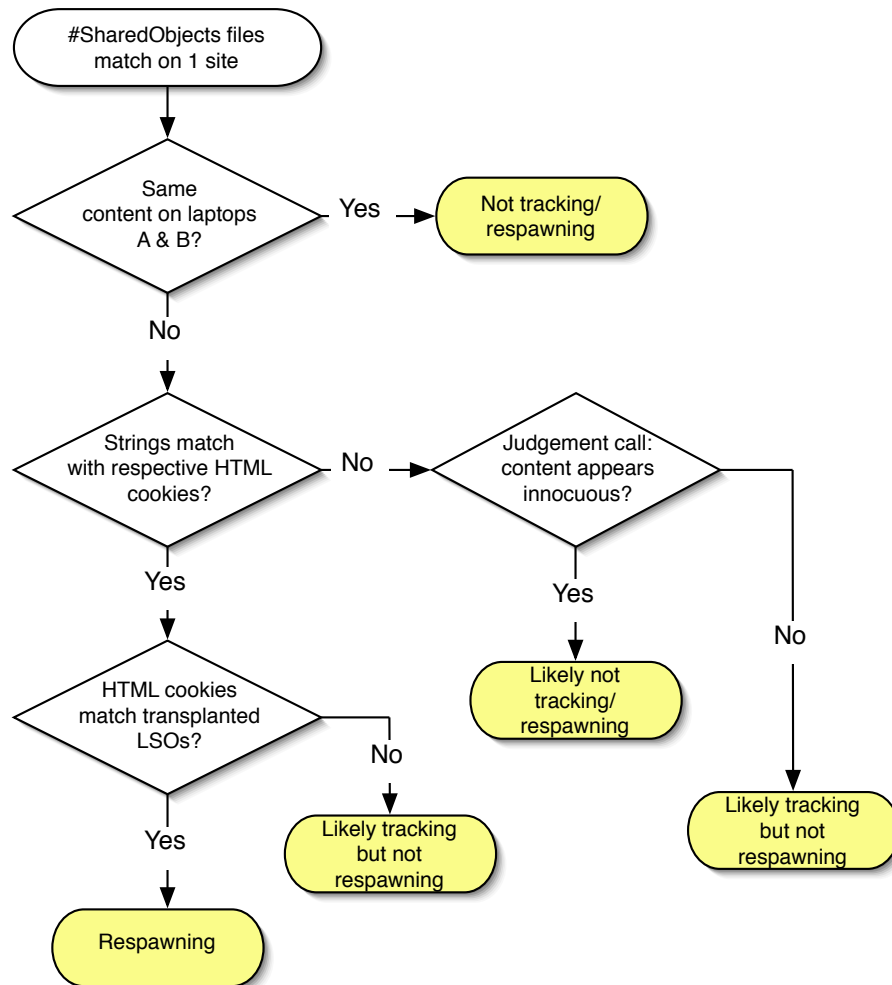


Figure 7.1: Flow chart of website classification based on SharedObjects

using LSOs. We also only reported cookies saved, not cookies set: we logged several sites that set LSOs but then deleted them. Transient LSOs can not be used in uniquely identifying users over time or for respawning, so they are not our concern. Finally, we turned on popup blocking in the browsers to reduce caching issues, which could also undercount the prevalence of LSOs from blocked popups.

7.4 Results

As described in the Methods Section, we visited each site in three “sweeps” for a total of nine times:

- Sweep 1, three visits from laptop A
- Sweep 2, three times from laptop B
- Sweep 3, three times on laptop A with the Flash cookies from laptop B

For quantifying HTTP cookies on sites there was no advantage to any sweep in particular; we arbitrarily chose to report results from the final sweep. We did see a small variation between sweeps, for example HTTP cookie use ranging from 92% to 95% of the 2009 top 100 sites. Similarly, we report statistics from the final sweep in the discussion of the Flash activity in the `sys` directory. In our discussion of the `SharedObjects` directory we contrast sweep 1 and sweep 2, and then check against results from sweep 3 to look for respawning.

7.4.1 Use of HTTP Cookies

Cookies are ubiquitous. Of the 2010 top 100 sites, only two never used cookies (`wikipedia.org` and `craigslist.org`). Similarly, 95% depending of the 2009 top 100 sites use cookies. Cookie use drops to only 59% of the random 500 sites.

Not only do fewer randomly selected sites use cookies, they also set fewer cookies than popular sites. We used Internet Explorer, which stores cookies in text files grouped by the name of the domain that set them. For example, the list of cookie files from a popular site might look like this:

```
cupslab@ad.yieldmanager[2].txt
cupslab@www.yahoo[2].txt
cupslab@doubleclick[1].txt
cupslab@yahoo[1].txt
cupslab@voicefive[1].txt
```

Here we see five different domains that set cookies (`ad.yieldmanager`, `doubleclick`, `voicefive`, `www.yahoo`, and `yahoo`). There is some overlap here — `www.yahoo` and `yahoo` are from the

same company. But as is the case in this example, in general the number of domains setting HTTP cookies is roughly equal to the number of different entities setting HTTP cookies on the computer.

The contents of an HTTP cookie might include something like this:

```
fpmS
u_30345330=%7B%22lv%22%3A1279224566%2C%22uvc%22%3A1%7D
www.yahoo.com/
1024
410443520
30163755
2720209616
30090329
*
fpps
_page=%7B%22wsid%22%3A%2230345330%22%7D
www.yahoo.com/
1024
410443520
30163755
2720209616
30090329
*
```

This is a snippet that shows two different HTTP cookies, *fpmS* and *fpps*, served by Yahoo. In Internet Explorer's implementation each cookie file may contain multiple cookies separated by asterisks.

As we summarize in Table 7.2, we found an average of 6.7 files for the 2010 top 100 sites, 5.8 for the 2009 top 100 sites, and 2.5 for the random 500. We observed a maximum of 34 different cookie files on the 2010 top 100 sites. We observed a maximum of 24 with the 2009 top 100, and 30 with the random 500. Users might be surprised to learn that a visit to their favorite site registers with dozens of different entities.

Table 7.2: HTTP Cookies

Data set	% sites with cookies	Avg. # domains	Max. # domains
2010 Top 100	98%	6.7	34
2009 Top 100	95%	5.8	24
Random	59%	2.5	30

7.4.2 Use of LSOs

69% of the 2010 top sites had some LSO activity, by which we mean at least created a directory to store Flash cookies, even if they never actually wrote any files. Similarly, so did 65% of the 2009 top sites, and half as many of the random sites with 33%. There are two directories that LSOs may be written to, `sys` and `#SharedObjects`.

`sys` directories

68% of the 2010 top sites wrote Flash cookies in the `sys` directory, as did 65% of the 2009 top 100 sites, and 33% of the randomly selected sites. Flash cookies in the `sys` directory are highly structured. Every website writing to the `sys` directory had a `settings.sol` file at the top level, which contains standard variables and settings for them, for example setting `defaultklimit` to 100, or `allowThirdPartyLSOAccess` set to true. There is nothing in any top level `settings.sol` file that suggests these files are used for tracking or re-spawning.

43% of the top 2010, 33% of the top 2009 sites, and 10% of the randomly selected sites have additional `settings.sol` files, usually named after the site itself. These files are also innocuous, containing fewer variables and settings. There is nothing that appears to be used in tracking or re-spawning in the `sys` directories at this time.

`#SharedObjects` directories

20% of the 2010 top 100 sites stored Flash cookies in the `#SharedObjects` directory, as did 16% of the 2009 top 100 sites and 8.2% of the randomly selected sites. These are the sites we are most interested in as sources of either respawning HTTP cookies due to LSOs, or as ways to individually identify users. We discuss these in more detail below.

7.4.3 Mismatched Sites

We compared the contents of Flash cookies in `#SharedObjects` directories on two identically configured laptops. However, we did not always find identical files on both laptops. For example, one site contained two Flash cookies on Laptops A and B, but contained an additional two Flash cookies just on Laptop B. In cases like this we could and did compare the Flash cookies in common, and in some cases could make some good guesses about tracking in the mismatched Flash cookies, but we could not follow the protocol we outlined in the methods section. Our results are necessarily more subjective in this section.

The 2010 top 100 sites included 20 sites with `#SharedObjects`, of which six did not have matching file names. Of these six, we believe four are individually identifying. The 2009 top 100 sites included 16 sites with `#SharedObjects`, of which four did not have matching file names. Of these four, we believe two are individually identifying. The random 500 sites include 41 sites with `#SharedObjects`, of which nine did not have matching file names. Of these nine, we believe five are individually identifying.

Why do we see so many mismatches between the two laptops? First party `#SharedObjects` remained stable. Third party `#SharedObjects` come from advertisers, and advertising rotates. Even though we collected data on both laptops only a few days apart, advertising — and advertising partners — can change over the course of a few minutes.

7.4.4 Prevalence of Unique Identifiers and Respawning in LSOs

We found paired LSOs with matching file names on 14 of the 2010 top 100 sites, 12 of the 2009 top 100 sites, and 32 of the random 500 sites. Tracking requires unique identifiers. Any LSO that set identical content on both laptops could not use that content for tracking or for respawning. For example, we found a variable named `testValue` set to `test` on both laptops, which is not cause for concern. Because `test` is identical on both laptops, it cannot be used to distinguish computers on future visits. Not all unique identifiers are used for tracking, but all tracking via LSOs requires a unique identifier. We found matching content on both laptops for six of the 2010 top 100 sites, four of the 2009 top 100 sites, and twenty of the random 500 sites. These sites are neither tracking nor respawning. See Figures 7.2, 7.3, and 7.4 for details.

We found a few sites where there were differences in content on the two laptops, but we did

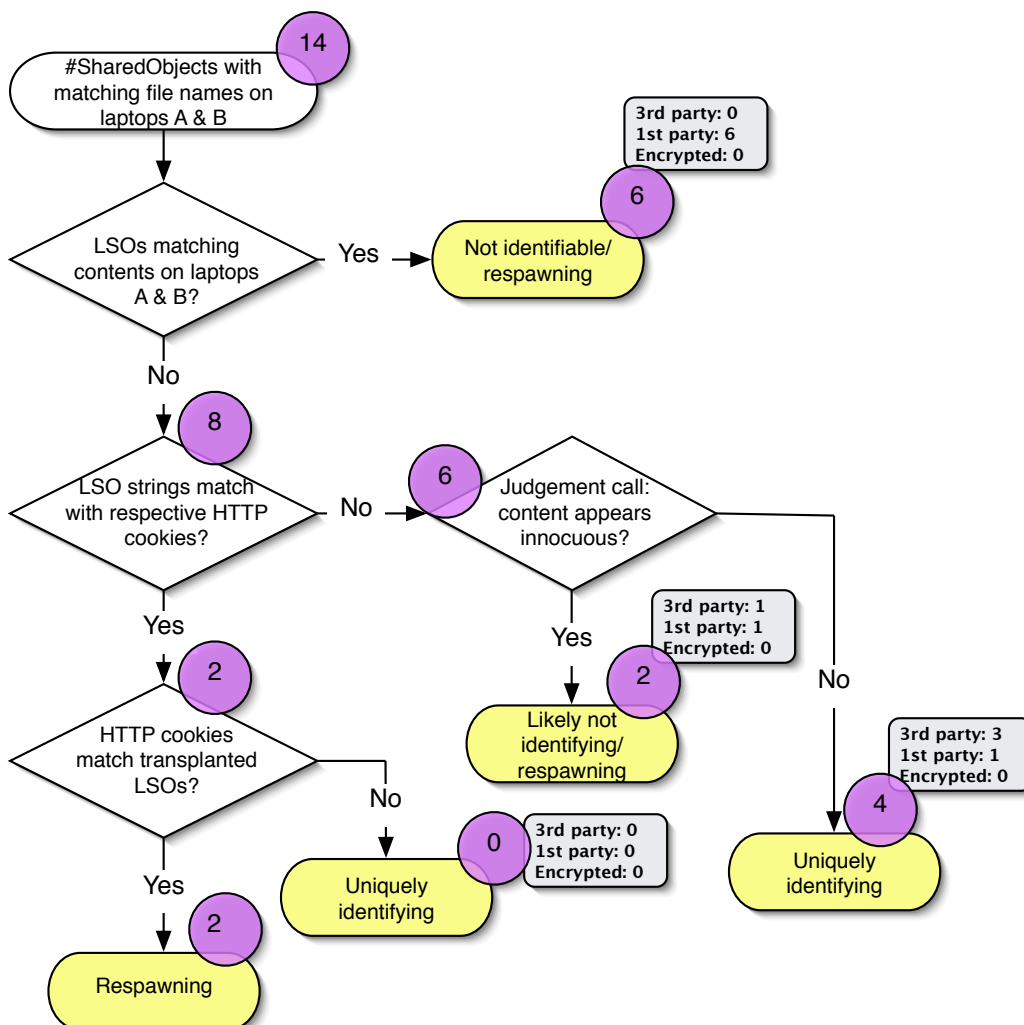


Figure 7.2: Analysis of the 100 most popular websites in 2010. Purple circles contain the number of sites that fall into a given category.

not believe the LSOs were used for tracking. For example, one LSO contained a Unix epoch-formatted time stamp with millisecond precision. It could be used for tracking — the odds of collisions with multiple users visiting a server at the same millisecond are low. But we expect it is just an exceedingly precise time stamp. This is a subjective judgement. We classified two LSOs in the 2010 top 100, one LSO in the 2009 top 100, and one LSO in the random 500 as innocuous.

Variable names like `userId` helped us theorize that many LSOs are used to identify users, rather than identifying creative content, but without knowledge we cannot conclude why LSOs contain unique identifiers, only to quantify how many do. We further investigated to see if con-

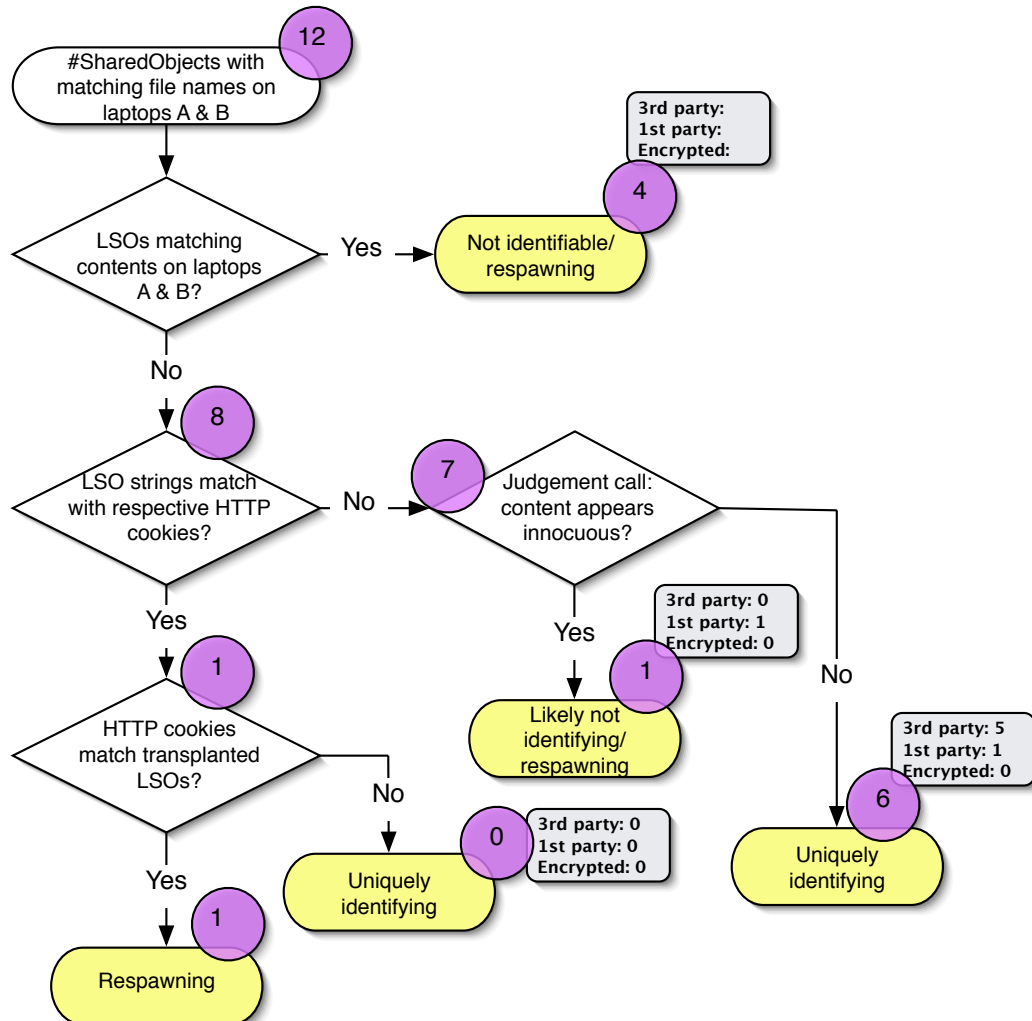


Figure 7.3: Analysis of the 100 most popular websites in 2009. Purple circles contain the number of sites that fall into a given category.

tent in LSOs matched content in HTTP cookies. If so, we performed analysis to see if respawning occurred, where LSOs are used to reinstate data after a user has deleted an HTTP cookie. For example, we found one LSO that contains a variable named `uID` set to a unique a 10 digit integer. After we deleted all HTTP cookies and migrated LSOs from one laptop to the other and then revisited the site, the same 10 digit integer now appears in the new HTTP cookies. These are clear-cut cases of respawning.

