Carnegie Mellon University

CARNEGIE INSTITUTE OF TECHNOLOGY

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF Doctor of Philosophy

TITLE **Privacy Notice and Choice in Practice**

PRESENTED BY **Pedro Giovanni Leon-Najera**

ACCEPTED BY THE DEPARTMENT OF

Engineering and Public Policy

Lorrie Cranor ADVISOR, MAJOR PROFESSOR	September 29, 2014 DATE
Douglas Sicker DEPARTMENT HEAD	September 29, 2014 DATE

APPROVED BY THE COLLEGE COUNCIL

Vijayakumar Bhagavatula DEAN	October 10, 2014 DATE

Privacy Notice and Choice in Practice

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

Pedro Giovanni Leon-Najera
B.S., Telecommunications Engineering, Universidad Nacional Autónoma de México
M.S., Information Security Technology and Management, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA
December, 2014

The views and conclusions contained in this document are those of the author, and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government, or any other entity.

Keywords: behavioral advertising, human-computer interactions, online tracking, privacy choice, privacy disclosure, privacy notice, privacy nudges, privacy regulation, standardized disclosures, targeted transparency, usability, user interface.

Abstract

In the United States, notice and choice remain the most commonly used mechanisms to protect people's privacy online. This approach relies on the assumption that users provided with notice will make informed choices that align with their privacy expectations.

The goal of this research is to empirically inform industry and regulatory efforts that rely on notice and choice to protect people's online privacy. To do so, we present a set of case studies covering different aspects of privacy notice and choice in four domains: online behavioral advertising (OBA), online social networks (OSN), financial privacy notices, and websites' machine-readable privacy notices.

We investigate users' privacy preferences, information needs, and ability to exercise choices in the OBA domain. Based on our results, we provide recommendations to improve the design of notice and choice methods currently in use in this domain. In the context of OSNs, we explore the effect of nudging notices designed to encourage more thoughtful disclosures among Facebook users and recommend changes to the Facebook user interface aimed to mitigate problematic disclosures. We demonstrate how standardized notices enable large-scale evaluations and comparisons of companies' privacy practices and argue that standardized privacy notices have an enormous potential to improve transparency and benefit users, privacy-respectful companies, and oversight entities.

We argue that, in today's complex Internet ecosystem, an approach that relies on users to make privacy decisions should also empower them with user-friendly interfaces, relevant information, and the tools they need to make privacy decisions. Finally, we further argue that notice and choice are necessary, but not sufficient to protect online privacy, and that government regulation is necessary to establish necessary additional protections including access, redress, accountability, and enforcement.

To my wife Alejandra and my son Josué, without whom this would not have been possible;

to my father Mario and my mother María Antonia, to whom I owe everything;

and to my sister Bella Saraí and my brother Yonathan, who are the reason I started this.

Acknowledgments

I could not have asked for a better adviser than Dr. Lorrie Faith Cranor. She knows the right doses of advice to give throughout the Ph.D. process. Despite her busy schedule, she was always available to help me. She has made the CUPS Lab a friendly and collaborative work environment. To me, she has been both an academic and parenting mentor. I thank her for having contributed to a life-changing and rewarding experience.

My thesis committee helped me to improve this work. Dr. Alessandro Acquisti encouraged me to see my research through the lens of behavioral economics. It was a privilege to collaborate with him on different projects, some of which are part of this thesis. Dr. Joel Reidenberg helped me to refine the legal accuracy of my claims and provided constant feedback throughout. Dr. Jon Peha encouraged me to think more deeply about the public policy implications of this thesis.

Dr. Granger Morgan has made EPP a unique place to do research that informs policy making. I thank the EPP staff, especially Vicky Finney, for always being there to help me. I thank my EPP classmates, Rebecca Balebako, Cristian Bravo, Frankie Catota, Alfredo Galván, Juan Roldán, and Mili Tamayao for their friendship and constant encouragement. You made EPP a great place to be part of.

I was fortunate to work with many talented people who helped me to produce more and better quality research. I particularly thank my co-authors Alyssa Au, Alessandro Acquisti, Rebecca Balebako, Lujo Bauer, Lorrie Cranor, Justin Cranshaw, Mihai Christodorescu, Alain Forget, Jim Graves, Manoj Hastak, Candice Hoke, Kelly Idouchi, Alfred Kobsa, Saranga Komanduri, Abigail Marsh, Robert McGuire, Aleecia McDonald, Carolyn Nguyen, Ashwini Rao, Norman Sadeh, Richard Shay, Florian Schaub, Manya Sleeper, Blase Ur, Yang Wang, and Guzi Xu.

I further thank my friends Robert Gormley, Cindy Gormley, Matthew Gormley, Candice Gormley, Jessica Perez, Emanuel Perez, William Griffith, Jeanne Griffith, Emily Griffith, Timothy Griffith, Kuo Chiang and Yi Fen Yeh (Emma) for their selfless help. Thank you for embracing my family as your own family and for always making us feel at home in Pittsburgh.

Last, but not least, I want to thank my wife Alejandra Penilla, my son Josué León, my father Mario León, my mother Maria Antonia Nájera, my sister Bella León, and my brother Yonathan León. Thank you for your tolerance and emotional support. You are my everyday motivation and this would not have been possible without you.

This research was funded in part by a grant from The Privacy Projects, the IWT SBO Project on Security and Privacy for Online Social Networks (SPION), a Google focused research award on privacy nudges, and by NSF grants DGE0903659, CNS1012763, and CNS0831428. I was partially supported by the National Council on Science and Technology (CONACYT), a Carnegie Institute of Technology Dean's fellowship, and a Bertucci's graduate fellowship.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The U.S. Government has recognized both benefits and privacy risks of new online business models enabled by the Internet [1]. Current proposals from both regulators and industry to protect people's online privacy are still developing and can be informed by the empirical research that we present in this work. Currently, most privacy-protection methods rely on the principles of notice and choice, requiring users to make privacy choices informed by companies' privacy disclosures. However, very little research has been done about how companies implement notice and choice mechanisms and how effective those mechanisms are from a user's standpoint. This is the focus of our research, which within the scope of four application domains and following an empirical approach, provides recommendations to further the design of privacy protections based on notice and choice.