In the 2010 top 100, we found four LSOs with unique identifiers that did not match HTTP content, and two that respawned. In the 2009 top 100, we found six LSOs with unique identifiers

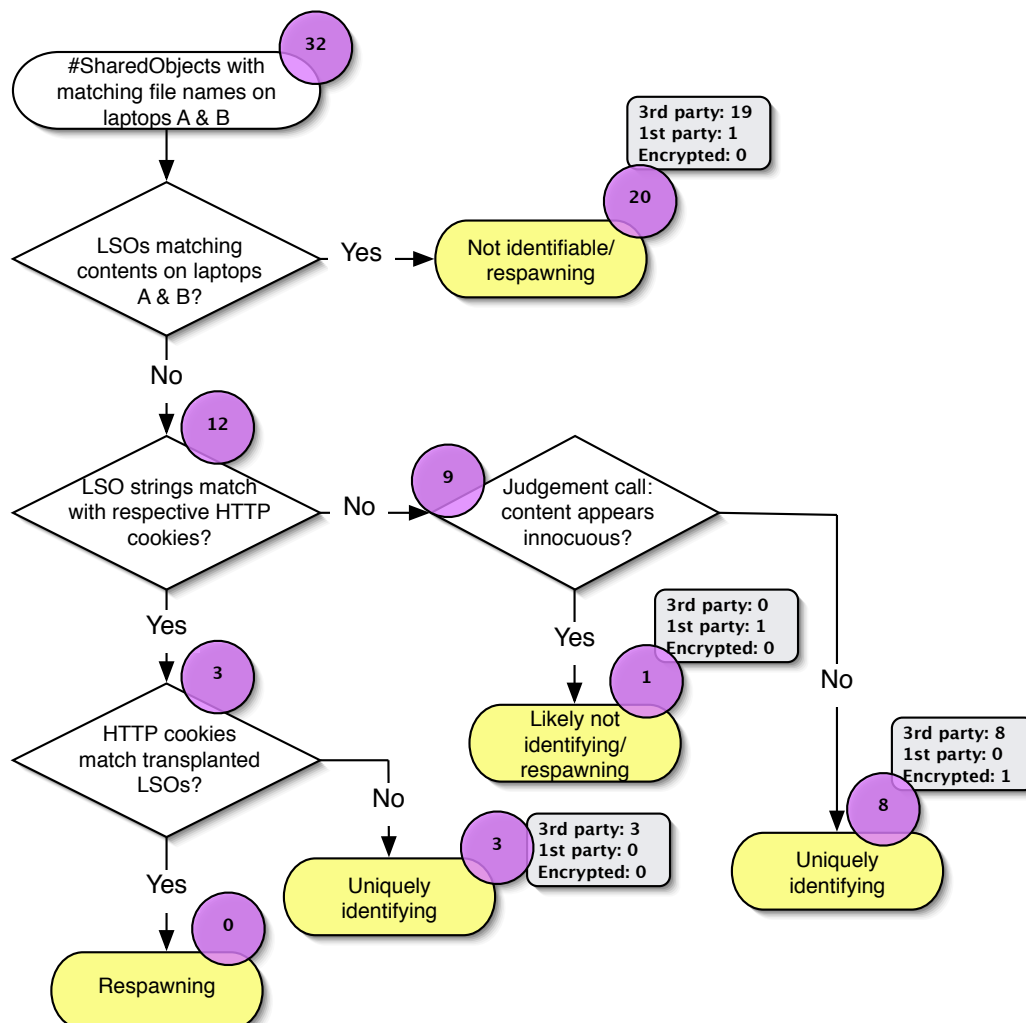


Figure 7.4: Analysis of the 500 randomly selected websites. Purple circles contain the number of sites that fall into a given category.

that did not match HTTP content, and one that respawned. In the random 500, we found eight unique identifiers that did not match HTTP content, three that did match HTTP content but did not respawn, and no respawning.

7.5 Discussion

With the methodology we used, it appears respawning is less substantially prevalent now than it was a year ago during the first Soltani paper, or even half a year ago for data collected in the

Wall Street Journal series. However, we should highlight a few ways in which we cannot directly compare results:

- We visited just the top level of a domain. The Soltani work loaded 20 pages per site. It is possible that additional interaction would find more respawning.
- We did not interact with Flash content.
- We relied exclusively on files saved to disk, and did not analyze network traffic.

Essentially, we took a wholesale approach investigating 630 sites while the Soltani research took a deeper look at 100 or 50 sites per study. Our approaches have different strengths and weaknesses, and we recommend looking at them as complementing each other.

While we found only two documented cases of respawning, unique identification remains vibrant. Respawning allowed software engineers to easily reuse existing code. With a little more work, a company using LSOs for unique identification can save the exact same data on their servers they would have with respawning. There is no functional difference: they still use LSOs to subvert users' attempts to manage cookies for privacy.

Unique identifiers usually came from third-party advertisers, especially for the random sites where we saw 100% of unique identifiers were third-party. 7% of web users delete Flash cookies, which sounds low, but leads to over-counting ad views by 25% — and this is substantially more reliable than the 30% deletion rate for HTTP cookies [27]. Advertisers may argue that tracking via Flash cookies is the same as tracking via HTTP cookies. Clearly this is not the case: user education lags for Flash cookies, tools are not as developed, and the furor in the press demonstrates user surprise. Moreover, the whole reason advertisers use LSOs for tracking is because HTTP cookies are not the same: if they were truly the same, advertisers should have no objection to exclusively using HTTP cookies for tracking and analytics.

One of our challenges with data analysis is that we received different cookies on the same site at different times, often because advertising rotates over time. Our experience with rotating advertisement is not unique to sites using Flash cookies and highlights an interesting problem. If a user visits a site, say the New York Times, it is possible her data will only go to the New York Times. But then reloading the site, with a new advertising partner, could create an entirely different set of data flows to tens of companies, and link to previously and subsequently collected

data in entirely different ways. Yet users are supposed to understand their privacy risks by reading privacy policies, which remain static regardless of the actual data flows occurring. Privacy practices are so fluid that first party sites frequently have no idea what their partners are doing — or indeed, who their partners are, as ads are subcontracted out to the highest bidder. Privacy policies do not actually reflect how data is collected and where it is used in practice. Can users make informed decisions in this environment?

Finally, as an observation, it is difficult to find calls for a purely industry self-regulation to Internet privacy credible when industry demonstrates such willingness to violate user intent and privacy. No malice is required anywhere in the chain: it is easy to imagine advertising engineers using a clever tactic to avoid data loss with an esoteric data store designed by a company prior to Adobe's purchase in 1996, long before the Internet was a household word. But the effects on user privacy are the same regardless of how decisions are made.

7.6 Recommendations

We have several recommendations to reduce Flash cookie abuse. Two are comparatively easy to implement, and are in Adobe's hands. As an immediate stop-gap measure, we suggest Adobe improve the user interface on the Flash opt out page.¹ As one journalist wrote, "the controls are so odd, the page has to tell you that it actually is the control for your computer, not just a tutorial on how to use the control" [128].

Second, Adobe could add clear language to their terms of service barring Flash cookie use for any data that is not necessary for Flash components. Those who violate the terms of service could lose their developers' licenses. Since it appears large advertisers are the ones most at issue, this could be a highly effective measure.

Cached Flash cookies pose a threat to privacy. Occasionally we quit Internet Explorer, deleted the directories containing Flash files, yet the next time we opened Internet Explorer we had cached content from the previous site we visited save to our disk. Even with pop-up blocking and quitting the browser between each site we visited we still had cached data persist about 6% of the time. We did not test any PETs or any in-browser tools, but we expect cached data gets past

¹http://www.macromedia.com/support/documentation/en/flashplayer/help/settings_manager03.html

them and users who attempt to delete Flash cookies are not always successful in doing so. We recommend a well-documented method to flush all Flash cookies, and recommend PETs authors invoke it prior to clearing Flash cookies.

Browser makers could dramatically improve privacy by effectively turning Flash cookies into session cookies, and providing better user controls. However, unless users know to look for settings, this only helps a very small minority. Asking browser makers to expending engineering resources for problems they did not create seems unsatisfying, but they do have the ability to improve user experience. Flash cookies are only one of many types of tracking technologies and browser vendors may need to keep adjusting to prevent new approaches from being used to track users without users' knowledge. However, since major browser makers also own advertising networks, they may lack incentives to act.

Finally, since a comparatively small number of sites use #SharedObjects LSOs, it might be worth making just those variety of Flash cookies opt-in by default. This would require substantially more effort, and opt-in is not a panacea, as it can be difficult for users to understand what they are being asked to do or to undo decisions later.

Chapter 8

Policy Options

Should the United States continue the current policy of industry self-regulation for online privacy? In the sections that follow we consider three different possible approaches: retaining the status quo (abbreviated as **S** below), modifying self-regulation (**M**), or replacing self-regulation (**R**). We present arguments on both sides of these approaches from proponents (**P**) and opponents (**O**). We explore how the research in this dissertation contributes to understanding these issues.

8.1 Status Quo

The first option we consider is continuing the policies already in place: notice and choice, with FTC action against transgressing companies. We have heard or read each of the arguments for and against that we highlight below. These are not exhaustive lists, which could fill several books, but rather an attempt to capture some of the most familiar arguments at this time. Most debate centers on whether we should keep the current approach.

8.1.1 Proponents

Some of the arguments from those in favor retaining the status quo include:

- **SP₁**: The status quo is fine as it is, as demonstrated by the millions of people who use the Internet. There is nothing broken so there is no need to fix anything.
- **SP₂**: A libertarian, hands-off approach to the Internet is the reason why it has thrived. Regulations would stifle innovation, and by the time legislation can be enacted it is already out-of-date.
- **SP₃**: We need more time for self-regulation to work. FTC and industry guidelines for behavioral advertising are not even a year old.
- **SP₄**: Privacy is a harm. Protecting privacy causes damage in one (or more) ways: by reducing efficiency in the market place; by alienating people who would learn they have no need to be embarrassed if only they knew their neighbors are just as human as they are; by fostering anti-social or illegal behaviors that thrive in secrecy.
- **SP₅**: Privacy is dead. Attempting to regulate online privacy is a naive waste of time and resources that could better be applied to understanding and navigating a post-privacy world.

- **SP₆**: Privacy is only a problem for old people. Rather than addressing concerns of a shrinking minority, it is effective to just wait for older generations to die off.
- **SP₇**: Privacy preferences are so individual that there is no way to pass laws that suit everyone. The best approach is to build Privacy Enhancing Technologies (PETs) and empower users to implement their own decisions.

8.1.2 Opponents

Some of the arguments for changing the status quo include:

- **SO₁**: Notice and choice only works if users have the knowledge, ability, and willingness to engage in informed consent. This has not been the case in practice.
- **SO₂**: Industry self-regulation has fallen into disrepute with examples of problems due to lack of regulation including Enron, Lehman Brothers, and Deepwater Horizon. Privacy issues are largely invisible, and trusting corporations has not worked out well in other contexts.
- **SO₃**: The status quo is causing market failures due to information asymmetries. For example, Microsoft believes they could find more customers for their Health Vault product if only they were able to convey their privacy and security practices.
- **SO₄**: Public outrage when people learn about practices is one sign that people are surprised by and do not accept current data practices. Backlash against Facebook's Beacon is one example.
- **SO₅**: A philosophical or ethical view that privacy is a basic right, meriting full protection. Treating what is sometimes exceedingly personal data as a commodity is at odds with human dignity.

8.1.3 Research

Our research does not address all of the views above. However, we can contribute in several areas. Our user studies on behavioral advertising (Chapters 5 and 6) argue against **SP₁**, that Internet use equates to a lack of problems. Although millions of people use the Internet, that does not mean

they understand or endorse current data practices. Instead, we found a mix of views. A subset of users welcome relevant advertising, though sometimes due to the mistaken belief that their data is protected by law, as others have also found [67]. We interviewed one participant who described withdrawing from online life due to privacy concerns and 40% of participants in our behavioral advertising study said they would be more careful online if they knew advertisers were collecting data. 15% in the behavioral advertising study and 11% in our study contrasting advertising types self-reported that they would not use sites with targeted advertising. This echoes findings in 2003 that half of users self-reported they would prefer to stop using sites with targeted advertising rather than paying for privacy or allow data collection [147], and a study in 2009, finding that over two-thirds of Americans would prefer not to have targeted ads [148]. We are concerned that advertising can create a chilling effect online, and may reduce the desirability of broadband access at the same moment we are investing public funds to build out additional infrastructure. We only studied current Internet users so we do not know if privacy concerns are slowing the rate of Internet adoption. If so, if those concerns may not match actual practices, as we saw with people who clear cookies in a misguided attempt to avoid spam (Chapter 5). How views of Internet privacy affect adoption rates would be an interesting area for further study.

SP₂, the claim that regulation is too slow and law makers know too little, contains a tacit assumption that therefore industry moves quickly and with knowledge. There are certainly times this is true, and any approach that can harness industry's ability to innovate has a tremendous advantage. However, our research touches upon examples where industry has responded slowly and with the clunky touch of design by committee. In Chapters 3 and 4, we studied formats built upon P3P, and noted that Internet Explorer makes cookie decisions based on P3P. Few people have any idea that Internet Explorer uses P3P Compact Policies: the settings are buried deep within the IE user interface [74]. Most users simply inherit default settings and do not realize they could configure their browser settings to match their personal privacy preferences. Furthermore, IE does not implement full P3P parsing, only compact policies dealing with cookies. Microsoft technical support web pages even advise P3P Compact Policy authors to use policies that do not reflect their practices as a work-around for an iFrame bug in IE, and many Compact Policies are crafted in ways that bypass IE controls [88]. As another example of usability issues with industry solutions, in Chapter 5 we find that 89% of respondents do not understand the text description

of the NAI Opt-Out page. And no wonder: different member companies have different practices. Opting-out of behavioral advertising means, variously, that all tracking stops, or tracking moves from an individual level to aggregate data with others who have opted out, or that tracking is completely unchanged. We did not even test the mechanism of opt-out cookies, whereby users must accept cookies and not delete them in order to eliminate seeing behavioral ads based on cookies. Again, this is hard to describe to users because it is a rather counterintuitive idea, and on the surface makes all the sense of clicking the “Start” button to shut down a Windows computer. Could Congress do worse? Quite possibly: one current proposal is a cookie-based “do not track” list, where once again we have a *Through the Looking Glass* suggestion that users must register themselves with advertisers in order to avoid tracking.

Our research does not directly address the conflicting arguments in **SP₃** and **SP₅**, that we need to give privacy self-regulation more time to work or that privacy is already dead with nothing further to do. We found user perceptions that some advertising methods are concerning, with privacy and security ranking among the top ways sellers could attract more purchases. Users appear to believe there are differences between sites, and have not yet given up on the idea of online privacy. However, our work on Flash cookies (Chapter 7) is only one recent example of research showing technologies that allow advertisers to collect data in ways consumers do not understand [57, 45]. Users still struggle to understand first-party cookies and do not understand the rich and broad data collected about them. Online privacy may not be dead, but it is diminished beyond what users’ perceive.

Similarly, we cannot speak directly to **SP₄** or **SP₆**, that privacy is a harm or only of concern to old people, but we see no evidence to suggest that is how users perceive things. We heard quite a bit about privacy invasion as a harm, or about advertisements as a harm (including a subset of people who wished for better-targeted ads). We never heard anyone express the view that they had too much privacy or that society would benefit from a reduction in privacy. Our user studies skew toward younger populations, and we do not have statistical significance to contrast to, say, people over 50 years of age. However, this means that if theories of privacy concern dying off over time are true, there is a long wait for the demise of the primarily 20- and 30-year-olds we studied. It would also suggest even more concern in older populations than we observed with younger participants. Other studies find comparatively little difference in privacy values between

age cohorts with high levels of privacy concern among young people [68, 87].

Our research supports the basis for **SP₇**, the notion that privacy preferences are not monolithic. We found a wide variety of preferences for behavioral advertising (Chapters 5 and 6). If there were a uniform, optimal level of privacy we could make a bright line rule and be done, but our research finds an even split between people who welcome the advantages of targeted advertising and those who very much want to avoid it as a harm, with a middle group who are uninterested in providing data to advertisers because they see no advantage to “better” ads when they prefer to ignore ads in the first place. Moreover, even individual users have nuanced preferences, for example, yes to targeted ads but only when not searching for gifts for family members who use the same computer. PETs sound like a fantastic solution. However, there are a few important obstacles to PETs. First, designing usable PETs is hard, requiring a mix of technical and design skills, plus an understanding of mental models of privacy (see Chapter 5 for discussion of some misperceptions, and Chapters 3 and 4 for attempts to communicate privacy information that we found did not work well) [66]. Second, funding usable PET design is difficult. For-profit models like TRUSTe are met with suspicion, academic funding is currently scarce, and open source volunteer efforts rarely attract people with design and interaction skills. Third, PETs often require cooperation from the technologies they attempt to protect against, which can hinder adoption. Fourth, most PETs require educated users. At the very least, users typically have to be aware there is an issue and take proactive measures to enable PETs. Finally, PETs become a privacy arms race. There have been gaps when PETs did not exist to meet privacy threats, as we mention in Chapter 7, so relying on PETs includes an acceptance that there may be periods of inadequate protection. When PETs work, they can be transformative: TOR, built on US military technology, helped Iranian protestors get their voices heard around the world while reducing risk of death at home. IE’s adoption of P3P compact policies changed the privacy landscape even without users being aware of how IE handles third party cookies. The Tsai et. al. work on providing salient privacy indicators during search showed people will even pay a premium for privacy, if information is conveyed in a useable way [146].

Each chapter of this thesis in some way supports **SO₁**, the assertion that notice and choice are not working in practice. In Chapters 3 and 4, we found that even when people are instructed to read privacy policies, they dislike the process and have a difficult time answering basic questions

about how their data are collected and used. This is congruent with studies that show participants did not understand when terms of service informed them they were about to install malware [58] and many studies that highlight how difficult it is to use privacy policies for decision making [59, 73, 115, 123]. Privacy policies also take a long time to read, which may itself be a barrier to usability (see Chapter 2). The good news from this is that subsequent privacy policy formats are now improved [80, 77]. Since notice is unlikely to go away, even if self-regulation undergoes serious change, it is worthwhile to seek improvements to privacy policy formats. However, people misunderstand the phrase “privacy policy” and believe it means they are protected [147] which may reduce their incentives to read privacy policies. Further, as we discuss in Chapter 2, if users read the privacy policies for each site they visit annually, they would spend about as much time reading policies as they currently do visiting websites. Even reducing the time to read privacy policies by an order of magnitude may be too much of a time burden on users. If users do not read policies and make decisions based upon them, the notice and choice framework fails. Reading is not sufficient. As we detail in Chapter 5, at this moment users are not clear on how even first party cookies work, confuse cookies and history, and make choices to clear or keep cookies based on misunderstandings, thereby not enacting their privacy preferences. This strongly suggests readers lack the background knowledge required to understand privacy policies, and suggests a need for education if we are to retain the notice and choice approach. A majority of users believe online privacy is a right, which may also undermine their willingness to read policies (see Chapters 5 and 6). Finally, we find industry self-regulation attempts do a poor job of communicating with users (Chapter 5) and we confirm companies are using Flash cookies to subvert attempts at user control of privacy (Chapter 7).

We have no research-based response on the efficacy of self-regulation, **SO₂**. Industry practices such as using Flash cookies to track users (Chapter 7) and setting P3P compact policies to avoid IE cookie management [88] create the impression that industry does not take self-regulation seriously. However, the FTC may take action in the future, which would be in keeping with the enforcement aspect of self-regulation.

We also have no research to establish or refute **SO₃**, that information asymmetries are causing market failures. We document some of the difficulties reading privacy policies (Chapters 2, 3 and 4) as well as misconceptions and lack of basic knowledge (Chapter 5). While we can establish

some of the conditions required for SO_3 , we have no counterfactual to study.

User surprise, SO_4 , does surface in our user studies (Chapters 5 and 6). Due to reactions of disbelief in lab studies, we presented common scenarios as hypotheticals. We found that most participants understand behavioral advertising is happening now or can happen in the future, though only 18% welcome behavioral ads. 40% believe there are ads based on email content, but 29% think such a thing would never happen. Only 9% think it is ok to have ads based on email as long as the service is free, yet 41% have Gmail accounts. Our research does not establish any causal links between user surprise and potential user rejection of technologies, but we do see reduced acceptance for email-based advertising and for less well-known types of targeted ads like DPI-based.

We find many people do see online privacy as a right, SO_5 , and they may be less likely to take affirmative steps to protect their privacy because they already believe it is protected by law (Chapter 5). How people perceive privacy may contribute to a debate of whether privacy *should* be a right, but not whether it is or is not a right, which our research does not address.

8.2 Modify Self-Regulation

As an alternative to continuing the current self-regulation approach, some suggest keeping the same general framework but attempting to improve specific aspects.

8.2.1 Proponents

Some of the arguments from those in favor modifying self-regulation include:

- MP_1 : Companies innovate when threatened with regulation. There is no need to actually pass new laws, since a climate of legislative uncertainty provides ample motivation. New laws would reduce incentives for innovation by establishing exactly what is minimal required.
- MP_2 : Privacy policies are the right approach, but we should mandate better formats for privacy policies.

- **MP₃**: Redefine success: rather than expect individuals to read privacy policies, instead rely on a few people to read policies and bring problems to public attention.

8.2.2 Opponents

Generally the same critiques of the status quo carry over into discussions of marginal changes, plus we have heard one rebuttal:

- **MO₁**: Data breach notification was supposed to change user behavior, but instead the pervasiveness of data breach convinced people they are powerless and have no better alternatives. Similarly, improving privacy disclosure will not change user behavior.

8.2.3 Research

Our research does not investigate how the pace of privacy innovation changes in response to laws, **MP₁**. This would be an interesting area for future work, perhaps by looking at changes in response to state privacy laws, both after they were enacted and during consideration.

Our research does not speak to whether privacy policies are the “right” approach, **MP₂**. We did find that some formats designed to improve usability failed to do so (Chapters 3 and 4). We again refer to other work that shows promise in this area [80, 77]. We offer the following three suggestions, borne from observation rather than research. First, if formats are to be mandated, it is better to test them for usability issues first rather than after the fact as happened with financial notices [119]. Second, users may have already been “trained out of” clicking the privacy policy link, aware the link leads to painful legalese. Something similar has happened with online help systems, with users ignoring help buttons just as they would skip over ads. It would be interesting to perform eye tracking studies to see if users notice privacy policy links. Finally, one of the most contentious and useful functions of standardized formats is to highlight specific information and de-emphasize the rest. The things people are interested in learning have already changed as technology changes. For example, there were no social media sites at the time P3P became a W3C specification. We recommend that if formats are mandated, that mandate include a process of periodic review and updates.

We find it novel to think of the free rider problem as a feature, as with suggestions to depend on a small number of readers to flag problems, **MP₃**. It might be interesting to estimate how

many people read policies, understand where there are problems, and have the time and ability to communicate those findings broadly. We have no research insights as to how practical this approach is, though we have seen it work in practice upon occasion.

While privacy breaches do not convince users to migrate en masse to different vendors, **MO**₁, others have found they do have small negative effects on breaching companies [4]. Perhaps more importantly, data breach laws may affect companies' internal practices [121]. This suggests two areas of interesting future research: first, are there any consequences for companies named in Flash cookie lawsuits? We find that much Flash tracking came from third party sites (see Chapter 7) so even an organized consumer boycott would not have an effect, but first party sites could cut ties with accused advertising networks. Second, how, if at all, does the act of writing a privacy policy change internal corporate practices? It may be that privacy policies are exceedingly valuable, just not as mechanisms of consumer notice.

8.3 Replace Self-Regulation

As a third broad approach, some call for an end to the industry self-regulation approach by creating new laws or regulations. Those who wish to end self-regulation believe it has failed to protect privacy, or that is unlikely to provide sufficient protection in a reasonable time frame. People who believe self-regulation is working, or who think privacy is not something society needs to protect, tend to oppose the imposition of new protections.

8.3.1 Proponents

Possible new approaches could stem from:

- **RP**₁: Regulation from the FTC. The FTC has the institutional history and staff to best understand current data practices, and is empowered to protect consumers.
- **RP**₂: Legislation from states. State laws enacted to protect online privacy can lead the way on privacy, as they have with other online topics like spam. Creating a patchwork of laws can lead to benefits from quick evolution to a model law. Multiple laws can create pressure for a preemptive federal law.

- **RP₃**: Legislation from Congress. The Internet is unquestionably involved in interstate commerce and it makes sense to have unified, national laws. If industry power is too strong to create meaningful legislation, it is still worthwhile to have something weak in place that can be readily modified during the next legislative window.

8.3.2 Opponents

Most of the arguments in favor of the status quo are also arguments against new regulation. We do not repeat them here.

- **RO₁**: Any legislation limited data use is unconstitutional due to free speech protections.
- **RO₂**: The FTC lacks the authority to act.
- **RO₃**: Privacy laws may create a “ceiling” instead of a “floor” — once companies learn the minimum they must do for privacy protections, there will be no incentive to do more.
- **RO₄**: Uncertainty on an optimal level of privacy means it is premature to craft new laws.

8.3.3 Research

Our research does have implications for specifics in various legislative and regulatory proposals, but little to say about the advisability or viability of new laws. As discussed, we find many flaws with the status quo, but that does not mean legislation or regulation as contemplated in **RP₁ – RP₃** would therefore necessarily be an improvement. Similarly, we have no research-based insights about the likelihood of courts upholding legislative or regulatory authority as in **RO₁ – RO₂**, or about the likely effects of law as a “ceiling” or “floor” as in **RO₃**. We do note other research that finds within companies privacy concerns are not limited to only following current law, and we have better corporate policies as a result [16].

We have less concern over uncertainty about privacy preferences, as in **RO₃**, than we do with our finding that different people want different things (Chapters 5 and 6). It may not be economically viable to have advertising networks that exclusively work with data from the roughly 20% who want targeted advertisements. Economic analysis for the long-term economic viability for advertising networks, advertisers, and first party sites hosting behavioral advertisements would be an interesting area for future work.

8.4 The Road Forward

Courts see privacy as a boolean: something is secret or it is known. Our study participants have a different concept. People distinguish between at least three categories of secret, published, and a middle area where knowledge is held by a trusted few or even a community. Internet privacy runs headlong into the difference between these conceptions of privacy.

It seems most likely that we will have a patchwork quilt of privacy regimes on top of our patchwork quilt of some data types protected more than others: self-regulation based around privacy policies and FTC enforcement, *plus* new regulations, *plus* new legislation, *plus* a race between privacy enhancing technologies and privacy invading technologies. More and different actors are likely to get involved in setting policy: the FCC due to the broadband plan, the Department of Commerce via the NTIA, the military's interest in civilian Internet security may implicate privacy issues, and so forth. Internet privacy has international dimensions, as seen when the State Department found itself issuing statements about Google's decision to turn off search filtering in China in retaliation for Chinese efforts to gather information from dissidents' Gmail accounts. Internet privacy is likely to become more complicated as more interested parties seek ways to influence policy.

P3P received great attention as an example of free-market innovation that demonstrates the lack of need for regulation, yet government sites have substantially higher adoption rates than corporate sites [40]. Unlike industry, federal and many state government websites have legal requirements to provide privacy policies in machine-readable format. This illustrates one case where a mix of corporate innovation plus a legal requirement was more effective than just a purely free market solution. Adoption of technological solutions appear to require either market incentives or mandates. With consolidation, most of the stakeholders in a position to enact privacy protections have financial incentives not to. A recent account details how Microsoft diminished and eliminated new privacy features in IE after acquiring the aQuantive advertising network for \$6 billion dollars [156]. One possible hybrid approach is to mandate adoption of mature PETs, since it appears adoption lags if it is purely optional.

In this thesis we have established several areas where industry self-regulation is failing to meet user's preferences for privacy. How can online privacy needs be met? That is an area for future work — and will likely be several people's life work.

Appendix A

Statistical Tests for Online Privacy Policies and Formats

This appendix expands upon Chapter 4 and includes supporting statistical details.

A.1 Statistical Significance

As mentioned in Chapter 4, we performed all tests of statistical significance at the $\alpha = 95\%$ confidence level. We performed ANOVA analysis for both time data and psychological acceptability, which we recorded on a seven point Likert scale and treated as continuous variables. Accuracy questions were categorical data (either accurate or false) so we used Chi Squared tests. Details of that analysis follows.

A.1.1 Accuracy

As mentioned in Chapter 4, accuracy scores are all reported as the percentage of people who answered the question correctly. Answers are always either Yes, No, or the policy Does Not Say (DNS). In some cases participants may have been confused about when to use Does Not Say, so we also reported and tested statistical significance for the combined percentage of participants who answered correctly with those who answered Does Not Say for all questions except the question on cookies. We tested for statistically significant differences in mean accuracy rates by company (Table A.1) and by format (Table A.2).

Table A.1: Statistical Significance Tests for Accuracy Questions by Company

Question	d.f.	χ^2 value	p	Significant?
Cookies	5	12.16	.033	✓
Opt Out Link	5	108.31	< .001	✓
Opt Out Link with DNS	5	53.44	< .001	✓
Share Email	5	22.43	< .001	✓
Share Email with DNS	5	37.05	< .001	✓
Telemarketing	5	24.99	< .001	✓
Telemarketing with DNS	5	44.34	< .001	✓

A.1.2 Time

We recorded time in milliseconds and reported it in minutes to assist readability. With such a fine grain unit of measure time is nearly continuous and we used ANOVA for analysis. We tested

Table A.2: Statistical Significance Tests for Accuracy Questions by Format

Question	d.f.	χ^2 value	p	Significant?
Cookies	2	28.95	< .001	✓
Opt Out Link	2	40.80	< .001	✓
Opt Out Link with DNS	2	53.44	< .001	✓
Share Email	2	1.90	.387	
Share Email with DNS	2	0.20	.903	
Telemarketing	2	50.08	< .001	✓
Telemarketing with DNS	2	0.20	.217	

for statistically significant differences in mean times to answer by company (Table A.3) and by format (Table A.4).

Table A.3: Statistical Significance Tests for Time to Answer by Company

Question	d.f.	F value	p	Significant?
Cookies	5	1.18	.320	
Opt Out Link	5	5.58	< .001	✓
Share Email	5	1.81	.109	
Telemarketing	5	1.75	.122	

Table A.4: Statistical Significance Tests for Time to Answer by Format

Question	d.f.	F value	p	Significant?
Cookies	2	4.50	< .012	✓
Opt Out Link	2	3.59	.028	✓
Share Email	2	0.15	.864	
Telemarketing	2	8.59	< .001	✓

A.1.3 Psychological Acceptability

As described in Chapter 4, we asked a series of questions to capture subjective impressions of the privacy policies. Responses were on a seven point Likert scale which is sufficient granularity to treat them as continuous variables. We performed ANOVA analysis to test for statistically significant differences in mean Likert scores by company (Table A.5) and by format (Table A.6).

Table A.5: Statistical Significance Tests for Psychological Acceptability by Company

Topic	Question	d.f.	F value	<i>p</i>	Significant?
Finding Info.	Explained thoroughly	5	1.9	.038	✓
Finding Info.	Confident understood	5	1.9	.099	
Finding Info.	Easier to understand	5	1.6	.148	
Finding Info.	Hard to find	5	.75	.589	
Trust	Feel secure	5	7.0	< .001	✓
Trust	Protect more	5	3.9	.020	✓
Enjoyment	Pleasurable	5	1.7	.135	
Enjoyment	Likely to read	5	2.4	.096	

Table A.6: Statistical Significance Tests for Psychological Acceptability by Format

Topic	Question	d.f.	F value	<i>p</i>	Significant?
Finding Info.	Explained thoroughly	2	1.6	.203	
Finding Info.	Confident understood	2	.33	.722	
Finding Info.	Easier to understand	2	2.89	.051	
Finding Info.	Hard to find	2	.60	.549	
Trust	Feel secure	2	14.4	< .001	✓
Trust	Protect more	2	8.0	< .001	✓
Enjoyment	Pleasurable	2	.62	.539	✓
Enjoyment	Likely to read	2	2.4	.032	

Appendix B

Behavioral Advertising Study

This is the Willingness To Accept (WTA) version of the user study described in Chapter 5. These questions appeared slightly differently to online participants than they do here in a version formatted for the printed page. For example, the page titles shown were never shown to participants, and were purely for the researchers' convenience.

In the Willingness to Pay (WTP) version, question 14 reads: "Would you pay an additional \$1 per month to your Internet service provider (ISP) to avoid having your favorite news site collect your data for behavioral advertisements?" Otherwise the two versions are identical.

Online Advertising

Page One

1. How frequently do you buy things over the Internet? (Required)

- ☐ A few times per week
- ☐ A few times per month
- ☐ A few times per year
- ☐ Never buy things online
- ☐ Other (please explain)

2. How frequently do you buy things based on Internet advertisements? (Required)

- ☐ A few times per week
- ☐ A few times per month
- ☐ A few times per year
- ☐ Never buy things based on Internet advertisements
- ☐ Other (please explain)

3. How frequently do you make purchases based on advertising in email? (Required)

- ☐ A few times per week
- ☐ A few times per month
- ☐ A few times per year
- ☐ Never buy things based on advertising in email
- ☐ Other (please explain)

4. What is a computer "cookie"? (Please answer on the basis of your own knowledge. Do not look up answers. We want to know what you think, not what someone else thinks! It is ok to say you are not sure.) (Required)

5. What could sellers do to entice you to purchase more products online? (Required)

	Matters a lot	Matters	Matters a little	Does not matter
Free shipping	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online coupons	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Local pickup	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Web discounts	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Improved fraud protection for credit card transactions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will not share your data with advertising partners	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Easy-to-use website	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Clear information about products	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No hassle return policy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Products recommended based on your past purchases	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Products recommended based on your friends' past purchases	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will only retain data about your purchases for three months	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No spam policy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6. Is there something else sellers could do to encourage you to buy online?