The remaining of this Chapter provides background on the legal framework in the United States to protect privacy and discusses the current state of privacy self-regulation based on notice and choice. Within this context, we then introduce the scope and goals of this research. Finally, we provide a roadmap of the thesis.

1.1 U.S. Privacy Regulation Framework

The United States follows a sectorial approach to protect privacy with existing federal and state laws that protect personal information in specific domains. Notable examples of industry-specific federal laws include the Fair Credit Reporting Act (FCRA), the Gramm-Leach Bliley Act (GLBA), the Health Insurance Portability and Accountability Act (HIPAA), and the Family Educational Rights and Privacy Act (FERPA). In addition, the Electronic Communications Privacy Act (ECPA) protects electronic communications from wiretapping, and the Video Privacy Protection Act (VPPA) protects video rental records from unauthorized disclosures. Swire et al. provide a comprehensive list of current U.S. privacy-related laws [2].

With a few exceptions, including the Children's Online Privacy Protection Act (COPPA), which protects privacy of Internet users under the age of 13; and the California Online Privacy Protection Act of 2003 (CalOPPA), which mandates the placement of privacy policies on websites that collect personal information; explicit laws to protect online privacy in the U.S. do not exist. Nevertheless, the Federal Trade Commission (FTC) has played an important role in protecting consumers' online privacy over the last decade. Under section 5 of the FTC Act of 1914, which empowers the FTC to prevent "unfair or deceptive acts or practices in or affecting commerce," the FTC has used its authority to take action against companies with deceptive privacy policies, companies that have failed to protect consumers personal information, and companies that have taken actions affecting consumers personal information. Examples include charges against Google,[1] Facebook,[2] and online advertisers,[3] among others.[4] In addition, the FTC has been active organizing privacy round tables and workshops, and has released reports and guidelines intended to improve transparency of companies' practices to protect users' privacy [3,4,5].

1.2 Transparency-based Regulations and Privacy Self-regulation

Transparency is a common regulation mechanism in use by governments. In the U.S. many laws require both government and private companies to publicly disclose information that is believed to affect people's welfare and decisions. For example, in the financial domain, the Securities and Exchanges Acts of 1933 and 1934 require publicly traded companies to disclose their financial statements to protect investors [6]. In the public health domain, the Nutrition Labeling and Education Act of 1990 requires the use of nutrition labels on products to empower users to make healthier choices and reduce food-related chronic diseases [7]. Similarly, 47 U.S. states have enacted data-breach notification laws with the goal of protecting people's personal information [8]. Fung et al. provide a comprehensive review of both national and international transparency-based regulations [9].

The overall goal of transparency-based regulations is to create an incentive for information disclosers to improve their practices in a manner that better serves the public interest. Specifically, the assumption is that people informed with relevant information and provided with viable alternatives will select those alternatives that better align with their expectations, incentivizing information disclosers to adjust their practices accordingly [9].

In the U.S., both industry and government have strongly relied on the idea of providing notice and choice to Internet users as a mechanism to protect online privacy. Furthermore, industry self-regulates, designing and implementing its own notice and choice mechanisms. Reliance on self-regulation has been justified based on market efficiency grounds. From an economics' perspective, the assumption is that informed users will choose to deal with companies with privacy practices that align with their privacy expectations [10] and users will choose companies with good privacy practices while rejecting those with bad ones, leading the market to an optimal level of privacy [11]. Therefore, industry advocates claim that self-regulation is a cost-effective solution because it achieves the goal of protecting privacy without requiring costly government interventions. From a pragmatic perspective, an argument is often made that privacy is a subjective, changing, and context-dependent concept [12, 13, 14], making it unfeasible for laws to fully capture everyone's expectations of privacy. Industry advocates further argue that restricting the free flow of information is not only impossible, but will stifle innovation with negative consequences for the U.S. economy [15].

[1] http://www.ftc.gov/opa/2012/08/google.shtm; http://www.ftc.gov/opa/2011/03/google.shtm

[2] http://ftc.gov/opa/2011/11/privacysettlement.shtm

[3] http://www.ftc.gov/opa/2011/11/scanscout.shtm

[4] http://www.ftc.gov/opa/reporter/privacy/privacypromises.shtml

1.3 The Need for Improved Notice and Choice

Self-regulation efforts have been ineffective due to the lack of enforcement [16, 17, 18]. Empirical evidence further suggests that privacy policies, a widely used notice mechanism, are largely ineffective. In particular, Internet users don't like reading privacy policies or terms of services [10, 19], they don't understand them [20], and misunderstand their purposes [21]. Furthermore, behavioral decision research has found that human beings are subject to cognitive limitations and behavioral biases that prevent them from making optimal decisions [22, 23, 24]. Privacy scholars have attributed to these biases the disconnection between privacy intentions and privacy behaviors, a phenomenon termed "the privacy paradox" [25].

Problems with the implementation of notice mechanisms have resulted in significant information asymmetries between users and companies. Specifically, while companies collecting users' information know exactly what information is collected, how they use it, and how much it is worth, users are often required to make privacy decisions in a vacuum of relevant information.

These problems in combination with technological advances in collection, aggregation, sharing, and dissemination of Internet users' information have incentivized U.S. policy makers to consider the creation of statutory privacy protection requirements. In particular, an area that has received a lot of attention from regulators and media is online behavioral advertising (OBA), the practice of tracking an individual's online activities in order to deliver advertising tailored to the individual's interests [3]. A number of privacy bills have been introduced by representatives of the U.S. Congress in the last few years.[5]

In February of 2012 the White House introduced a Privacy Bill of Rights asking online companies to implement enforceable codes of conduct based on the FIPPs, including individual control, transparency, access, security, and accountability as essential privacy enhancing principles [1]. Similarly, the FTC released a report that contemplates regulation and enforceable self-regulation [5]. While the enactment of government legislation that considers enforcement mechanisms can incentivize companies to comply with established requirements, the problem of designing effective notice and choice remains unaddressed. In particular, notice and choice can only deliver the expected benefits if they provide relevant information and meaningful choices for users.

1.4 Research Questions and Scope

The overall goal of this thesis is to inform with empirical research the design of privacy notice and choice mechanisms that can better reach Internet users. In this section we state our research questions and discuss the scope and methods used to answer them. The main research questions addressed in this thesis are shown in Table 1.1. To answer these questions we employed various research methods and covered different aspects of privacy notice and choice in four domains: websites' machine-readable privacy notices, financial institutions' standardized notices, online behavioral advertising (OBA), and online social networks (OSN). We selected these domains for the following reasons:

[5] Some examples are: Do Not Track Me Online Act of 2011; Do Not Track Kids Act of 2011; Best Practices Act of 2010; Commercial Privacy Bill of Rights Act of 2011; Consumer Privacy Protection Act of 2011

Websites' privacy policies. Websites are the main method to access the Internet and therefore users interact with them every day. In addition, many websites use the Platform for Privacy Preferences (P3P), a machine-readable standard developed by the industry, to communicate privacy policies. An industry-developed, machine-readable standard, allows for a large-scale evaluation of privacy policies under an industry self-regulated environment.

Financial institutions' privacy notices. By their nature, financial institutions collect sensitive information from users. Furthermore, the fact that many financial institutions have been incentivized to implement online standardized notices, provides a unique opportunity to evaluate, at large-scale, the privacy notices of companies operating under a government-regulated environment.

Online behavioral advertising (OBA). The practice of tracking users online activities has important consequences for privacy. As a result, OBA has triggered most of the privacy debate over the last several years and regulators have urged the advertising industry to develop better privacy protections. Furthermore, surveys of consumers consistently report that users are concerned about being tracked online and current privacy protections developed by the ad industry mostly rely on notice and choice.

Online social networks (OSN). In particular, we study Facebook, the most popular social network with more than 1.2 billion users worldwide. Media stories and research, including ours, have shown that Facebook users struggle to protect their privacy. Therefore, exploring mechanisms to assist these users with their online disclosures is a goal that is worth pursuing.

Research Question	Thesis Chapter
Q1. What is the effect of industry self-regulation and government-regulation on companies' implementations of privacy notices?	Chapters 2, 3, and 4
Q2. What are the strengths, weaknesses, and opportunities for improvement of current online notice and choice mechanisms?	Chapters 2, 3, 4, 5, 6, 7, and 8
Q3. Can contextual, in-time notices mitigate users' regrettable information disclosures?	Chapter 9

Table 1.1: Mapping between research questions and thesis' chapters

Table 1.2 shows the relationship between the covered application domains, research methods used, and our research questions.

1.5 Thesis Roadmap

In this chapter we have provided an overview of the strategy to protect privacy in the United States and its relationship to principles of privacy notice and choice. We have further introduced the scope, specific research questions, and overarching goal of this thesis. In Chapters 2 through 9 we present each of the eight studies that we conducted. In Chapter 2, we present a large-scale evaluation of websites' usage of Platform for Privacy Preferences (P3P) compact policies (CPs). In Chapter 3, we present a large-scale evaluation of financial institutions standardized privacy disclosures. In Chapter 4, we present a manual evaluation of 75 online advertising companies' privacy policies. In Chapter 5, we discuss users' knowledge and perceptions of online behav-

Application Domain	Research Methods Used	Research Question
Websites' privacy policies	Large-scale, automatic evaluation of machine-readable privacy policies	Q1 and Q2
Financial institutions' privacy notices	Large-scale, automatic evaluation of standardized privacy notices	Q1 and Q2
Online behavioral advertising (OBA)	Usability testing, semi-structured interviews, online surveys, and manual evaluation of privacy policies	Q1 and Q2
Online social networks (OSN)	Online field study	Q3

Table 1.2: Mapping between covered application domains, research methods and research questions.

ioral advertising (OBA) collected through semi-structured interviews. In Chapter 6, we present a 45-participant laboratory study evaluating the usability of nine popular tools to limit OBA. In Chapter 7, we present a 1,505-participant online study evaluating users' understandings and reactions to OBA privacy disclosures. In Chapter 8, we present two large-scale online studies with 2,912 and 1,882 participants, respectively, investigating the effect of advertising companies' practices on users' willingness to share information with those companies. In Chapter 9, we designed and tested nudging notices through a 28-participants field study with active Facebook users. In each chapter, we include a detailed discussion of background, related work, and research methods applicable to the study presented in that chapter. In Chapter 10, we discuss the public policy implications of this research. Finally, in Chapter 11, we present conclusions and opportunities for future research.

Chapter 2

The Misrepresentation of Website Privacy Policies Through the Misuse of P3P Compact Policy Tokens

Platform for Privacy Preferences (P3P) compact policies (CPs) are a collection of three-character and four-character tokens that summarize a website's privacy policy pertaining to cookies. User agents, including Microsoft's Internet Explorer (IE) web browser, use CPs to evaluate websites' data collection practices and allow, reject, or modify cookies based on sites' privacy practices. CPs can provide a technical means to enforce users' privacy preferences if CPs accurately reflect websites' practices. Through automated analysis we can identify CPs that are erroneous due to syntax errors or semantic conflicts. We collected CPs from 33,139 websites and detected errors in 11,176 of them, including 134 TRUSTe-certified websites and 21 of the top 100 most-visited sites. Our work identifies potentially misleading practices by web administrators, as well as common accidental mistakes. We found thousands of sites using identical invalid CPs that had been recommended as workarounds for IE cookie blocking. Other sites had CPs with typos in their tokens, or other errors. 98% of invalid CPs resulted in cookies remaining unblocked by IE under it's default cookie settings. It appears that large numbers of websites that use CPs are misrepresenting their privacy practices, thus misleading users and rendering privacy protection tools ineffective. Unless regulators use their authority to take action against companies that provide erroneous machine-readable policies, users will be unable to rely on these policies.