7. Do you want websites you visit to show you ads that are tailored to your interests? (Required)

☐ Yes☐ No☐ Other (please explain)

8. Do you want websites you visit to show you discounts that are tailored to your interests? (Required)

☐ Yes☐ No☐ Other (please explain)

9. Do you want websites you visit to show you news that is tailored to your interests? (Required)

☐ Yes☐ No☐ Other (please explain)

Untitled Page

10. What is your favorite news website? (Required)

11. How frequently do you visit this website? (Required)

- ☐ A few times per day
☐ Once a day
☐ A few times per week
☐ A few times per month
☐ Once a month
☐ Less frequently than once a month
☐ Other (please explain)

12. Some websites use behavioral advertising to change which ads they display in response to your online activities over time and across multiple websites. Imagine your favorite news website offered you a choice of receiving behavioral advertising or randomly chosen ads, which would you chose? (Required)

- ☐ Behavioral advertising
☐ Random advertising

13. How strongly do you agree or disagree with the following statements: (Required)

	Strongly Agree	Agree	Mildly Agree	Neutral	Mildly Disagree	Disagree	Strongly Disagree
I want the benefits of relevant advertising	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I do not care if advertisers collect data about my search terms	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I ignore ads, so I do not care if ads are targeted to my interests or if ads are random	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am protected by law against advertisers collecting data about me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Behavioral targeting works poorly and I get ads that are not relevant to me, even when they are supposed to be	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would stop using any site that uses behavioral advertising	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would watch what I do online more carefully if I knew advertisers were collecting data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Someone keeping track of my activities online is invasive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I do not care if advertisers collect data about which websites I visit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I ignore ads, so there is no benefit to me if ads are targeted to my interests	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I ignore ads, so there is no harm to me if ads are targeted to my interests	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

14. Would you accept a discount of \$1 per month off your Internet service provider (ISP) bill to allow your favorite news site to collect your data for behavioral advertisements? (Required)

- ☐ Yes
☐ No
☐ Other (please explain)

15. Some websites may offer you a choice of paying for content or receiving content for free in exchange for letting them send you targeted advertising. How strongly do you agree or disagree with the following statements? (Required)

	Strongly Agree	Agree	Mildly Agree	Neutral	Mildly Disagree	Disagree	Strongly Disagree
It is not worth paying extra to avoid targeted ads	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Advertisers will collect data whether I pay or not, so there is no point paying	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Privacy is a right and it is wrong to be asked to pay to keep companies from invading my privacy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Companies asking me to pay for them not to collect data is extortion	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I hate ads and would pay to avoid them	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

20. If you use more than one web browser on your primary computer, why do you do so?

21. About how frequently do you clear cookies from the computer you typically use? (Required)

- ☐ Never
- ☐ Less than once a year
- ☐ About every year
- ☐ A few times a year
- ☐ Monthly
- ☐ A few times a month
- ☐ A few times a week
- ☐ Daily
- ☐ More frequently than daily
- ☐ Whenever I close the web browser
- ☐ Other (please explain)

22. Why? (Required)

23. Do you use any software that deletes cookies for you? (Required)

- ☐ Not sure
- ☐ No
- ☐ Yes (please name or describe)

24. Do you use a "private browsing" mode in your web browser? (Required)

- ☐ Not sure
- ☐ No
- ☐ Yes

25. People make different decisions about cookies. Please indicate if the following statements reflect decisions you make. (Required)

	True	False	Does not apply to me
I delete cookies so no one else can see which links I have visited on my computer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I delete cookies to avoid malware (viruses, spyware, phishing, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I delete cookies to save space	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I delete cookies out of habit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I delete cookies but only on shared computers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I do not delete cookies because they make web pages load faster	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I do not delete cookies because they keep me signed in to websites	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I do not delete cookies because I want to see which links I have visited before	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I accept cookies because sites do not work otherwise	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I accept cookies because saying "yes" or "no" to each cookie takes too long or is a hassle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I reject cookies from sites I don't trust	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I delete cookies so companies cannot track me over time	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

26. Sometimes you hear about web browser history. Are cookies and history the same? (Required)

☐ Yes

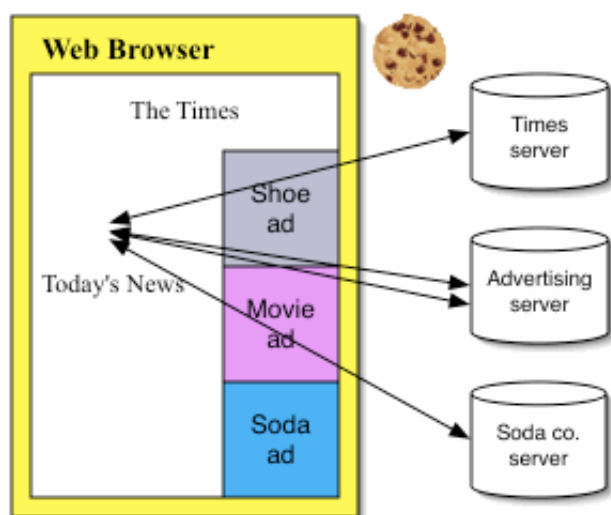
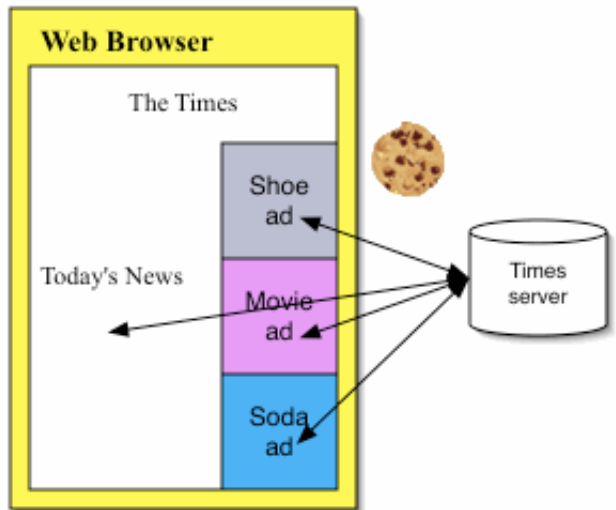
☐ No (please explain how they differ)

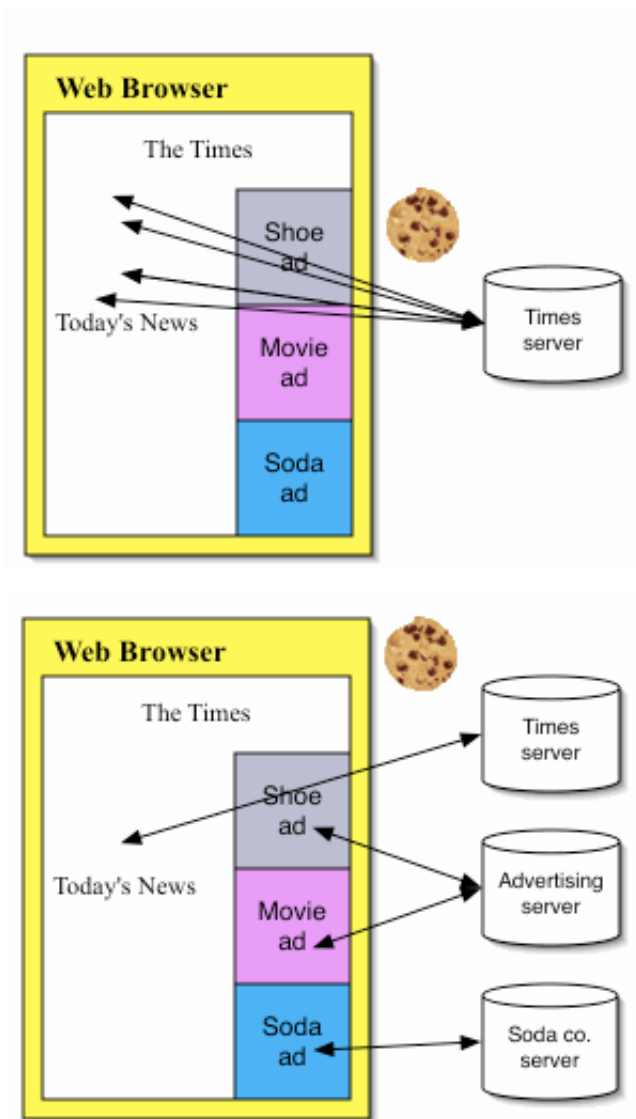
27. Please indicate if the following statements about computer cookies are true, false, or if you are not sure. (Required)

	True	False	Unsure
Cookies are small bits of data stored on my computer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A website I visit can read every cookie I have, no matter which website the cookie is from	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cookies let me stay logged in over time without needing to enter my password every time I visit a site	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Advertisers can use cookies on multiple websites to learn which sites I visit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cookies let web browsers' forward and backward arrows work correctly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cookies store a history of websites I have visited in the past	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cookies change the color of hyperlinks to websites I have already visited	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
By law, cookies may not contain credit card information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cookies contain information from when I first purchased my computer, including my name and home address	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The PATRIOT ACT allows law enforcement officials to read my cookies if I exchange email with someone on the terrorist watch list	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cookies are a type of spyware	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cookies let people send me spam	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cookies let websites display more quickly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cookies enable personalized advertising based on my prior behavior online	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cookies may be combined with other data that identifies me by name	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If I do not accept cookies, websites cannot tell where I am physically located	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cookies enable personalized content like color schemes or what type of information I want to see on a website	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Model of Cookies

Please refer to the images below to answer questions at the bottom of this page. Imagine you are using a standard web browser to visit The Times website, which has ads as depicted in the diagrams. There are no other non-visible components to the webpage.





28. If you were visiting The Times website, which diagram do you think best illustrates how cookies usually work? (Required)

- ☐ Figure 1
- ☐ Figure 2
- ☐ Figure 3
- ☐ Figure 4

29. With the diagram you selected, what would happen when a web browser loads the page?

(Required)

- ☐ All the cookies shown in the illustration will be transmitted automatically when a user visits this page.
- ☐ Some of the cookies shown in the illustration will be transmitted automatically when a user visits this page and some won't be transmitted until a user clicks on the ads on this page.
- ☐ None of the cookies shown in the illustration will be transmitted until a user clicks on links or ads on this page.

Not sure

Other (please explain)

30. Which, if any, of the diagrams above could not happen? (Please check all that apply.)

(Required)

- ☐ Figure 1 could not happen
- ☐ Figure 2 could not happen
- ☐ Figure 3 could not happen
- ☐ Figure 4 could not happen
- ☐ All figures are possible

Untitled Page

Complete this section only if: "If you were visiting The Times website, which diagram do you think best illustrates how cookies usually work?" matches: 'Figure 4'

The image below shows search results on a popular website. We have added red boxes and letters A through F to identify portions of the web page. Please refer to the image to answer the questions below.

The image is a screenshot of a Google search results page for the query "advertising". The page is divided into several sections, each identified by a red box and a letter:

- A**: The top navigation bar and search bar area, including links for Web, Images, Videos, Maps, News, Shopping, Gmail, and more, along with the Google logo and the search input field.
- B**: The main search results area, showing the first few results for "advertising". The results include sponsored links, a Wikipedia entry, and several other websites.
- C**: The right-hand sidebar, which contains additional sponsored links and a "Sponsored Links" section.
- D**: The "Searches related to advertising" section, which lists various related search terms like "types of advertising", "advertising careers", "advertising history", "advertising tips", "advertising agency", "advertising campaigns", "advertising costs", and "advertising jobs".
- E**: The bottom navigation bar, which includes links for "Google Home", "Advertising Programs", "Business Solutions", "Privacy", and "About Google".
- F**: The search bar area at the bottom of the page, which includes the search input field and the "Search" button.

31. How many ads are there in each section in the image above? If there are any ads in a section, please describe what the ads are for. It's ok to say you aren't sure.

	# Ads	Description of ads (it's ok to say you aren't sure)
A		
B		
C		
D		
E		

32. Imagine you visited google.com and performed the search shown above. Assume you do not have any ad blocking software and have your browser configured to accept all cookies. In section B shown above, are there any companies able to set cookies on your computer? If so, which company or companies? (Required)

☐ No

☐ Yes (which company or companies?)

33. Imagine you visited google.com and performed the search shown above. Assume you do not have any ad blocking software and have your browser configured to accept first party cookies, but reject third party cookies. In section B shown above, are there any companies able to set cookies on your computer? If so, which company or companies? (Required)

☐ No

☐ Yes (which company or companies?)

Advertising Identification

Complete this section only if: "If you were visiting The Times website, which diagram do you think best illustrates how cookies usually work?" matches: 'Figure 4'

The image below is the home page for a popular website. We have added red boxes and letters A through G to identify portions of the web page. Please refer to the image to answer the questions below.

The image shows the Yahoo! homepage as of March 20, 2010. Red boxes and letters A through G are used to identify specific areas of the page:

- A**: A small red box in the top left corner, near the "Make Yahoo! your homepage" banner.
- B**: A red box around the "Web Search" button.
- C**: A red box around the "Today" section, which features a large image of Katherine Heigl and the headline "Katherine Heigl endures dress disaster".
- D**: A red box around the "Trending Now" section, which lists various news items and a Mazda advertisement.
- E**: A red box around the "More Yahoo! Sites" section, which lists various services like Answers, Autos, Finance, etc.
- F**: A red box around the "Yahoo! for Your Business" section, which lists services like Small Business Solutions, Advertise with Us, etc.
- G**: A red box around the footer area, which contains copyright information and links to Privacy Policy, About Our Ads, etc.

34. How many ads are there in each section in the image above? If there are any ads in a section, please describe what the ads are for. It's ok to say you aren't sure.

	# Ads	Description of ads (It's ok to say you aren't sure)
A		
B		
C		
D		
E		
F		
G		

35. The Mazda ad in box D can set a cookie on my computer, but only if I click on the ad. If I do not click on the ad, it cannot set a cookie. (Required)

- ☐ True
☐ False
☐ Other (please explain)

36. Imagine you visited yahoo.com as shown above. Assume you do not have any ad blocking software and have your browser configured to accept all cookies. In section D shown above, are there any companies able to set cookies on your computer? If so, which company or companies? (Required)

- ☐ No
☐ Yes (which company or companies?)

37. Imagine you visited yahoo.com and as shown above. Assume you do not have any ad blocking software and have your browser configured to accept first party cookies, but reject third party cookies. In section D shown above, are there any companies able to set cookies on your computer? If so, which company or companies? (Required)

- ☐ No
☐ Yes (which company or companies?)

38. From section B shown above, if your browser accepts all cookies, which of the following companies can set cookies on your computer? (Required)

	Can set cookies	Cannot set cookies
Yahoo!	<input type="radio"/>	<input type="radio"/>
Facebook	<input type="radio"/>	<input type="radio"/>
Google	<input type="radio"/>	<input type="radio"/>
HotJobs	<input type="radio"/>	<input type="radio"/>
AdWords	<input type="radio"/>	<input type="radio"/>
A weather company	<input type="radio"/>	<input type="radio"/>

39. In box B, the last item in "MY FAVORITES" is Weather. Imagine you visited yahoo.com and did not log in to Yahoo! Would the weather report be customized for your city? Please explain why or why not. (Required)

- ☐ Yes (please explain why)
- ☐ No (please explain why)

40. In box D, what does the top image show? (Required)

- ☐ Printers
- ☐ Cars
- ☐ Guitars
- ☐ Penguins

NAI opt out

The image below is the home page for a popular website. Please refer to the image to answer the questions below.

NAI
Network Advertising Initiative

Consumer Opt-Out | Privacy

Home Managing Your Privacy Participating Networks About Us Contact Us Blog

Overview
Principles Overview
Opt-Out
Enforcement
Opt-out Problems
FAQs
Learn More

Opt Out of Behavioral Advertising

The NAI Opt-out Tool was developed in conjunction with our members for the express purpose of allowing consumers to "opt out" of the behavioral advertising delivered by our member companies.

Using the Tool below, you can examine your computer to identify those member companies that have placed an advertising cookie file on your computer.

To opt out of an NAI member's behavioral advertising program, simply check the box that corresponds to the company from which you wish to opt out. Alternatively, you can check the box labeled "Select All" and each member's opt-out box will be checked for you. Next click the "Submit" button. The Tool will automatically replace the specified advertising cookie(s) and verify your opt-out status.

Opting out of a network does not mean you will no longer receive online advertising. It does mean that the network from which you opted out will no longer deliver ads tailored to your Web preferences and usage patterns.

If you have any questions, please visit our [FAQ section](#).

Opt-Out Status

Vibrant Media More Information	No Cookie You have not opted out and you have no cookie from this network.	Opt-Out <input type="checkbox"/>
[x+1] (formerly Poindexer Systems) More Information	Active Cookie You have not opted out and you have an active cookie from this network.	Opt-Out <input type="checkbox"/>
Yahoo! Ad Network More Information	Active Cookie You have not opted out and you have an active cookie from this network.	Opt-Out <input type="checkbox"/>
TACODA Audience Networks More Information	No Cookie You have not opted out and you have no cookie from this network.	Opt-Out <input type="checkbox"/>
Tribal Fusion More Information	No Cookie You have not opted out and you have no cookie from this network.	Opt-Out <input type="checkbox"/>

Select all Clear Submit

41. Based on the image above, if you visited this web site, what would you think it is? Check as many as apply. (Required)

- ☐ A scam website to collect your private information
- ☐ A scam website designed to find out which websites you visit
- ☐ A website that lets you see fewer online ads
- ☐ A website that lets you tell companies you do not want to see ads from them, but you will still see as many ads overall
- ☐ A website that lets you tell companies not to collect data about you
- ☐ A website that allows companies to profile you, but not show you ads based on those profiles
- ☐ Other (please explain)

42. What would happen if you checked the opt out box for Yahoo! Ad Network? (Required)

43. What would happen if you clicked the opt out box for Vibrant Media? (Required)

44. If you saw this website like this, would you use it? (Required)

- ☐ Yes
- ☐ No
- ☐ Not sure (please explain)

Untitled Page

For the following questions, please read the description of a possible scenario and then answer if you think this is something that could happen or will not happen.