2.1 Introduction

The Platform for Privacy Preferences (P3P) is a World Wide Web Consortium (W3C) recommendation for specifying website privacy policies in a machine readable format. Developed as part of an industry self-regulatory effort and published in 2002, it provides two privacy policy formats: full policies and compact policies (CPs). P3P full policies are XML files that represent website privacy policies in detail. P3P CPs summarize website privacy policies regarding cook-

This chapter is based on "Token Attempt: The Misrepresentation of Website Privacy Policies Through the Misuse of P3P Compact Policies Tokens" [26].

ies using a string of three-character and four-character tokens [27]. Internet Explorer (IE) makes cookie-filtering decisions by comparing a website's CPs with user-configured privacy preferences [28]. If a CP does not reflect the website's actual privacy practices then that CP is not useful for decision making. In 2002, regulators from several countries agreed that a P3P policy is legally binding and "constitutes a representation to consumers on which they can be expected to rely" [29].

In this paper, we present the results of our automated evaluation of P3P CPs collected from 33,139 websites. Verifying the accuracy of CPs requires comparing the computer-readable statements in a CP with a website's actual privacy practices with respect to cookies. Sometimes this can be done by reading the statements that sites make in their privacy policies. However, sometimes privacy policies do not have sufficient details, and they may not be completely accurate. Even when accurate details are available, reviewing the accuracy of CPs is a very labor-intensive process. Therefore we looked for ways to identify errors in P3P CPs that can be determined based on the syntax errors and conflicting tokens within CPs themselves, without having to review thousands of privacy policies, and without requiring first-hand knowledge that these policies are accurate. We developed heuristics to detect three categories of CP errors: $invalid\ tokens$, $missing\ tokens$, and $conflicting\ tokens$.

We found that nearly 34% of the CPs evaluated in August 2010 have at least one error in these categories, and more than half of those with errors omit required information. In addition to syntax and semantic errors, we found that 79% of CPs lack a corresponding full P3P policy, which is required for P3P-compliance. Among the 100 most-visited websites[1] we found 48 sites with CPs, 21 of which exhibited CP errors that our automated analysis could detect. In addition, 41 of these popular sites with CPs did not have corresponding full P3P policies. The numbers reported in this paper should be considered a lower bound for the actual number of CPs with errors, as CPs may contain other types of errors that require manual comparison with human-readable privacy policies or other types of analysis to detect.

We analyzed the impact of CP errors on privacy and found that these errors could mislead users by misrepresenting privacy practices with respect to cookies. We also determined the implications of faulty CPs for user agent behavior. We analyzed practices that appear to be deliberately designed to bypass IE default privacy filters and found that more than 97% of incorrect CPs would bypass these default filters. Our work identifies potentially misleading practices by web administrators, as well as common accidental mistakes.

This paper is organized as follows. In Section 2.2, we discuss the details of P3P and compact policies and review related work. In Section 2.3, we describe our study methodology. In Section 2.4, we introduce and define common P3P CP errors, present data on the frequency with which each type of error appears, and discuss the privacy implications of these errors. In Section 2.5, we analyze the impact of incorrect CPs on IE and discuss evidence that incorrect CPs are being used to bypass IE privacy filters. In Section 2.6, we analyze the CPs of three groups of websites: popular sites, network advertisers that offer opt-out cookies, and TRUSTe-certified sites. Finally, we present our conclusions in Section 2.7. We also include seven appendixes that provide detailed information about CP tokens and data from our analysis.

[1] Most-visited websites as of August 2010, according to http://www.quantcast.com/top-sites-1.

2.2 Background and Related Work

In this section we provide an overview of the Platform for Privacy Preferences (P3P) 1.0. In addition, we provide background on P3P compact policies and discuss related work.

2.2.1 The P3P Specification

P3P is a standard for specifying websites' privacy policies in a machine-readable XML format that can be processed and acted upon by automated tools [30], [27]. P3P allows user agents to automatically evaluate privacy policies against users' preferences. For example, users can set up their privacy preferences using user agents embedded in their web browsers so that their browsers will warn about mismatches with privacy preferences or block cookies at sites where mismatches occur [31]. When P3P is embedded in search engines, search results can be annotated automatically with privacy information [32]. P3P enables automatic generation of standardized "nutrition label" privacy notices, which are more understandable and easier to read than conventional policies [33]. It also allows automated tools to collect privacy policy data for analysis [34].

Published in 2002, the P3P 1.0 specification defines two types of P3P policies: full policies and compact policies. P3P full policies are written in XML format and use a defined vocabulary and a data schema to represent human-readable privacy policies in a machine-readable language. Compact policies summarize privacy practices associated with the use of cookies; they are transmitted in response to HTTP requests using HTTP headers. Full P3P policies are the authoritative source for website information management practices. The P3P specification requires compact policies to have corresponding full P3P policy files [27].

The P3P specification defines a protocol for requesting and transmitting P3P policies via HTTP. Retrieving a P3P policy requires a two-step process. P3P user agents issue requests for P3P policy reference files stored in a "well-known" location and also check for P3P HTTP headers and P3P metadata embedded in HTML content. P3P policy reference files contain references to the location of full P3P policy files. After fetching a P3P policy reference file, user agents are able to locate and retrieve a full P3P policy file [27].

W3C maintains an online validation tool that checks for syntax errors in full P3P policies and CPs.[2]

2.2.2 P3P Compact Policies

Compact policies (CPs) are defined in the P3P specification as an optional performance optimization. CPs are optionally served when a website transmits a cookie. They provide a lightweight mechanism to provide information about a website's privacy practices with respect to cookies and help user agents quickly decide how to process cookies. The P3P specification requires that sites that choose to deploy CPs make an effort to do so accurately. According to the P3P specification, "if a web site makes compact policy statements it MUST make these statements in good faith" [35]. Despite being an optional component of P3P, CPs are important because they are used by IE 6, 7, and 8 to determine whether to block or modify cookies.

[2] http://www.w3.org/P3P/validator.html

P3P specifies a set of CP *tokens* associated with nine types of P3P policy elements. Valid CPs must have at least five of these elements. The valid *tokens* for each element, the corresponding full-P3P policy elements, and a short description of each are provided in Appendix A.1.