45. Imagine you visit the New York Times website. One of the ads is for Continental airlines. That ad does not come to you directly from the airline. Instead, there is an ad company that determines what ad to show to you, personally, based on the history of prior websites you have visited. Your friends might see different ads if they visited the New York Times. (Required)

- ☐ This happens a lot right now
- ☐ This happens a little right now
- ☐ This does not happen now but could happen in the future
- ☐ This will never happen because it is not allowed by law
- ☐ This will never happen because there would be consumer backlash against companies that engaged in this practice
- ☐ This will never happen because it would not be profitable
- ☐ Other (Please explain)

46. How would you feel about this practice? Choose as many as apply. (Required)

- ☐ Glad to have relevant advertisements about things I am interested in instead of random advertisements
- ☐ Wouldn't even notice the advertisements, just ignore them
- ☐ It's creepy to have advertisements based on sites I've visited
- ☐ No one should use data from Internet history
- ☐ Other (please explain):
- ☐ It's ok as long as the New York Times is free

47. Imagine you are online and your email provider displays ads to you. The ads are based on what you write in email you send, as well as email you receive. (Required)

- ☐ This happens a lot right now
- ☐ This happens a little right now
- ☐ This does not happen now but could happen in the future
- ☐ This will never happen because it is not allowed by law
- ☐ This will never happen because there would be consumer backlash against companies that engaged in this practice
- ☐ This will never happen because it would not be profitable
- ☐ Other (Please explain)

48. How would you feel about this practice? Choose as many as apply. (Required)

- ☐ Glad to have relevant advertisements about things I am interested in instead of random advertisements
- ☐ Wouldn't even notice the advertisements, just ignore them
- ☐ It's creepy to have advertisements based on my emails
- ☐ No one should use data from email because it is private like postal mail
- ☐ Other (please explain):
- ☐ It's ok as long as the email service is free

49. Please indicate how much you agree or disagree with the following statements (Required)

	Strongly Agree	Agree	Mildly Agree	Neutral	Mildly Disagree	Disagree	Strongly Disagree
Online advertising is annoying	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising is necessary for the Internet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising lets me read sites without paying money	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising is beneficial because ads are a source of information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising is better than television or billboards because it is easier to ignore online ads	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising tends to be related to the website it is on	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising tends to be for products or services I am interested in	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Compared to other types of advertising like television or magazines, online advertisement has more to do with what I want and is less random	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising is insulting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising makes the Internet slower	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising is distracting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising has too many unrelated and off topic ads	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising based on my actions online is creepy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising for medical products should be regulated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online advertising is just a fact of life	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

50. Imagine you are using a standard web browser (configured to accept third party cookies with no ad blocking.) You go to the Yahoo website and it contains an advertisement directly from Amazon. Which cookies can be set on your computer? (Required)

- ☐ None at all
☐ None at all unless you click something
☐ None at all unless you login to Yahoo
☐ None at all unless you login to Amazon
☐ Yahoo only, regardless of whether you are signed in to Yahoo
☐ Yahoo regardless of whether you are signed in to Yahoo, and Amazon only if you click the Amazon ad
☐ Yahoo regardless of whether you are signed in to Yahoo, and Amazon only if you login to Amazon
☐ Yahoo and Amazon, regardless of whether you are signed in or click anything
☐ Not sure
☐ Other (please explain)

51. Now imagine the ad for Amazon comes from an ad server that displays different ads each time you visit the Yahoo! site. Can the ad server set a cookie on your computer? (Required)

- ☐ Yes
☐ Yes, but only if I click the ad
☐ No
☐ Not sure
☐ Other (please explain)

52. Can the ad server read cookies set by Yahoo? (Required)

- ☐ Yes
- ☐ Yes, but only if I am logged in to Yahoo
- ☐ Yes, but only if I click the ad
- ☐ No
- ☐ Not sure
- ☐ Other (please explain)

Untitled Page

53. What is your current occupation?

- ☐ Administrative Support (eg., secretary, assistant)
- ☐ Art, Writing and Journalism (eg., author, reporter, sculptor)
- ☐ Business, Management and Financial (eg., manager, accountant, banker)
- ☐ Education (eg., teacher, professor)
- ☐ Legal (eg., lawyer, law clerk)
- ☐ Medical (eg., doctor, nurse, dentist)
- ☐ Science, Engineering and IT professional (eg., researcher, programmer, IT consultant)
- ☐ Service (eg., retail clerks, server)
- ☐ Skilled Labor (eg., electrician, plumber, carpenter)
- ☐ Student
- ☐ Other Professional
- ☐ Not Currently Working/Currently Unemployed
- ☐ Retired
- ☐ Other (please specify)
- ☐ Full-time parent or home maker

54. What is your gender?

- ☐ Female
- ☐ Male

55. What is your age?

56. Please select your race (check as many as apply)

- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Black or African American
- ☐ Latina/Latino or Hispanic
- ☐ Native Hawaiian or Other Pacific Islander
- ☐ White

57. Which of the following best describes your highest achieved education level?

- ☐ Some High School
- ☐ High School Graduate
- ☐ Some college, no degree
- ☐ Associates degree
- ☐ Bachelors degree
- ☐ Graduate degree (Masters, Doctorate, etc.)

58. What is your total household income?

- ☐ Less than \$12,500
- ☐ \$12,500 - \$24,999
- ☐ \$25,000 - \$37,499
- ☐ \$37,500 - \$49,999
- ☐ \$50,000 - \$62,499
- ☐ \$62,500 - \$74,999
- ☐ \$75,000 - \$87,499
- ☐ \$87,500 - \$99,999
- ☐ \$100,000 - \$112,499
- ☐ \$112,500 - \$124,999
- ☐ \$125,000 or More

59. In which country do you live?

- ☐ United States
- ☐ Other (please specify)

60. How many years have you been using the Internet?

61. How would you describe your computer use? Please check all that apply:

- ☐ Surf the web
- ☐ Send and read email
- ☐ Maintain a blog
- ☐ Use general office productivity programs (Word, Excel, PowerPoint, etc.)
- ☐ Use specialized programs (Photoshop, R, etc.)
- ☐ Web programming (HTML, CSS, AJAX, etc.)
- ☐ All other types of programming (Java, C, Lisp, etc.)
- ☐ Other (please describe)

62. Please provide any comments you may have about this study below [optional]

Appendix C

Targeted Advertising Comparison

This is the “cancer” version of the user study described in Chapter 6. The “scuba” version used examples about scuba with the following changes:

Prior to question 7, “For example, if you search for the word “scuba,” you might see ads for scuba gear. You would not see ads for scuba gear again while searching for other topics.”

Prior to question 13, “For example, if you purchase scuba gear from an online site, over time your friends might see ads for wetsuits.”

Prior to question 19, “For example, if you read an online article about scuba on a news site, then searched for “best reefs,” on a different website, next week you might see an ad for a scuba vacation.”

Prior to question 26, “For example, if you uploaded a photo of yourself diving, plus purchased a pair of pants online, your ISP might show you ads for wetsuits in your size.”

Prior to question 32, “For example, if you used an online mail service and exchanged email with a friend about your new interest in scuba, you might see ads for scuba instructors.”

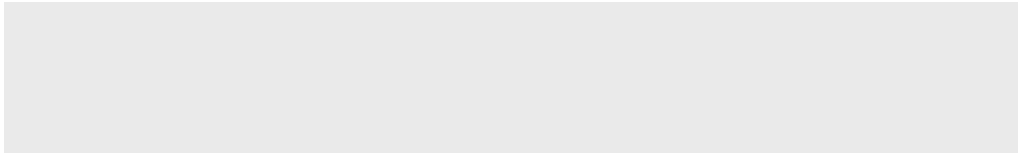
Online Advertising Study

Page One

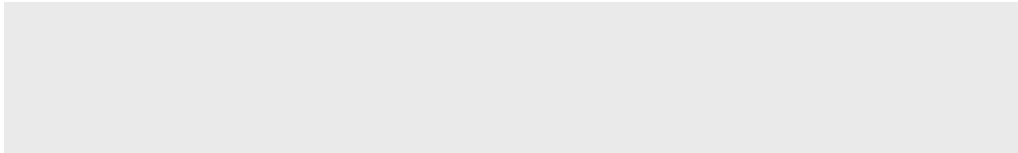
Thank you for taking our survey. Please note that it is only open to you once. If you take it a second time (even months later) we will have to reject the second HIT. At the end of the study you will get a secret code to enter into mturk so you can get paid.

Important: Please do not consult reference material (google, wikipedia, etc.) Just answer on the basis of your current knowledge. We are not interested in what someone else thinks - we want to know what you think! You may give an incomplete answer or say you do not know.

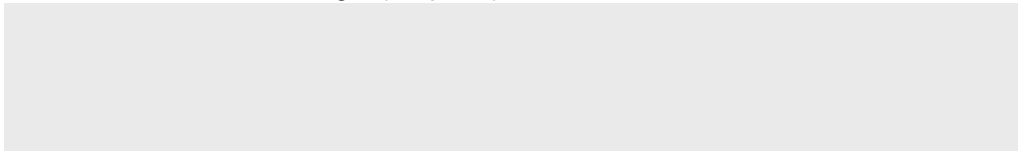
1. Some websites use "cookies". What is a cookie? (Required)

A large, light gray rectangular box intended for the user to type their answer to the first question.

2. Some websites have "third party cookies". What is a third party cookie? (Required)

A large, light gray rectangular box intended for the user to type their answer to the second question.

3. What is "behavioral advertising"? (Required)

A large, light gray rectangular box intended for the user to type their answer to the third question.

Contextual

Some websites use contextual advertising to change the ads they display in response to your current actions online. For example, if you search for the word "cancer," you might see ads for medical clinics. You would not see ads related to cancer again while searching for other topics.

4. Please check all of the ways you have heard about contextual advertising in the last 30 days. (Required)

- ☐ Not at all
- ☐ TV
- ☐ Radio
- ☐ Magazine (online or paper)
- ☐ Newspaper (online or paper)
- ☐ From a friend
- ☐ Blog
- ☐ A website describing the practice in general (please list the website(s) here)
- ☐ A website describing their own practices (please list the website(s) here)
- ☐ Other (please describe)

5. Please indicate how much you agree or disagree that the words below describe contextual advertising: (Required)

	Strongly disagree	Disagree	Mildly disagree	Neutral	Mildly agree	Agree	Strongly agree
Concerning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Creepy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Desirable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Entertaining	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Helpful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Invasive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pushy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reasonable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Affiliate

Some websites use affiliate marketing to change which ads they display in response to online activities within a group of friends. For example, if you purchase a book about brain cancer from an online site, over time your friends might see ads for home nursing care. Similarly, you may see ads related to your friends' purchases or websites they visited.

6. Please check all of the ways you have heard about affiliate marketing in the last 30 days.
(Required)

- ☐ Not at all
- ☐ TV
- ☐ Radio
- ☐ Magazine (online or paper)
- ☐ Newspaper (online or paper)
- ☐ From a friend
- ☐ Blog
- ☐ A website describing the practice in general (please list the website(s) here)
- ☐ A website describing their own practices (please list the website(s) here)
- ☐ Other (please describe)

7. Please indicate how much you agree or disagree that the words below describe affiliate marketing: (Required)

	Strongly disagree	Disagree	Mildly disagree	Neutral	Mildly agree	Agree	Strongly agree
Concerning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Creepy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Desirable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Entertaining	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Helpful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Invasive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pushy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reasonable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Behavioral

Some websites use behavioral advertising to change which ads they display in response to your online activities over time and across multiple websites. For example, if you read an online article about a link between cell phones and cancer on a news site, then searched for "brain cancer," on a different website, next week you might see an ad for a drugs that reduce side effects from chemotherapy.

8. Please check all of the ways you have heard about behavioral advertising in the last 30 days. (Required)

- ☐ Not at all
- ☐ TV
- ☐ Radio
- ☐ Magazine (online or paper)
- ☐ Newspaper (online or paper)
- ☐ From a friend
- ☐ Blog
- ☐ A website describing the practice in general (please list the website(s) here)
- ☐ A website describing their own practices (please list the website(s) here)
- ☐ Other (please describe)

9. Please indicate how much you agree or disagree that the words below describe behavioral advertising: (Required)

	Strongly disagree	Disagree	Mildly disagree	Neutral	Mildly agree	Agree	Strongly agree
Concerning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Creepy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Desirable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Entertaining	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Helpful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Invasive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pushy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reasonable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. The Children's Online Privacy Protection Act of 1998 (COPPA) went into effect on the twenty first of April in 2000 to protect children under the age of 13. On what date did COPPA go into effect? (Required)

- ☐ 4/13/00
- ☐ 4/21/00
- ☐ 4/13/98
- ☐ 4/21/98

DPI

Some websites use deep packet inspection (DPI) to change the ads they display in response to all of your actions online. For example, if you exchanged email with a friend about your Aunt's death from cancer, you might see ads for genetic testing to assess your own cancer risk.

11. Please check all of the ways you have heard about deep packet inspection (DPI) in the last 30 days. (Required)

- ☐ Not at all
- ☐ TV
- ☐ Radio
- ☐ Magazine (online or paper)
- ☐ Newspaper (online or paper)
- ☐ From a friend
- ☐ Blog
- ☐ A website describing the practice in general (please list the website(s) here)
- ☐ A website describing their own practices (please list the website(s) here)
- ☐ Other (please describe)

12. Please indicate how much you agree or disagree that the words below describe deep packet inspection (DPI): (Required)

	Strongly disagree	Disagree	Mildly disagree	Neutral	Mildly agree	Agree	Strongly agree
Concerning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Creepy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Desirable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Entertaining	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Helpful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Invasive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pushy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reasonable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Demographics

13. What is your gender?

- ☐ Female
☐ Male

14. What is your age?

Under 18
18 - 24
25 - 34
35 - 44
45 - 54
55 - 64
65 - 74
75 or over

15. Please select your race (check as many as apply)

- ☐ American Indian or Alaska Native
☐ Asian
☐ Black or African American
☐ Latina/Latino or Hispanic
☐ Native Hawaiian or Other Pacific Islander
☐ White

16. Which of the following best describes your highest achieved education level?

Some High School
High School Graduate
Some college, no degree
Associates degree
Bachelors degree
Graduate degree (Masters, Doctorate, etc.)

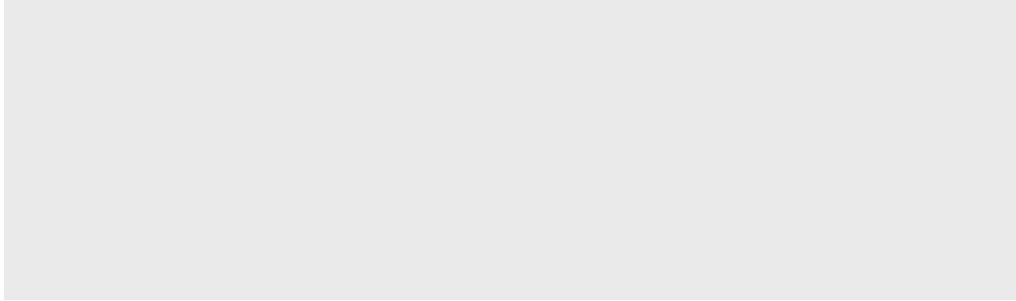
17. What is your total household income?

Less than \$12,500
\$12,500 - \$24,999
\$25,000 - \$37,499
\$37,500 - \$49,999
\$50,000 - \$62,499
\$62,500 - \$74,999
\$75,000 - \$87,499
\$87,500 - \$99,999
\$100,000 - \$112,499
\$112,500 - \$124,999
\$125,000 or More

18. In which country do you live?

- ☐ United States
☐ Other (please specify)

19. Please provide any comments you may have about this study below [optional]



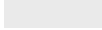
Important! To prevent fraud, you must enter your secret code and the last four digits of your phone number here and when you return to mturk to get paid. Please copy and paste the code below.

Your secret code:

20. What is your secret code (shown above)? (Required)



21. What are the last 4 digits of your phone number? (Required)



Appendix D

Sites Visited to Quantify Flash Cookie Use

We analyzed three data sets based on Quantcast's list of the million most visited websites. Two were top 100 most visited sites as of July 2009 and 2010. They differed by 31 sites, shown in bold. The final data set is 500 sites we randomly selected from the Quantcast list of one million.

Table D.1: Quantcast's Top 100 most visited websites as of July 8, 2010

about.com,	adobe.com,	amazon.com,
americangreetings.com,	answers.com,	aol.com,
ap.org,	apple.com,	ask.com,
associatedcontent.com,	att.com,	bankofamerica.com,
bbc.co.uk,	bestbuy.com,	bing.com,
bizrate.com,	blinkx.com,	blogger.com,
blogspot.com,	bluemountain.com,	break.com,
careerbuilder.com,	causes.com,	chase.com,
chinaontv.com,	city-data.com,	cnet.com,
cnn.com,	comcast.com,	comcast.net,
craigslist.org,	dailymotion.com,	digg.com,
drudgereport.com,	ebay.com,	ehow.com,
evite.com,	examiner.com,	facebook.com,
flickr.com,	formspring.me,	go.com,
godaddy.com,	google.com,	hp.com,
hubpages.com,	huffingtonpost.com,	hulu.com,
ign.com,	imdb.com,	latimes.com,
legacy.com,	linkedin.com,	live.com,
mapquest.com,	match.com,	merriam-webster.com,
metacafe.com,	microsoft.com,	monster.com,
msn.com,	mtv.com,	mybloglog.com,
myspace.com,	netflix.com,	nytimes.com,
optiar.com,	pandora.com,	paypal.com,
people.com,	photobucket.com,	reference.com,
reuters.com,	simplyhired.com,	suite101.com,
target.com,	thefind.com,	tmz.com,
tumblr.com,	twitpic.com,	twitter.com,
typepad.com,	usps.com,	walmart.com,
washingtonpost.com,	weather.com,	weatherbug.com,
webmd.com,	wellsfargo.com,	whitepages.com,
wikia.com,	wikipedia.org,	windows.com,
wordpress.com,	wunderground.com,	yahoo.com,
yellowpages.com,	yelp.com,	youtube.com,
zynga.com		

Table D.2: Quantcast's Top 100 most visited websites as of July, 2009

about.com,	adobe.com,	amazon.com,
------------	------------	-------------

Continued on Next Page...