2.2.3 Internet Explorer Cookie Blocking

IE 6, 7, and 8 allow users to set their privacy preferences, which are then used to evaluate websites' CPs and perform cookie filtering. IE considers cookies that are used to collect personally identifiable information (PII) without providing users the choice to opt-in/opt-out to be *unsatisfactory cookies*. IE performs cookie filtering based on six privacy levels that the user can set up; namely, *Block All Cookies*, *High*, *Medium High*, *Medium* (default level), *Low*, and *Accept All Cookies* [28]. The vast majority of users do not change the default privacy settings, so analyzing filtering conditions at the *medium* level allows us to determine the privacy impact of invalid CPs that bypass IE filters. Under the medium setting, unsatisfactory first-party cookies are converted into session cookies and unsatisfactory third-party cookies are rejected. In addition, third-party cookies not accompanied by CPs are rejected. When cookies are rejected, website functionality may be impaired and sites may be unable to collect some of the data they use for operational or business purposes. Because of this cookie-blocking feature, many website administrators have implemented CPs on their websites.

2.2.4 Related Work

The first large-scale automated analysis of P3P policies found that approximately 10% of 5,739 websites tested in 2003 were P3P-enabled [34]. In addition, 85 websites had only compact policies and were therefore not P3P-compliant, and about a third of the full P3P policies had technical errors. Reay et. al. performed a study of over 3,000 full and compact P3P policies. They found high rates of websites posting CPs without corresponding full P3P policies, as well as many sites that had conflicts between full and compact P3P policies [36].

Cranor, et al. performed an error analysis on P3P full policies. They found that 73% of the 14,720 full P3P policies analyzed contained syntax errors. They compared full P3P policies with their corresponding natural language privacy policies at 21 popular websites and found conflicts between the P3P and natural language policy at each of these sites. Most conflicts were associated with the *PURPOSE*, *CATEGORIES*, and *RECIPIENTS* elements [32].

Reidenberg and Cranor studied the accuracy of P3P user agents [37]. They suggested that inaccurate representations by user agents could undermine the purpose of the P3P standard. From a legal standpoint, the enforceability of an agreement based on a P3P user agent's simplified representation of a website privacy policy is uncertain [38]. In the United States, the Federal Trade Commission Act of 2006 empowers the FTC to prevent "unfair or deceptive acts or practices in or affecting commerce...." The FTC has used this authority to take action against companies with deceptive privacy policies. While the FTC has not taken such actions on the basis of deceptive machine-readable privacy policies to date, it appears to be within the FTC's authority to do so [29].

2.3 Methodology

We collected CPs from four data sets. First, we collected 52,156 URLs containing CPs from the Privacy Finder search engine cache in January 2010. The cache was initially seeded in 2005 through a web crawl [32] and has grown over time as a result of Privacy Finder users' searches. Second, on July 3, 2010 we collected a list of the 3,417 websites holding TRUSTe privacy seals, as reported by the membership list posted on the TRUSTe website [39]. Third, we collected a list of the 100 most-visited websites from *www.quantcast.com* on August 10, 2010. Fourth, we collected a list of 75 network advertisers offering opt-out cookies that can be set using the Beef Taco (Targeted Advertising Cookie Opt-Out) Firefox extension as of June 23, 2010. We removed duplicate domains from multiple datasets, so we had a final list of 55,636 unique URLs.

In August 2010 we used our web crawler to visit our list of 55,636 URLs and attempt to download CPs. Some datasets had URLs that were not fully qualified (for example, foo.com instead of `http://www.foo.com`) so we prepended http:// and http://www. and tried again if sites failed. When we contacted sites that no longer responded, we retried twice. When we contacted sites that gave other errors, we retried up to ten times. Some of the sites on this list were no longer available, and some that previously had provided CPs were no longer providing them. We were able to collect CPs from 33,139 sites.

At sites with P3P CPs we also checked for full P3P policies. The crawler checked for a P3P policy reference file in the P3P well-known location, HTTP header, and page content. If found we parsed this file to retrieve the location of the full P3P policy (the policy is either embedded in the policy reference file or referenced by a URL in the policy reference file). We then retrieved the full P3P policy and file and verified that it contained a P3P policy. 7,016 of the sites with CPs also had full P3P policies.

Some domains we analyzed had more than one host on the domain with a compact policy in our Privacy Finder cache dataset, for example both `http://www.x.example.com` and `http://www.y.example.com`. We report data in both aggregated form where each domain is represented only once for each unique CP found regardless of the number of hosts (just one entry for example.com if all hosts in example.com have the same CP) and in non-aggregated form where each host is represented (x.example.com and y.example.com are separate). Furthermore, if a single domain exhibited different incorrect CPs, we reported them separately and count both of them in the aggregated data set. There are $19,820$ domains in our aggregated data set.

We define a set of P3P CP errors and group them into three categories: *invalid tokens*, *missing tokens*, and *conflicting tokens*. We define each type of error in Section 2.4.1. We developed scripts to parse CPs and detect each type of error we defined. We also developed a script to check each CP to determine whether it would be considered satisfactory by IE under the default privacy setting [28].

2.4 Compact Policy Errors

In this section we present the results of our evaluation of 33,139 collected CPs. We define three categories of errors, and for each type provide data on the frequency of occurrence. We suggest probable underlying causes of these errors: typos, lack of understanding of the P3P specification, or attempts to avoid web browser cookie filtering. We further evaluate the impact of these errors on users' privacy.

We found 11,176 CPs with errors, about 34% of the URLs we analyzed. If we aggregate these invalid CPs by unique domain names, the number of incorrect CPs is reduced to 4,696. Domain names with more than one website account for at least 57% of the total invalid CPs. If companies hosting multiple sites correct their CPs, the CPs with errors could be reduced to less than 15% of the URLs we analyzed. Table 2.1 summarizes the CP errors we found.

According to the P3P specification, CPs must be accompanied by a full P3P policy. Therefore, we investigated how many of the collected CPs have a corresponding full P3P policy. We found that only 21% of those websites providing CPs also provided full P3P policies.

Type of error	**Number of URLs**	**Percent of URLs with Errors**	**Number of Domains**	**Percent of Domains with Errors**
Any problem	11,176	33.7%	4,696	23.7%
Invalid tokens				
Invalid tokens	3,839	11.6%	794	4.0%
CURa (no other problems)	5,295	16.0%	2,557	12.907%
Missing tokens	6,402	19.3%	3,319	16.8%
Conflict between tokens				
IVA	3,977	12.0%	923	4.7%
CON	3,899	11.8%	835	4.2%
IVD	1,059	3.2%	669	3.4%
TEL	340	1.0%	178	0.9%
NID	366	1.1%	168	0.9%
NOR	345	1.0%	99	0.5%
NON	32	0.1%	27	0.1%
Total analyzed	33,139 URLs		19,820 Domains	

Table 2.1: Summary of CP errors. Some CPs have errors in more than one category. CPs that contain the invalid $CURa$ token and no other errors are not included in the "any problem" count.

2.4.1 Invalid Tokens

Definition

All valid CP tokens are shown in Appendix A.1. Some tokens can optionally be accompanied by an attribute that specifies that a data practice is performed always, on an opt-in basis, or on an opt-out basis (a, i, or o). No other tokens or token-attribute combinations are valid. An *invalid token* error occurs when a CP includes an invalid token or an invalid token-attribute combination.

Tokens that do not specify the optional attribute default to the $always$ attribute (a). The CUR (current purpose) token does not take an optional attribute, but we found $CURa$ is commonly used in CPs. As CUR is the only $PURPOSE$ element that does not allow an attribute, we believe web administrators may be mistakenly adding the invalid a attribute. Since CUR already means that data is always collected for the current purpose this does not change the semantics of the CP, though it is technically invalid. In this paper we report the frequency of $CURa$ separately from other invalid tokens and do not count this error in our overall error numbers.

Evaluation

Of the total CPs evaluated, 11.6% contained invalid tokens. An additional 16% contained the invalid and harmless CURa token and no other invalid tokens, but we do not include these URLs in our count of sites with invalid tokens. Companies publishing multiple websites under a single domain name account for 79% of the invalid token errors.

Analysis

Many invalid token errors are likely to be accidental. We found that several CPs include the CUS string which is not a valid token. We believe that web administrators might have intended to use the TAI (tailoring) token but wrote CUS (customize) instead, which is an option that appeared in early drafts of the P3P specification. Another common syntax error is swapping letters within tokens. For example, web administrators tend to write OPT instead of OTP, TIA instead of TAI and so on. Web administrators commonly add attributes to tokens that do not accept them. For example, they use $CAOo$ instead of CAO (providing user access to contact information). Another common token found with invalid attributes was OUR, which is the only recipient token that does not accept an attribute. Some invalid CPs contain multiple valid tokens concatenated together without the required white-space separator. More concerning are the CPs that contain completely invalid strings such as $AMZN$ or $VPRT$.

While some of the invalid token errors are likely typos and many appear to be harmless, these errors may cause user agents to incorrectly interpret a CP, which could confuse and mislead users. In addition, it appears that some of the completely-invalid tokens are being used by sites to prevent their cookies from being blocked without providing a meaningful CP.

2.4.2 Missing Tokens

Definition

According to the P3P specification, each statement in a CP that does not include the NID (no user-identifiable data collected) token should include at least one *PURPOSE*, *RECIPIENT*, *RETENTION*, *CATEGORIES* and *ACCESS* token. A *missing tokens* error occurs when a CP without a NID token does not contain at least one token from each of the above five categories.

Evaluation

Missing-token errors are the most common type of error we found. These errors occur in 19% of the total analyzed CPs and account for more than 57% of the incorrect CPs.

Analysis

Missing tokens make it impossible to use the CP to determine a website's data practices with respect to cookies. For example, a CP that contains no *CATEGORIES* token fails to communicate the type of information associated with a cookie. Because P3P declarations are positive, the absence of a token is an indicator that a website does not engage in a particular practice. Therefore, sites that are missing *CATEGORIES* tokens are effectively stating that they do not collect any category of data. Furthermore, IE analyzes *CATEGORIES*, *PURPOSE* and *RECIPIENT* tokens to make cookie-blocking decisions. As we discuss later, if these tokens are omitted, IE may incorrectly accept a cookie that would otherwise be blocked.

2.4.3 Conflicting Tokens

Definition

The CP semantics should guarantee consistency between different tokens. For example if a CP includes the *ALL* token, indicating users have access to all of their personally identifiable information, the CP cannot include the *NON* token as well, as that means users have no access to their personally identifiable information. Similarly, if a CP includes the *NOR* token, which means the website does not store permanently the information collected, it cannot include any other *RETENTION* element tokens. If a website declares that it does not collect personally identifiable information (PII) then the CP should not include tokens that suggest collection of such information. Several conflicting token scenarios are defined in the following paragraphs.

Conflict with the NID token. The non-identifiable token *NID* should be used only when the website does not collect any type of PII and associate it with a cookie. There are many tokens that suggest collection of this information. In particular *PHY* (information to locate or contact an individual in the physical world), *ONL* (information to locate or contact an individual on the Internet, e.g. email address), *FIN* (financial information), *LOC* (physical location data) and *GOV* (government identifier, e.g. social security number) tokens are directly associated with the collection of PII. Similarly, *CAO* (contact and other information) and *IDC* (online and physical contact information) *ACCESS* tokens should appear only if PII is collected by the website and associated with a cookie. Furthermore, the *IVA* (individual analysis), *IVD* (individual decision), *CON* (contact) and *TEL* (telemarketing) *PURPOSE* tokens require PII and should not be in the CP if the *NID* token is also in the CP.

Conflict with the IVA/IVD tokens. Performing individual analysis (*IVA*) or making individual decisions (*IVD*) requires identifying a particular individual. The *IVA* and *IVD* tokens must be accompanied by at least one of the following *CATEGORIES* tokens: *PHY, ONL, FIN, PUR, GOV.*

Conflict with the CON token. The contact token, *CON*, requires enough information to contact the individual either by electronic or physical means. *CON* must be accompanied by at least a *PHY* or *ONL* token.

Conflict with the TEL token. Telemarketing requires phone numbers, which are part of the physical category. Therefore, the *TEL* token must be accompanied by a *PHY* token.