Table D.2 – Continued

answers.com,	aol.com,	apple.com,
ask.com,	att.com,	bankofamerica.com,
bestbuy.com,	bizrate.com,	blogger.com,
blogspot.com,	capitalone.com,	careerbuilder.com,
chase.com,	city-data.com,	classmates.com,
cnet.com,	cnn.com,	comcast.com,
comcast.net,	craigslist.org,	dailymotion.com,
dell.com,	digg.com,	discovery.com,
ebay.com,	ehow.com,	evite.com,
expedia.com,	ezinearticles.com,	facebook.com,
flickr.com,	geocities.com,	go.com,
google.com,	homedepot.com,	hp.com,
huffingtonpost.com,	hulu.com,	imdb.com,
jcpenny.com,	linkedin.com,	live.com,
lowes.com,	mapquest.com,	merriam-webster.com,
metacafe.com,	microsoft.com,	mlb.com,
monster.com,	msn.com,	mtv.com,
myspace.com,	netflix.com,	nih.gov,
nytimes.com,	oprah.com,	pandora.com,
paypal.com,	people.com,	photobucket.com,
pogo.com,	pronto.com,	reference.com,
reuters.com,	scribd.com,	sears.com,
shopzilla.com,	simplyhired.com,	smarter.com,
tagged.com,	target.com,	ticketmaster.com,
time.com,	tripadvisor.com,	tripod.lycos.com,
twitter.com,	typepad.com,	ups.com,
usmagazine.com,	usps.com,	verizon.com,
verizonwireless.com,	vzw.com,	walmart.com,
weather.com,	webmd.com,	wellsfargo.com,
whitepages.com,	wikimedia.org,	wikipedia.org,
windows.com,	wordpress.com,	wunderground.com,
yahoo.com,	yellowpages.com,	youtube.com,
zimbio.com		

Table D.3: Random selection of 500 sites

24hourpet.com,	350smallblocks.com,	411webdirectory.com,
72712.com,	787787.com,	aalas.org,
aartkorstjens.nl,	abbottbus.com,	accutronix.com,
ad-mins.com,	adaholicsanonymous.net,	adamscountyhousing.com,
adorabubbleknits.com,	advanceexpert.net,	agnesfabricsshop.com,
air-land.com,	alignmed.com,	allstarsportspicks.com,
almostfrugal.com,	amandabeard.net,	amazingamberuncovered.com,
amigofoods.com,	ancestryhost.org,	appcelerator.com,
ar-10-rifles.com,	arcadianhp.com,	archerairguns.com,
ariionkathleenbrindley.com,	arizonabattery.com,	arizonahealingtours.com,
asbj.com,	asiainc-ohio.org,	askittoday.com,
askmd.org,	asla.org,	astonhotels.com,
atbfinancialonline.com,	athenscountyauditor.org,	auburncountryclub.com,
auctioneeraddon.com,	autorepairs-guide.info,	avistarentals.com,
awildernessvoice.com,	azbiz.com,	babygotfat.com,

Continued on Next Page...

Table D.3 – Continued

backwoodssurvivalblog.com,	badvoter.com,	bargainmartclassifieds.com,
battlestargalactica.com,	beaconschool.org,	beatport.com,
beechwoodcheese.com,	benedictinesisters.org,	best-hairy.com,
bestshareware.net,	bethpage.coop,	bf1systems.com,
bibleclassbooks.com,	bibleverseposters.com,	bird-supplies.net,
blackopalmine.com,	bladesllc.com,	blogmastermind.com,
bluetoothringtones.net,	body-piercing-jewellery.com,	bookjobs.com,
boulevardsentinel.com,	boyntonbeach.com,	bradcallen.com,
brealynn.info,	brill.nl,	broncofix.com,
buckstradingpost.com,	bucky.com,	buyhorseproperties.com,
bwcnfarms.com,	cabands.com,	cabins.ca,
cafemomstatic.com,	capitalgainsmedia.com,	cardiomyopathy.org,
careerstaffingnow.com,	carrollshelbymerchandise.com,	cashloanbonanza.com,
cateringatblackswan.com,	cdcoupons.com,	charterbank.com,
charterco.com,	chashow.org,	cheapusedcars.com,
childrensheartinstitute.org,	christmas-trees-wreaths-decorations.com,	clarislifesciences.com,
claytonihouse.com,	clcofwaco.org,	clean-your-pcc1.com,
cloningmagazine.com,	clubdvsx.com,	codeproject.com,
coltbus.org,	coltranet.com,	columbusparent.com,
complxregionalpainsyndrome.net,	computervideogear.com,	conservativedvds.com,
cookbooksforsale.com,	coolatta.org,	corvettepartsforsale.com,
countrymanufacturing.com,	cpainquiry.com,	crazyawesomemeyeah.com,
crbna.com,	creatupropiaweb.com,	credit-improvers.net,
credicaredirect.com,	crowderhitecrews.com,	culttvm2.com,
curepeyronies.net,	curiousinventor.com,	dansdidnts.com,
dardenrestaurants.com,	datingthoughts.com,	dcso.com,
de.ms,	dealante.com,	dealsoutlet.net,
delti.com,	desktops.net,	detroitmasonic.com,
digitalmania-online.com,	disasterreliefeffort.org,	dividend.com,
dmvedu.org,	dobbstireandauto.com,	dodgeblockbreaker.com,
donlen.com,	donnareed.org,	dorpexpress.com,
dukeandthedoctor.com,	dvdsetcollection.com,	easypotatosalad.com,
educationalrap.com,	elmersgluecrew.com,	emailfwds.com,
emailsparkle.com,	empty.de,	ereleases.com,
escapefate.net,	eurekasprings.org,	evanity.com,
expowest.com,	eyesite.org,	fashionreplicabags.com,
fast-guardcleaneronpc.net,	fatlove.net,	fearrington.com,
fitnesshigh.com,	flatpickdigital.com,	fleetairarmarchive.net,
florahydroponics.com,	floridafishinglakes.net,	flyingbarrel.com,
foodtimeline.org,	foreclosuredlist.com,	foreclosurepulse.com,
forzion.com,	fourreals.com,	free-party-games.com,
freepetclinics.com,	freshrewardscore.com,	fretwellbass.com,
fukushima.jp,	fullertontitans.com,	fundmojo.com,
fusioncrosstraining.com,	ga0.org,	gaara.ws,
ganstamovies.com,	gemission.org,	genesearch.com,
gerdab.ir,	getanagentnow.com,	girlfights.com,
globalfire.tv,	gmail.com,	gogivetraining.com,
gold-speculator.com,	goldenstaterails.com,	gomotobike.com,
goodseed.com,	googgpillz.com,	gordonbierschgroup.com,
gotostedwards.com,	goutresource.com,	graceandtruthbooks.com,
grooveeffect.com,	hairybulletgames.com,	halffuneralchapel.com,
hallmarkchannel.tv,	hammondstar.com,	happyshoemedia.com,
healthcaresalaryonline.com,	hills.net,	historyofnations.net,
hoover-realestate.com,	horseshoes.com,	hostpapa.com,
hoveringads.com,	howyouspinit.com,	hp-lexicon.com,
hsbc.com.mx,	hvk.org,	icdri.org,
idxcentral.com,	ieer.org,	iflextoday.com,
indianapolis.com,	infiniofdenver.com,	inhumanity.com,
inria.fr,	intelos.com,	iphonealley.com,
iris-photo.com,	itmweb.com,	itvs.com,
itw.com,	ivanview.com,	jacksoncountygov.com,
japanautopages.com,	jesus-passion.com,	jetbroadband.com,
jimmycanon.com,	josejuandiaz.com,	joybauernutrition.com,
junohomepage.com,	jwsuretybonds.com,	kbduct.com,
kimballarea.com,	kitten-stork.com,	knittingpureandsimple.com,
kpcstore.com,	lacosteshoes.us,	lafarge-na.com,
lakeareavirtualtours.com,	latinrank.com,	layover.com,
life-insurance-quotes-now.com,	lifepositive.com,	liftopia.com,
like.to,	lintvnews.com,	logodogzprintz.com,
lstractorusa.com,	ltwell.com,	lydiasitaly.com,
madisonindiana.org,	magnetnetworks.com,	marketminute.com,
mastiffrescue.org,	maurywebpages.com,	mayoarts.org,
mcperson.edu,	mcswain-evans.com,	measurebuilt.com,

Continued on Next Page...

Table D.3 – Continued

meiselwoodhobby.com,	menalive.com,	merbridal.com,
michiganford.com,	microcenter.com,	miltonmartintoyota.com,
minki.net,	mirdrag.com,	missourimalls.net,
mistercater.com,	mitutoyo.com,	mmodels.com,
modbee.com,	moforaja.com,	moldingjobs.com,
moneytip.com,	moselhit.de,	motomatters.com,
motosolvang.com,	movefrontlisten.com.com,	mule.net,
mundofree.com,	my-older-teacher.net,	mycomputerclub.com,
mylexia.com,	mypickapart.com,	mystic-nights.com,
mysticalgateway.com,	mysticlake.com,	mytableware.com,
nationalcoalition.org,	naturalmedicine.com,	ncbeachbargains.com,
ncgold.com,	nec.jp,	nekoarcnetwork.com,
newcracks.net,	newlawyer.com,	newmacfurnaces.com,
newscoma.com,	nexstitch.com,	nhlottery.com,
nittygrittyinc.com,	nobledesktop.com,	nottslad.com,
npg.org.uk,	nscale.org.au,	nwlanews.com,
ocharleydavidson.com,	offscreen.com,	oixi.jp,
olympus-imaging.com,	omahaimpound.org,	onelasvegas.com,
onepaycheckatime.com,	optimost.com,	orchidphotos.org,
outbackphoto.com,	ownacar.net,	ownthenight.com,
p2pchan.info,	parkcityinfo.com,	parksandcampgrounds.com,
paulrevereraiders.com,	pedalmag.com,	pennhealth.com,
performancehobbies.com,	perthmilitarymodelling.com,	pet-loss.net,
petworld.com,	pgamerchandiseshow.com,	planfor.fr,
plantronics.com,	pngdealers.com,	polapremium.com,
policespecial.com,	pphinfo.com,	promotersloop.com,
promusicaustralia.com,	prophecykeepers.com,	prostockcars.com,
psychprog.com,	puppyluv.com,	puppystairs.com,
q102philly.com,	qdobamail.com,	quickappointments.com,
quickertek.com,	quickfinder.com,	raleyfield.com,
raphaelsbeautyschool.edu,	rareplants.de,	rax.ru,
readingequipment.com,	realtracker.com,	rentonmclendonhardware.com,
restaurantsonlinenow.com,	resveratrol20.com,	reu.org,
revengeismydestiny.com,	ripcordarrowrest.com,	rpmrealty.com,
rrmusic.com,	rumc.com,	russellrowe.com,
russianbooks.com,	sacramentoconventioncenter.com,	salonhogar.net,
santaslodge.com,	scalemodeltoys.com,	scanner-antispyspyh4.com,
scmo.org,	scsgenealogy.com,	scottpublications.com,
sdchina.com,	search4i.com,	searchgenealogy.net,
section4wrestling.com,	seelyewrightofpawpaw.net,	seewee.net,
sheisladyboy.com,	shipleydonuts.com,	shootangle.com,
shouldersurgery.org,	simcomcity.com,	simplesignshop.com,
socalmls.com,	sohojobs.org,	southwestblend.com,
spanderfiles.com,	spatechla.com,	squireparsons.com,
srtk.net,	standup2cancer.org,	start-cleaning-business.com,
statenotary.info,	stimuluscheck.com,	stjosephccschool.net,
stmaryland.com,	storagedeluxe.com,	stranges.com,
sud.org.mx,	sudzfactory.com,	summer-glau.net,
sungardpsasp.com,	sureneeds.com,	sweetdealsandsteals.com,
sweettattianna.com,	swingstateproject.com,	syque.com,
tackletog.com,	tamusahr.com,	tasteequip.com,
tecnocino.it,	tempgun.com,	texasthunder.com,
the-working-man.com,	theacademic.org,	theacorn.com,
theauctionblock.org,	thedailymaverick.co.za,	thedigitalstory.com,
thelator.com,	thegardenhelper.com,	thegriddle.net,
thegunninghawk.com,	theinductor.com,	theliterarylink.com,
themainmarketplace.com,	themodelbook.com,	thenextgreatgeneration.com,
thepromenadebolingbrook.com,	therichkids.com,	threebarsranch.com,
thunderracing.com,	tickledpinkdesign.net,	tj9991.com,
todaywebspecial.com,	top-forum.net,	toponlinedegreechoices.com,
tracksideproductions.com,	trafficinteractive.com,	transfermarkt.de,
treadmillstore.com,	tri-une.com,	tropicalfishfind.com,
trycovermate.com,	ttsky.com,	twaa.com,
twtastebuds.com,	ualpaging.com,	uniquetruckaccessories.com,
univega.com,	unon.org,	uprius.com,
usaplforum.com,	uscoot.com,	v-picks.com,
vacuumtubeonline.com,	valueoasis.com,	vandykerifles.com,
vcbank.net,	vet4petz.com,	vidaadois.net,
videocelebs.org,	visitshenandoah.com,	vitamin-supplement-reference.com,
vitruvius.be,	walmartdrugs.net,	wcha.org,
weddingnet.org,	wefong.com,	wegotrecords.com,
weplay.com,	wetzelcars.com,	wi-fihotspotlist.com,
wiara.pl,	wildfoodadventures.com,	willyfogg.com,

Continued on Next Page...

Table D.3 – Continued

windsorhs.com, woodauto.com, woolrichhome.com, worlds.ru, ymcatriangle.org, zabaware.com,	wippit.com, woodenskis.com, worldcrops.org, wwwcoder.com, youthoutlook.org, ziua.ro	womantotal.com, woollydesigns.com, worldmapfinder.com, wxc.com, ywcahotel.com,
--	--	--

Acknowledgments

I have been fortunate to work with amazing scholars during my education. My advisor, Lorrie Faith Cranor, carefully guided my research. Lorrie is a rare advisor who understands incoming students need structure, and later-stage students need to learn how to approach research more independently. Lorrie has tremendous demands on her time, but always makes time for her students. Money was never a barrier to doing the best possible research: Lorrie found a way to fund everything I asked for. Similarly, Lorrie called me two weeks before classes started to announced she had applied for funding on my behalf, secured it, and would I be interested in being her student? This extraordinary conversation changed my life in ways I have not yet begun to understand. On an even more personal note, I am particularly grateful that Lorrie had the flexibility to let me work remotely for much of the past two years. I know it took more of her time and effort to do so. I will miss Lorrie's daily insights, wit, humor, and example of how to navigate the academic privacy sphere. It has been a privilege to be her student.

My three committee members also shaped this thesis. Alessandro Acquisti improved the opportunity cost of the time to read privacy policies material. Discussions with Alessandro formed the basis of the willingness to pay v. willingness to accept research on behavioral advertising. Alessandro is not only wise and insightful, he is one of the kindest people I have met. Deirdre K. Mulligan influenced this paper in ways she is not even aware of: I met with some of her students and walked away mulling how to make invisible privacy issues visible. Deirdre encouraged me to tell a better story of what I found, and her comments at my proposal dramatically improved my conclusions. Jon M. Peha started me on the path to the Ph.D. when he misunderstood why I was in his office and thought I was asking to be his student. Five years later I confess: no, I was just trying to figure out what EPP does. Being a Ph.D. student had never even occurred to me.

And so it has been: Jon has seen where I was going years before I have. Jon's mentorship has gone above and beyond what committee members are asked to do, while never interfering with Lorrie's role as my advisor.

And what does EPP do? The Engineering & Public Policy department is amazing. The department performs policy-relevant research on technical topics. I could not have wished for a better home. Our department head, Granger Morgan, admitted me despite my non-traditional background for a non-traditional field. EPP faculty are warm and supportive of the students and genuinely want to see us succeed. To call out just a few: H. Scott Matthews' class gave me the tools for the value of time to read privacy policies paper. Michael L. DeKay taught me to perform the survey work and statistical analysis, building upon an excellent foundation from David Banks. Francisco Veloso helped me understand cluster analysis. Mitchell Small helped me survive encountering Weibull distributions for the first time with classmates yawning at the review of basic material. Keith Florig, Elizabeth Casman, and Anny YuShang Huang made TA'ing tremendous fun.

EPP's students, and the students from the CUPS lab, helped me with every phase of this thesis. Thanks in particular to Janice Tsai who paved the way for me in so many areas. I benefited tremendously by watching Janice "go first" in nearly every area, and from her wisdom at each milestone. Robert Reeder was a delight to collaborate with, and has been a great help as I try to transition out of the Ph.D. Patrick Kelley valiantly attempted to help me move toward visual grace and saved more posters from my lack of style than I can count. Serge Egelman made every project more fun than it would have been. Robert McGuire, Keisha How, and Gregory Norcie did tremendous work from keeping the servers running to writing custom software to hand-coding tedious results. Steve Sheng, Ponnurangam Kumaraguru, and Pitikorn Tengtakul provided feedback and helpful comments on preliminary drafts of several sections. Alexandre M. Mateus, Hugo Horta, and Cristian Bravo-Lillo kept me smiling even through Pittsburgh winters.

Many thanks to Michelle McGiboney and Suzy Bausch of The Nielsen Company for assistance interpreting their research. Gregory Kesden is not even from my department, yet stepped in to make my life a better place at a crucial moment. Paul Mazaitis helped with research that is not part of this thesis but will, some day, see the light of day thanks to his efforts. Chris Hoofnagle and Ashkan Soltani were very generous with their time and experience about LSOs. Eden Fisher

supported me from the very beginning, and will never buy me another lunch or tea, not even at Amy Enrico's Tazza D'Oro where I wrote so many papers. Jazz Ayvazyan saved the day by setting up Adobe Connect for my defense. Thanks to so very many people at CDT and Adobe, particularly Deneb Meketa.

My personal views of Internet privacy are shaped by my experiences in online communities, some of whom kindly beta tested surveys. You are unhinged & wonderful people: thank you all. Particular thanks to nloof, bren, Dale, and the many denizens of thang, plus derekn and the Friday night gang for putting up with me curled up doing what turned out to be some of my most productive work. Kevin Cooney, Jay S. Laefer, and Jenny Gove worked very hard to make the Google Tech Talks possible. Jay gets special mention for keeping me saner for the past very many years. Jody Myers helped me stay organized and focused.