Conflict with the NON/NOR tokens. The *NON* token indicates users are not allowed to access any data collected about them. None of the remaining tokens in the ACCESS element should appear in the CP with the *NON* token. Similarly, the *NOR RETENTION* token means the website does not store permanently any kind of PII. If there is a *NOR* token the CP should not contain any other *RETENTION* tokens.

Evaluation

The most common errors in this category are associated with the individual analysis (IVA) token. We found these errors in 12% of the analyzed CPs and 35% of the invalid CPs. The second most common type of error in this category is associated with the *CON* token, which represents more than 11% of the collected CPs and more than 34% of the total incorrect CPs. Conflicts involving the remaining tokens are relatively rare, each occurring in less than 4% of the CPs we analyzed.

Analysis

When a website incorrectly uses the NID token, users have conflicting information about whether or not their PII is being collected. Web administrators may misunderstand the definition of the NID token in the P3P specification and use it incorrectly. The most common NID conflict we found was with the ONL token (email address or other online contact information). This suggests that web administrators might be unaware that email addresses are considered identifiable information. Another common conflict is with the IVA token (individual analysis) which implies that identifiable information is used to perform an analysis. If the NID token is used and it is true that no PII is collected, the PSA token (pseudo-analysis) should be used instead of the IVA token.

The *NOR conflicting token* error leads to confusion about website retention practices. This is probably an accidental mistake, but could negatively affect users' decisions. We found that some CPs include NOR and IND tokens: while the company claims no retention of PII at the same time they claim that they store information indefinitely. Many of the invalid CPs in this error category include the BUS token, meaning the retention period is based on their business practices. While it might be the case that their business practices do not require the retention of personal information, including both of these tokens is confusing.

The *IVA, IVD, CON and TEL conflicting token* error categories can also lead to user confusion. These errors suggest either a lack of understanding of the P3P tokens or an attempt to misrepresent a site's practices.

Most of the websites in the *NON conflicting token* error category simultaneously include the CAO and NON tokens in their CPs. This suggests a possible misunderstanding of the NON token. It leaves users with uncertainty about the access options offered by the website.

2.4.4 CPs Without Full P3P Policies

The P3P specification requires websites implementing CPs to have a corresponding full P3P policy. Only 7,016 of the 33,139 URLs in our data set (21%) had full P3P policies. As shown in Table 2.2, CPs without errors were almost three times more likely to have full P3P policies than those with errors. 26.9% of error-free CPs had corresponding full P3P policies, while only 9.9% of CPs with errors had full P3P policies. For CPs with the two most common errors, invalid tokens and missing tokens, the percentage of URLs that present full P3P policies is even lower. This finding provides some evidence that websites with these types of errors may be providing inaccurate CPs to avoid having their cookies blocked.

We found full P3P policies for 17% of the 391 TRUSTe sites with CPs, 55% of the 11 network advertising sites with CPs, and 15% of the 48 most-visited sites with CPs. Appendices A.5, A.6, and A.7 show the details on most-visited sites, network advertiser sites, and TRUSTe sites respectively.

The CPs with IVD, TEL, and NID conflicting tokens were more likely to have corresponding full P3P policies. This result suggests that many of these websites may be making

Type of error	Number of CPs	Full P3P Policies	Ratio CPs/Full
None (valid CPs)	21,963	5,915	26.9%
Any problem	11,176	1,106	9.9%
Invalid tokens	3,839	255	6.7%
Missing tokens	6,402	469	7.03%
Conflicting tokens			
IVA	3,977	586	14.7%
CON	3,899	321	8.2%
IVD	1,059	424	40.0%
TEL	340	107	31.5%
NID	366	127	34.7%
NOR	345	64	18.6%
NON	32	1	3.1%

Table 2.2: Full P3P policies at websites with CPs.

good faith efforts to properly comply with the P3P specification and not just sending CPs to avoid cookie blocking. However, web administrators may not fully understand the meanings of these tokens.

2.4.5 Other Semantic Errors

In addition to the errors already discussed, CPs may contain other types of errors that we were unable to detect automatically. For example, CPs may be semantically inconsistent with human-readable policies posted on websites or with full P3P policies. In addition, CPs may fail to accurately represent website data practices. Evaluating semantic errors requires analyzing websites' human readable policies, which is not a task that can be automated. Given the large scale of our study, reading human readable policies for the thousands of sites we evaluated is impractical. Furthermore, human-readable policies do not always contain detailed information about data collection and treatment associated with the use of cookies, so even reading these policies would not necessarily resolve questions of P3P accuracy. Thus, a complete evaluation of semantic errors is beyond the scope of our study. However, to gain some insights into the prevalence of semantic errors, we manually compared CPs with full P3P policies and human-readable policies at 41 sites, including 11 most-visited web sites. As detailed in Appendix A.3 and A.4, 40 of these sites failed to provide full P3P policies, 15 failed to provide human-readable privacy policies, and 2 sites had CPs consisting only of meaningless, invalid tokens. When comparing the full policies and human-readable policies with their corresponding CPs, we found 4 sites with slight discrepancies, and 15 sites with major discrepancies. Furthermore, we were not able to compare 5 sites because their human-readable policies did not include any statement about cookies.

2.5 Compact Policies and Internet Explorer

Microsoft helped drive P3P adoption by using P3P CPs to make cookie-blocking decisions. However, the large number of CP errors and low rate of full P3P compliance suggest that many websites are adopting P3P CPs in order to avoid cookie blocking, but are not presenting accurate representations of their privacy practices. To gain additional insights into CP adoption, we analyzed the CPs we collected to determine whether IE would classify them as *satisfactory* under the default medium setting. We found that 99% of CPs collected would be considered satisfactory by IE. Of the 33,139 CPs we examined, only 118 error-free CPs and 263 CPs with errors were considered unsatisfactory, and thus would likely be blocked when the cookie was used in a third-party context. All but three of the unsatisfactory CPs with errors had missing-tokens errors.

IE cookie filters only look for combinations of tokens considered unsatisfactory. They ignore invalid tokens and do not check to make sure the minimum required tokens are present. They also do not look for token conflicts. If IE performed the same sort of checks we did in our analysis and treated CPs with these errors as unsatisfactory, we would expect the error rate to be reduce over time because companies would have an incentive to correct the errors in their CPs (although it is possible that we would then see an increased rate of other types of errors that are less-easily detectable through automated analysis).