My family got stuck listening to me talk and talk and talk about this thesis, particularly my mother, Ruth M. Schofield. It means something different as an older student to thank your parents in a thesis: thank you. From my father, I got my love of computers. From my mother, I got my concern for people being harmed. Bart Schofield, Scott and Lisa Schofield, and Charlie and Kate Jarest got pulled into tedious discussions about my thesis when they could manage to find me. My thanks to Nameer, Jean, and May Jawdat, plus John, Pris, Adam, Julia, and soon Joanna Peterson. I thank you all for your love and support, especially as I was absent or distracted at every holiday for the past six years.

Faisal N. Jawdat, my husband, will be the happiest person to see this thesis completed. When we were getting married I lured him to CMU for 18 months. That was over six years ago. He has been my cheerleader, sounding board, and best friend throughout. Congratulations to Faisal for surviving this process. I look forward to the decades ahead when neither one of us is in school.

Thanks to the Friedman Fellowship, which supported a summer internship prior to my Ph.D. and dramatically influenced my ability to become a Ph.D. student. This thesis was supported by CyLab at Carnegie Mellon under grant DAAD19-02-1-0389 from the Army Research Office, Microsoft Research, and Adobe Systems, Inc. The Barbara Lazarus Fellowship supported my first year in the Ph.D.

Bibliography

- [1] Tracking the trackers: Our method. *Wall Street Journal* (July 31 2010).
- [2] AAAA, ANA, BBB, DMA, AND IAB. Self-regulatory program for online behavioral advertising, 2009. <http://www.iab.net/media/file/ven-principles-07-01-09.pdf> Accessed 23 January 2010.
- [3] ACKERMAN, M. S., CRANOR, L. F., AND REAGLE, J. Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM Conference on Electronic Commerce* (November 1999). <http://doi.acm.org/10.1145/336992.336995> Accessed 1 March 2010.
- [4] ACQUISTI, A., FRIEDMAN, A., AND TELANG, R. Is there a cost to privacy breaches? an event study. In *Fifth Workshop on the Economics of Information Security* (2006).
- [5] ACQUISTI, A., AND GROSS, R. Predicting social security numbers from public data. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 106, 27 (July 2009). <http://www.pnas.org/content/106/27/10975.full.pdf+html> Accessed 25 August 2010.
- [6] ACQUISTI, A., AND GROSSKLAGS, J. Privacy and rationality in individual decision making. *Security & Privacy Magazine, IEEE* 3, 1 (January-February 2005), 26–33.
- [7] ACQUISTI, A., JOHN, L., AND LOEWENSTEIN, G. What is privacy worth? Tech. rep., Heinz College, Carnegie Mellon University, 2009.
- [8] ADOBE SYSTEMS. Flash player penetration, 2010. http://www.adobe.com/products/player_census/flashplayer/ Accessed 21 August 2010.
- [9] AKERLOF, G. A. The market for “lemons”: Quality uncertainty and the market mechanism, 1970.
- [10] ANDERSON, S. House subcommittees hold joint hearing on behavioral advertising. *Security, Privacy and the Law* (July 2009). <http://www.securityprivacyandthelaw.com/2009/07/articles/recent-legislation-1/house-subcommittees-hold-joint-hearing-on-behavioral-advertising/> Original testimony available from <http://www.youtube.com/watch?v=-Wk1p2qdbmw>. Accessed 9 November 2009.
- [11] ANTON, A., EARP, J. B., QINGFENG, H., STUFFLEBEAM, W., BOLCHINI, D., AND JENSEN, C. Financial privacy policies and the need for standardization. *IEEE Security & Privacy* 2, 2 (Mar-Apr 2004), 36–45.

- [12] ANTON, A. I., EARP, J. B., AND YOUNG, J. D. How Internet users' privacy concerns have evolved since 2002. Tech. Rep. Computer Science Technical Report TR-2009-16, North Carolina State, 2009. http://theprivacyplace.org/blog/wp-content/uploads/2009/07/tr_2009_16.pdf Accessed 3 March 2010.
- [13] ARIAS, M. L. Internet law – behavioral advertising in the United States. *Internet Business Law Services* (June 2009). http://www.ibls.com/internet_law_news_portal_view.aspx?s=latestnews&id=2237 Accessed 9 November 2009.
- [14] ATTRIBUTOR. Google ad server share now at 57%. microhoo less than 15% market share., December 2008. <http://www.attributor.com/blog/google-ad-server-share-now-at-57-microhoo-less-than-15-market-share/> Accessed 3 March 2010.
- [15] AXON, S. Chrome gains, IE slumps in browser wars. *The Social Media Guide*, May 2010. <http://mashable.com/2010/05/04/chrome-firefox-ie-stats/> Accessed 15 May 2010.
- [16] BAMBERGER, K. A., AND MULLIGAN, D. K. Privacy on the books and on the ground. *Stanford Law Review* 63 (2010). UC Berkeley Public Law Research Paper No. 1568385. <http://ssrn.com/abstract=1568385> Accessed 14 May 2010.
- [17] BARBARO, M., AND ZELLER JR, T. A face is exposed for AOL searcher no. 4417749. *New York Times* (August 2006).
- [18] BARON, M. G., AND BLEKHMAN, L. Evaluating outdoor recreation parks using tcm: On the value of time. In *North American Regional Science Meeting* (Charleston, South Carolina,, January 2002). http://ie.technion.ac.il/Home/Users/mbaron/E_21_Baron-Blekhman_Jan2_2002.pdf Accessed 1 Feb 2009.
- [19] BECKER, G. S. A theory of the allocation of time. *The Economic Journal* 75, 299 (1965). <http://www.jstor.org/stable/2228949> Accessed 1 Feb 2009.
- [20] BECKER, G. S., AND MURPHY, K. M. A simple theory of advertising as a good or bad. *The Quarterly Journal of Economics* 108, 4 (1993). <http://www.jstor.org/stable/2118455> Accessed 15 February 2009.
- [21] BENDRATH, R. Icons of privacy, May 2007. <http://bendrath.blogspot.com/2007/05/icons-of-privacy.html> Accessed 22 Feb 2009.
- [22] BENDRATH, R. Global technology trends and national regulation: Explaining variation in the governance of deep packet inspection. Tech. rep., Delft University of Technology, March 2009. Paper originally prepared for the International Studies Annual Convention, New York City, 15-18 February 2009 http://userpage.fu-berlin.de/~bendrath/Paper_Ralf-Bendrath_DPI_v1-5.pdf Accessed 3 March 2010.
- [23] BETTER ADVERTISING. Better advertising: About us, 2010. http://www.betteradvertising.com/about_us.html Accessed 5 March 2010.
- [24] BLODGET, H. Complete ceo: Ips sell clickstreams for \$5 a month. Seeking Alpha, March 2007. <http://seekingalpha.com/article/29449-compete-ceo-ips-sell-clickstreams-for-5-a-month> Accessed 10 April 2007.

- [25] BOORTZ, A. R. New federal privacy bill in the works: Behavioral advertising “beneficial,” but must be done “appropriately”. *AdLaw By Request* (August 2009). <http://www.adlawbyrequest.com/2009/08/articles/legislation/new-federal-privacy-bill-in-the-works-behavioral-advertising-beneficial-but-must-be-done-appropriately/> Accessed 9 November 2009.
- [26] BOUTIN, P. Biometrics firm confirms: User counts for websites are 2-4 times too high. *VentureBeat* (February 2010). <http://venturebeat.com/2010/02/16/biometrics-firm-confirms-user-counts-for-websites-are-2-4-times-too-high/> Accessed 13 June 2010.
- [27] BOUTIN, P. Flash cookies get deleted, skew audience stats as much as 25 percent. *VentureBeat* (April 2010). <http://venturebeat.com/2010/04/14/flash-cookies-get-deleted-skew-audience-stats-as-much-as-25-percent/> Accessed 13 June 2010.
- [28] BUREAU OF LABOR STATISTICS. Table b-3. average hourly and weekly earnings of production and nonsupervisory workers on private nonfarm payrolls by industry sector and selected industry detail, 2008. <http://stats.bls.gov/news.release/empst.t16.htm> Accessed 1 February 2009.
- [29] BUSINESS WIRE. European union issues guidance on privacy notices; new notices make it easier for consumers to understand, compare policies, January 2005. <http://www.tmcnet.com/usubmit/2005/jan/1104731.htm> Accessed 19 May 2009.
- [30] CARR, D. Integrating Flash content with the HTML environment, April 2008. https://www.adobe.com/devnet/dreamweaver/articles/integrating_flash_html.html Accessed 21 August 2010.
- [31] CARVER, R. P. Is reading rate constant or flexible? *Reading Research Quarterly* 18, 2 (Winter 1983). <http://www.jstor.org/stable/747517> Accessed 8 June 2008.
- [32] CENTER FOR INFORMATION POLICY LEADERSHIP. Ten steps to develop a multilayered privacy policy, 2007. www.hunton.com/files/tbl_s47Details%5CFileUpload265%5C1405%5CTen_Steps_whitepaper.pdf Accessed 12 July 2007.
- [33] CHENG, J. Lawsuit: Disney, others spy on kids with zombie cookies. *Ars Technica* (August 16 2010). <http://arstechnica.com/tech-policy/news/2010/08/lawsuit-disney-others-spy-on-kids-with-zombie-cookies.ars> Accessed 21 August 2010.
- [34] CLELAND, S. The blind eye to privacy law arbitrage by Google – broadly threatens respect for privacy. Testimony Before the House Energy & Commerce Subcommittee On Telecommunications and the Internet. Hearing on “What Your Broadband Provider Knows About Your Web Use: Deep Packet Inspection and Communications Laws and Policies”, July 2008. netcompetition.org/Written_Testimony_House_Privacy.pdf Accessed 14 May 2010.
- [35] CLIFFORD, S. Groups far apart on online privacy oversight. *The New York Times* (December 7 2009). <http://www.nytimes.com/2009/12/08/business/media/08adco.html> Accessed 3 March 2010.
- [36] CLIFFORD, S. A little ‘i’ to teach about online privacy. *The New York Times* (January 26 2010). <http://www.nytimes.com/2010/01/27/business/media/27adco.html> Accessed 5 March 2010.

- [37] CNN MONEY.COM. Online sales spike 19 percent, May 2007. http://money.cnn.com/2007/05/14/news/economy/online_retail/index.htm Accessed 1 February 2009.
- [38] COE, R. It's the effect size, stupid: What effect size is and why its important. *Annual Conference of the British Educational Research Association* (September 2002). <http://www.leeds.ac.uk/educol/documents/00002182.htm> Accessed 27 Dec 2007.
- [39] Children's Online Privacy Protection Act of 1998 (COPPA), Public Law No. 104-191, October 1998. www.cdt.org/legislation/105th/privacy/coppa.html Accessed 27 Mar 2007.
- [40] CRANOR, L. F., EGELMAN, S., SHENG, S., McDONALD, A. M., AND CHOWDHURY, A. P3P deployment on websites. *Electronic Commerce Research and Applications* (2007).
- [41] CRANOR, L. F., GUDURU, P., AND ARJULA, M. User interfaces for privacy agents. *ACM Transactions on Computer-Human Interaction (TOCHI)* (2006).
- [42] DOWNS, J. S., HOLBROOK, M. B., SHENG, S., AND CRANOR, L. F. Are your participants gaming the system? screening mechanical turk workers. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems* (New York, NY, USA, 2010), ACM, pp. 2399-2402.
- [43] DRICHOUTIS, A. C., LAZARIDIS, P., AND RODOLFO M. NAYGA, J. Consumers' use of nutritional labels: a review of research studies and issues, 2006.
- [44] DYER, D., DALZELL, F., AND OLEGARIO, R. *Rising Tide: Lessons from 165 Years of Brand Building at Procter & Gamble*. Harvard Business Press, 2004.
- [45] ECKERSLEY, P. How unique is your browser? In *Proceedings of the Privacy Enhancing Technologies Symposium (PETS)* (2010).
- [46] The endowment effect: It's mine, I tell you. *The Economist* (June 2008). http://www.economist.com/science/displaystory.cfm?story_id=11579107 Accessed 2 September 2008.
- [47] EGELMAN, S., CRANOR, L. F., AND CHOWDHURY, A. An analysis of p3p-enabled web sites among top-20 search results. In *Eighth International Conference on Electronic Commerce* (Fredericton, New Brunswick, Canada).
- [48] EGELMAN, S., TSAI, J., CRANOR, L. F., AND ACQUISTI, A. Timing is everything? The effects of timing and placement of online privacy indicators. In *CHI 2009* (Boston, MA, USA, April 2009).
- [49] FAUL, F., ERDFELDER, E., LANG, A. G., AND BUCHNER, A. G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39 (2007), 175-191.
- [50] FEDERAL TRADE COMMISSION. In brief: The financial privacy requirements of the Gramm-Leach-Bliley Act. <http://www.ftc.gov/bcp/online/pubs/buspubs/glbshort.htm> Accessed 21 Mar 2007.
- [51] FEDERAL TRADE COMMISSION. Internet service provider settles ftc privacy charges, March 2005. <http://www.ftc.gov/opa/2005/03/cartmanager.shtm> Accessed 1 February 2008.

- [52] FEDERAL TRADE COMMISSION. Protecting consumers in the next tech-ade: A report by the staff of the federal trade commission, March 2008. <http://www.ftc.gov/os/2008/03/P064101tech.pdf> Accessed 1 February 2009.
- [53] FEDERAL TRADE COMMISSION. FTC staff revises online behavioral advertising principles, February 2009. <http://www.ftc.gov/opa/2009/02/behavad.shtm> Accessed 15 May 2009.
- [54] FEDERAL TRADE COMMISSION. Sears settles FTC charges regarding tracking software, June 2009.
- [55] FEDERAL TRADE COMMISSION. Self-regulatory principles for online behavioral advertising: Tracking, targeting, and technology. Staff Report, February 2009. <http://www.ftc.gov/os/2009/02/P085400behavadreport.pdf> Accessed 9 November 2009.
- [56] GARFINKEL, S. *Database nation: the death of privacy in the 21st century*. O'Reilly & Associates, Inc., Sebastopol, CA, 2001.
- [57] GOMEZ, J., PINNICK, T., AND SOLTANI, A. Knowprivacy, June 2009. http://www.knowprivacy.org/report/KnowPrivacy_Final_Report.pdf Accessed 4 March 2010.
- [58] GOOD, N., DHAMIJA, R., GROSSKLAGS, J., THAW, D., ARONOWITZ, S., MULLIGAN, D., AND KONSTAN, J. Stopping spyware at the gate: a user study of privacy, notice and spyware. *SOUPS 2005: Proceedings of the 2005 symposium on Usable privacy and security* (2005), 43–52.
- [59] GRABER, M. A., D'ALESSANDRO, D. M., AND JOHNSON-WEST, J. Reading level of privacy policies on internet health web sites. *Journal of Family Practice* (July 2002).
- [60] U.S. Gramm-Leach-Bliley Financial Modernization Act of 1999, Public Law no. 106–102, November 1999.
- [61] GROSSKLAGS, J., AND ACQUISTI, A. When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. In *Workshop on the Economics of Information Security (WEIS)* (2007).
- [62] GUYTON, J. L., KOROBOW, A. K., LEE, P. S., AND TODER, E. J. The effects of tax software and paid preparers on compliance costs. *National Tax Journal* 58, 3 (2005).
- [63] HA, V., INKPEN, K., AL SHAAR, F., AND HDIEB, L. An examination of user perception and misconception of Internet cookies. In *CHI '06 Extended Abstracts of Human Factors in Computing Systems* (April 2006). <http://doi.acm.org/10.1145/1125451.1125615> Accessed 1 March 2010.
- [64] HARPER, S. R., AND KUH, G. D. Myths and misconceptions about using qualitative methods in assessment. In *Using qualitative methods in institutional assessment. New Directions for Institutional Research*, S. R. Harper and S. D. Museus, Eds., no. 136. Jossey-Bass, San Francisco, 2007, pp. 5–14.
- [65] HOCHHAUSER, M. Lost in the fine print: Readability of financial privacy notices, July 2001. <http://www.privacyrights.org/ar/GLB-Reading.htm> Accessed 27 Mar 2007.
- [66] HONG, J. I., NG, J. D., LEDERER, S., AND LANDAY, J. A. Personal privacy through understanding and action: Five pitfalls for designers. In *DIS '04: Proceedings of the 5th conference on Designing interactive systems* (New York, NY, USA, 2004), ACM, pp. 91–100. <http://doi.acm.org/10.1145/1013115.1013129> Accessed 7 Nov 2008.

- [67] HOOFNAGLE, C. J., AND KING, J. What Californians understand about privacy online, 2008 September. <http://ssrn.com/abstract=1262130> Accessed 7 Nov 2008.
- [68] HOOFNAGLE, C. J., KING, J., LI, S., AND TUROW, J. How different are young adults from older adults when it comes to information privacy attitudes and policies? <http://ssrn.com/abstract=1589864> Accessed 25 August 2010, April 2010.
- [69] HUANG, H.-J. Language-focus instruction in EFL writing : Constructing relative clauses in definition paragraphs. In *2008 International Conference on English Instruction and Assessment* (2008). <http://www.ccu.edu.tw/fllcccu/2008EIA/English/C16.pdf> Accessed 22 Feb 2009.
- [70] INITIATIVE, N. A. Faq, 2010. http://www.networkadvertising.org/managing/faqs.asp#question_19 Accessed 5 March 2010.
- [71] INTERACTIVE ADVERTISING BUREAU. Internet advertising revenues top \$21 billion in '07, reaching record high. news release, May 2008. http://www.iab.net/about_the_iab/recent_press_releases/press_release_archive/press_release/299609 Accessed 1 February 2009.
- [72] JENSEN, C., AND POTTS, C. Privacy policies examined: Fair warning or fair game? *GVU Technical Report 03-04* (February 2003). <ftp://ftp.cc.gatech.edu/pub/gvu/tr/2003/03-04.pdf> Accessed 1 January 2008.
- [73] JENSEN, C., AND POTTS, C. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2004), ACM, pp. 471–478.
- [74] JENSEN, C., POTTS, C., AND JENSEN, C. Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies* 63 (July 2005), 203–227.
- [75] KANTOR, A. Aol search data release reveals a great deal. USA Today, August 2006. http://www.usatoday.com/tech/columnist/andrewkantor/2006-08-17-aol-data_x.htm Accessed 1 Feb 2007.
- [76] KAY, M., AND TERRY, M. Textured agreements: Re-envisioning electronic consent. Technical report cs-2009-19, David R. Cheriton School of Computer Science, University of Waterloo, 2009.
- [77] KAY, M., AND TERRY, M. Communicating software agreement content using narrative pictograms. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems (CHI)* (2010), pp. 2715–2724.
- [78] KAYE, K. Yahoo opens behavioral ad data curtain, December 2010. <http://www.clickz.com/3635839> Accessed 5 March 2010.
- [79] KELLEY, P. G. Conducting usable privacy & security studies with amazon’s mechanical turk. In *Symposium on Usable Privacy and Security (SOUPS)* (Redmond, WA, July 2010).
- [80] KELLEY, P. G., BRESEE, J., REEDER, R. W., AND CRANOR, L. F. A “nutrition label” for privacy. In *Symposium on Usable Privacy and Security, (SOUPS)* (2009).
- [81] KITTUR, A., CHI, E. H., AND SUH, B. Crowdsourcing user studies with mechanical turk. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (2008), ACM, pp. 453–456. <http://doi.acm.org/10.1145/1357054.1357127> Accessed 5 March 2010.

- [82] KMETOVICZ, R. E. *New Product Development: Design and Analysis*. Wiley-IEEE, New York, 1992.
- [83] KOSARA, R., AND ZIEMKIEWICZ, C. Do mechanical turks dream of square pie charts? In *BELIV '10: BEyondtime and errors: Novel evaLuation methods for Information Visualization* (2010), ACM, pp. 373–382.
- [84] KRISTOL, D., AND MONTULLI, L. RFC2109 - HTTP state management mechanism. <http://www.faqs.org/rfcs/rfc2109.html> Accessed 10 May 2010.
- [85] LAUDON, K. C. Markets and privacy. *Communications of the ACM* 39, 9 (1996), 96.
- [86] LEMOS, R. MSN sites get easy-to-read privacy label. *CNET News.com* (2005). http://news.com.com/2100-1038_3-5611894.html Accessed 30 May 2007.
- [87] LENHART, A., AND MADDEN, M. Teens, privacy & online social networks. Pew Internet & American Life Project, April 2007.
- [88] LEON, P. G., CRANOR, L. F., McDONALD, A. M., AND MCGUIRE, R. Token attempt: The misrepresentation of website privacy policies through the misuse of P3P compact policy tokens. In *Proceedings of the 9th Workshop on Privacy in the Electronic Society (WPES)* (October 2010).
- [89] LEUNIG, T. Time is money: A re-assessment of the passenger social savings from victorian british railways. *The Journal of Economic History* (2006). Working paper. <http://www.lse.ac.uk/collections/economicHistory/pdf/LSTC/0905Leunig.pdf> Accessed 1 February 2009.
- [90] LUNNEY, G. H. Using analysis of variance with a dichotomous dependent variable: An empirical study. *Journal of Educational Measurement* 7, 4 (Winter 1970), 263–269.
- [91] MANCHANDA, P., DUBE, J.-P., GOH, K. Y., AND CHINTAGUNTA, P. K. The effect of banner advertising on internet purchasing. *Journal of Marketing Research XLIII* (February 2006), 98–108.
- [92] MARSH, JR., R. M. Legislation for effective self-regulation: A new approach to protecting personal privacy on the internet. *Mich. Telecomm. Tech. L. Rev* 543 (2009), 543–563. <http://www.mttlr.org/volfifteen/marsh.pdf>.
- [93] MCCARTHY, J. Truste decides its own fate today. *Slashdot* (November 1999). <http://slashdot.org/yro/99/11/05/1021214.shtml> Accessed 1 Jun 2008.
- [94] McDONALD, A., AND CRANOR, L. F. An empirical study of how people perceive on-line behavioral advertising. Tech. Rep. CyLab Technical Report 09-015, Carnegie Mellon, November 2009. http://www.cylab.cmu.edu/research/techreports/tr_cylab09015.html.
- [95] McDONALD, A., AND CRANOR, L. F. Cookie confusion: Do browser interfaces undermine understanding? *CHI 2010: Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (2010). To appear.
- [96] McDONALD, A. M., AND CRANOR, L. F. The cost of reading privacy policies. *I/S - A Journal of Law and Policy for the Information Society* 4, 3 (2008). <http://www.is-journal.org/V04I03/McDonald.pdf> Accessed 22 April 2010.

- [97] McDONALD, A. M., AND CRANOR, L. F. Americans' attitudes about internet behavioral advertising practices. In *Proceedings of the 9th Workshop on Privacy in the Electronic Society (WPES)* (October 4 2010).
- [98] McDONALD, A. M., AND CRANOR, L. F. Beliefs and behaviors: Internet users' understanding of behavioral advertising. In *38th Research Conference on Communication, Information and Internet Policy (Telecommunications Policy Research Conference)* (October 2 2010).
- [99] McDONALD, A. M., REEDER, R. W., KELLEY, P. G., AND CRANOR, L. F. A comparative study of online privacy policies and formats. In *Privacy Enhancing Technologies Symposium (PETS)* (August 5–7 2009).
- [100] MILLER, J. I. Note: "don't be evil": Gmail's relevant text advertisements violate google's own motto and your e-mail privacy rights. *Hofstra Law Review* 1607 (Summer 2004).
- [101] MORGAN, M. G., FISCHHOFF, B., BOSTROM, A., AND ATMAN, C. J. *Risk Communication: A Mental Models Approach*. Cambridge University Press, 2002.
- [102] MULLINS, R. Privacy group argues buzz breaks wiretap laws. *VentureBeat* (February 2010). <http://venturebeat.com/2010/02/17/privacy-group-argues-buzz-breaks-wiretap-laws/> Accessed 14 May 2010.
- [103] MY BYLINE MEDIA. The Flesch reading ease readability formula. <http://www.readabilityformulas.com/flesch-reading-ease-readability-formula.php> Accessed 9 Mar 2009.
- [104] NIE, N. H. Ten years after the birth of the internet: How do americans use the internet in their daily lives? Faculty Working Paper Stanford Institute for the Quantitative Study of Society, 2005. http://www.stanford.edu/group/siqss/research/time_study_files/ProjectReport2005.pdf Accessed 1 February 2009.
- [105] NIELSEN/NET RATINGS. Nielsen online reports topline u.s. data for march 2008. news release, April 14 2008. http://www.nielsen-online.com/pr/pr_080414.pdf Accessed 26 February 2009.
- [106] NIELSEN/NET RATINGS. Internet audience metrics, united states, February 2009. http://www.netratings.com/resources.jsp?section=pr_netv&nav=1 Accessed 26 February 2009 (site now updated with data reflecting the present time period).
- [107] OHM, P. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review (forthcoming)* (2010). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006 Accessed 10 November 2009.
- [108] ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. Oecd guidelines on the protection of privacy and transborder flows of personal data. http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html Accessed 1 Feb 2009.
- [109] OUT-LAW NEWS. Drop the jargon from privacy policies, says privacy chief, September 2005. <http://www.out-law.com/page-5791> Accessed 23 Mar 2007.
- [110] PETTY, R. D. Marketing without consent: Consumer choice and costs, privacy, and public policy. *Journal of Public Policy and Marketing* 19, 1 (Spring 2000).

- [111] PEW INTERNET & AMERICAN LIFE PROJECT. Demographics of internet users: Who's online, 2006. http://www.pewinternet.org/trends/User_Demo_4.26.07.htm Accessed 5 June 2007.
- [112] PEW INTERNET & AMERICAN LIFE PROJECT. Report: Internet, broadband, and cell phone statistics, 2009. <http://www.pewinternet.org/Reports/2010/Internet-broadband-and-cell-phone-statistics.aspx?r=1> Accessed 26 July 2010.
- [113] PITOFISKY, R. "self-regulation and privacy online" (prepared statement before the senate committee on commerce, science, and transportation, July 1999. <http://www.ftc.gov/os/1999/07/privacyonlinetestimony.pdf> Accessed 1 Feb 2009.
- [114] PITOFISKY, R. "privacy online: Fair information practices in the electronic marketplace" (prepared statement before the senate committee on commerce, science, and transportation), May 2000. <http://www.ftc.gov/os/2000/05/testimonyprivacy.htm> Accessed 1 Feb 2009.
- [115] POLLACH, I. What's wrong with online privacy policies? *Communications of the ACM* 30, 5 (September 2007), 103–108.
- [116] PRESCOTT, L. OS X share up 29% in past year, slowly chipping away at Microsoft. VentureBeat, February 2010. <http://venturebeat.com/2010/02/26/os-x-share-up-29-in-past-year-slowly-chipping-away-at-microsoft/> Accessed 15 May 2010.
- [117] PRIVACY GOURMET. Key findings of 2008 aol consumer survey on behavioral advertising, April 2008. <http://www.privacygourmet.co.uk/2008/04/key-findings-of.html> Accessed 25 Nov 2008.
- [118] REEDER, R. W., KELLEY, P. G., McDONALD, A. M., AND CRANOR, L. F. A user study of the expandable grid applied to P3P privacy policy visualization. In *WPES '08: Proceedings of the 7th ACM workshop on Privacy in the electronic society* (2008), ACM, pp. 45–54. <http://portal.acm.org/citation.cfm?id=1456403.1456413#> Accessed 22 Feb 2009.
- [119] REPORT BY KLEIMANN COMMUNICATION GROUP FOR THE FTC. Evolution of a prototype financial privacy notice, 2006. <http://www.ftc.gov/privacy/privacyinitiatives/ftcfinalreport060228.pdf> Accessed 2 Mar 2007.
- [120] RODGERS, Z. Iab tackles privacy worries with big display ad campaign, December 2009. <http://www.clickz.com/3635817> Accessed 5 March 2010.
- [121] ROMANOSKY, S., TELANG, R., AND ACQUISTI, A. Do data breach disclosure laws reduce identity theft? In *Seventh Workshop on the Economics of Information Security* (June 2008).
- [122] ROSS, J., ZALDIVAR, A., IRANI, L., AND TOMLINSON, B. Who are the turkers? worker demographics in amazon mechanical turk. Technical report socialcode-2009-01, Univeristy of California, Irvine, 2009.
- [123] RYKER, R., LAFLEUR, E., MCMAINIS, B., AND COX, K. C. Online privacy policies: An assessment of the fortune e-50. *JOURNAL OF COMPUTER INFORMATION SYSTEMS* 42, 4 (Summer 2002), 15–20.
- [124] SALTZER, J. H., AND SCHROEDER, M. D. The protection of information in computer systems. *Proceedings of the IEEE* 63 (September 1975), 1278–1308.

- [125] SAVAGE, S. J., AND WALDMAN, D. Broadband internet access, awareness, and use: Analysis of united states household data. *Telecommunications Policy* 29, 8 (2005).
- [126] SCHONFELD, E. Is beacon inflating facebook's visitor numbers? *Tech Crunch* (December 2007). <http://techcrunch.com/2007/12/06/is-beacon-inflating-facebooks-visitor-numbers/> Accessed 3 March 2010.
- [127] SHENG, X., AND CRANOR, L. F. An evaluation of the effect of US financial privacy legislation through the analysis of privacy policies. *I/S - A Journal of Law and Policy for the Information Society* 2, 3 (Fall 2006), 943–980.
- [128] SINGEL, R. Privacy lawsuit targets 'net giants over "zombie" cookies. *Ars Technica* (July 2010). <http://arstechnica.com/tech-policy/news/2010/07/privacy-lawsuit-targets-net-giants-over-zombie-cookies.ars> Accessed 22 August 2010.
- [129] SOLTANI, A., CANTY, S., MAYO, Q., THOMAS, L., AND HOOFNAGLE, C. J. Flash cookies and privacy, August 11 2009. <http://ssrn.com/abstract=1446862> Accessed 15 Apr 2010.
- [130] STAMPLEY, D. A. Managing information technology security and privacy compliance, 2005. <http://www.neohapsis.com/utility/NeoPrivacyWhitepaper.pdf> (linked to as "Privacy Compliance") Accessed 1 February 2008.
- [131] STONE, B. Ads posted on facebook strike some as off-key. *The New York Times* (March 3 2010). <http://www.nytimes.com/2010/03/04/technology/04facebook.html> Accessed 5 March 2010.
- [132] STORY, L. How many site hits? depends who's counting. *New York Times* (October 2007).
- [133] SULLIVAN, B. Mobile web best practices 2.0: Basic guidelines, W3C editor's draft, March 2008. <http://www.w3.org/2005/MWI/BPWG/Group/Drafts/BestPractices-2.0/ED-mobile-bp2-20080327#bp-cookies-recover> Accessed 21 September 2010.
- [134] SUTTE, J. D. Some quitting facebook as privacy concerns escalate. *CNN Tech* (May 2010). <http://www.cnn.com/2010/TECH/05/13/facebook.delete.privacy/index.html?iref=allsearch> Accessed 15 May 2010.
- [135] SWEENEY, L. /emphk-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems* 10, 5 (2002), 557–570.
- [136] SZOKA, B. M. Privacy polls v. real-world trade-offs. *The Progress & Freedom Foundation Progress Snapshot Paper* 5, 10 (November 2009). <http://ssrn.com/abstract=1502811> Accessed 4 March 2010.
- [137] THE CENTER FOR INFORMATION POLICY LEADERSHIP, H. . W. L. Multi-layered notices. <http://www.hunton.com/Resources/Sites/general.aspx?id=328> Accessed 23 Mar 2007.
- [138] THE OFFICE OF THE PRIVACY COMMISSIONER. Release of privacy impact assessment guide and layered privacy policy, August 2006. http://www.privacy.gov.au/news/06_17.html Accessed 22 Feb 2009.
- [139] THEURER, T. Performance research, part 3: When the cookie crumbles. *Yahoo! User Interface Blog* (March 2007). <http://yuiblog.com/blog/2007/03/01/performance-research-part-3/> Accessed 10 May 2010.

- [140] THOMPSON, R. Minimizing liability and productivity risks: How to control the impacts of spyware, hacker tools and other harmful applications. Computer Associates, October 2004. <http://www.ameinfo.com/pdfdocs/51515.pdf> 3 February 2009.
- [141] TREVELYAN, M. Stolen account prices fall as market flooded. [news.com.au](http://www.news.com.au/technology/story/0,25642,24023758-5014111,00.html), July 2008. <http://www.news.com.au/technology/story/0,25642,24023758-5014111,00.html> Accessed 1 Feb 2009.
- [142] TRUSTE. Truste program requirements. <http://www.truste.org/requirements.php> Accessed 19 Jan 2009.
- [143] TRUSTE. 2008 study: Consumer attitudes about behavioral targeting, March 2008. http://danskprivacynet.files.wordpress.com/2009/02/truste2008_tns_bt_study_summary1.pdf Accessed 9 November 2009.
- [144] TRUSTE. Truste launches pilot of behavioral advertising notice and choice program, January 2010. http://www.truste.com/about_TRUSTE/press-room/news_truste_oba_pilot_announcement.html Accessed 5 March 2010.
- [145] TRUSTE AND TNS. 2009 study: Consumer attitudes about behavioral targeting, March 2009.
- [146] TSAI, J., EGELMAN, S., CRANOR, L. F., AND ACQUISTI, A. The effect of online privacy information on purchasing behavior: An experimental study. In *The 6th Workshop on the Economics of Information Security (WEIS)* (2008). <http://weis2007.econinfosec.org/papers/57.pdf> Accessed 22 Feb 2009.
- [147] TUROW, J. Americans & Online Privacy: The System is Broken. Annenberg Public Policy Center Report, 2003.
- [148] TUROW, J., KING, J., HOOFNAGLE, C. J., BLEAKLEY, A., AND HENNESSY, M. Americans reject tailored advertising and three activities that enable it. Tech. rep., Annenberg School for Communications, University of Pennsylvania, September 2009. http://repository.upenn.edu/asc_papers/137/ Accessed 4 March 2010.
- [149] UNITED NATIONS GENERAL ASSEMBLY. Universal declaration of human rights, art. 12, 1948. <http://www.unhchr.ch/udhr/lang/eng.pdf> Accessed 1 Feb 2009.
- [150] U.S. CENSUS BUREAU. U.S. Census Bureau population and household economic topics, census 2000, 2000.
- [151] U.S. CENSUS BUREAU, POPULATION DIVISION. Annual estimates of the resident population by sex and five-year age groups for the united states: April 1, 2000 to july 1, 2009 (nc-est2009-01), June 2010.
- [152] VARIAN, H. R. Economic aspects of personal privacy. Faculty Working Paper, Department of Economics, Univ. of California at Berkeley, 1996. <http://people.ischool.berkeley.edu/~hal/Papers/privacy> Accessed 1 February 2009.
- [153] VILA, T., GREENSTADT, R., AND MOLNAR, D. Why we can't be bothered to read privacy policies: models of privacy economics as a lemons market. *ACM International Conference Proceeding Series* 5 (2003), 403–407.
- [154] W3C WORKING GROUP. The platform for privacy preferences 1.1 (P3P1.1) specification, November 2006. <http://www.w3.org/TR/P3P11/> Accessed 28 Mar 2007.

- [155] WEINREICH, H. Not quite the average: An empirical study of web use. *ACM Transactions on the Trans Web* 2, 1 (February 2008).
- [156] WINGFIELD, N. Microsoft quashed effort to boost online privacy. *Wall Street Journal* (August 2 2010). http://online.wsj.com/article/SB10001424052748703467304575383530439838568.html?mod=ITP_pageone_0 Accessed 15 August 2010.