After observing a large number of identical CPs in our data set, we suspected that web administrators might be copying these CPs from a common source. We used a search engine to track down the source of some of the most common CPs in our data set.

We discovered that Microsoft's support website recommends the use of invalid CPs as a work-around for a problem in IE. Specifically, a FRAMESET or parent window that references another site inside a FRAME considers the referenced site as a third-party, even if it is first-party content located on the same server [40]. Microsoft suggests the following invalid CP: $CAO\ PSA\ OUR$. This CP is clearly invalid since it does not contain any $RETENTION$ or $CATEGORIES$ tokens. Even if the CP were valid, Microsoft's recommendation undermines the purpose of P3P since it encourages web administrators to use CPs that do not represent their actual data practices. We found several technical blogs recommending similar solutions [41], [42]. Some of them suggested the exact CP described above and referred to the Microsoft support website as the source of their advice [43]. This CP occurred 2,756 times in our data set. Only 31 of these CPs had corresponding full P3P policies. Nearly 25% of the invalid CPs used these tokens, representing 43% of invalid CPs in the missing-tokens error category. We did not find this CP at any of the 100 most-visited websites.

In an article titled "P3P in IE6: Frustrating Failure" posted in the O'Reilly blog, the author suggests another "trick" to bypass IE6 privacy filters. He recommends adding a P3P CP header that "enables your cookie to survive any privacy setting" [44]. The CP suggested is: $NOI\ ADM\ DEV\ PSAi\ COM\ NAV\ OUR\ OTRo\ STP\ IND\ DEM$. This CP does not contain any of the errors we tested for. However, if web administrators blindly post it without confirming that it matches their site's actual practices, they will mislead users. This CP occurred 4,360 times in our data, representing 13% of the CPs analyzed. Only 12 of these CPs had corresponding full P3P policies. We did not find this CP at any of the 100 most-visited websites.

We manually analyzed 30 privacy policies of websites that use the CP recommended by the O'Reilly blog. 14 of these 30 websites were randomly taken from the domains that present the

Compact Policy listed in http-stats.com/P3P	Number of occurrences in collected CPs	Errors	IE Satisfactory cookie?
CAO DSP COR CUR ADM DEV TAI PSA PSD IVAi IVDi CONi TELo OTPi OUR DELi SAMi OTRi UNRi PUBi IND PHY ONL UNI PUR FIN COM NAV INT DEM CNT STA POL HEA PRE GOV	2,738	None	YES
NOI ADM DEV PSAi COM NAV OUR OTRo STP IND DEM	4,360	None	YES
NON DSP CURa OUR NOR UNI	0	Missing tokens	YES
ALL CURa ADMa DEVa TAIa OUR BUS IND PHY ONL UNI PUR FIN COM NAV INT DEM CNT STA POL HEA PRE LOC OTC	102	None	YES
BUS CUR CONo FIN IVDo ONL OUR PHY SAMo TELo	293	Missing tokens	YES
CAO DSP COR CURa ADMa DEVa OUR IND PHY ONL UNI COM NAV INT DEM PRE	553	None	YES
NOI NID ADMa OUR IND UNI COM NAV	464	None	YES
ALL DSP COR CURa ADMa DEVa TAIa PSAa PSDa IVAa IVDa CONa TELa OUR STP UNI NAV STA PRE	24	Conflicting IVA, IVD, CON, and TEL tokens	YES
NOI DEVa TAIa OUR BUS UNI STA	359	None	YES
CAO PSA OUR	2,756	Missing tokens	YES

Table 2.3: Frequency of "common" CPs from http-stats.com in collected CPs

most URLs with this CP. The remaining 16 evaluated websites were randomly chosen from the set of websites sending this CP. From the 30 manually-analyzed websites, only one had a valid P3P full policy. However, the valid full P3P policy did not match the corresponding CP, and when we returned to the website in September we discovered the full P3P policy had been removed. We were unable to locate any human-readable privacy policy at 15 of these websites. Furthermore, none of the human-readable privacy policies we found properly matched the corresponding CP. Apendix A.3 summarizes the results of this manual evaluation.

From the 15 websites that exhibited a human-readable policy, only 10 mention the use of cookies. However, we found semantic inconsistencies between all of these policies and their corresponding CPs. Most of these policies made no mention of data practices related to the NOI access token; ADM, DEV, or $PSAi$ purpose tokens; STP or IND retention tokens; or DEM categories token. Many of them made references to other cookie-related data practices not captured by the CP.

We performed a Google search for "ie blocking iframe cookies" and found a number of sites suggesting similar solutions. For example, another blog post recommended a CP that we found 300 times in our data set [42]. On the other hand, we also found a popular question-and-answer site that advised web administrators to create CPs that accurately reflect their site's privacy policy: "The tags are not only a bunch of bits, they have real world meanings, and their use gives you real world responsibilities! For example, pretending that you never collect user data might make the browser happy, but if you actually collect user data, the P3P is conflicting with reality. Plain and simple, you are purposefully lying to your users...." Immediately under the example CP was the warning: "Note that the combination of P3P headers in the example may not be applicable on your specific website; your P3P headers MUST truthfully represent your own privacy policy!" [45]. This warning must have been effective as we did not find the example CP in our data set.

We examined the top 10 P3P header values listed at http-stats.com, a website that crawls the web and compiles data on HTTP header values. These headers included the CPs recommended in the O'Reilly blog and on the Microsoft support website. As shown in Table 2.3, we found multiple instances of nine of these CPs in our data set and detected errors in four of these CPs. All of these CPs are considered satisfactory by IE.

2.6 Compact Policies and Popular Websites

In this section we present the results of our evaluation of the top-100 most-visited websites, 3,417 TRUSTe-certified websites, and 75 network advertising websites offering opt-out cookies that can be set using the Beef Taco Firefox extension. We found that only 391 of the evaluated TRUSTe sites had CPs and 134 of those had errors. 48 of the 100 most-visited sites had CPs and 21 had errors. 11 of the evaluated network advertising sites had CPs and only one had an error.

The top-visited domains with CPs are listed in Appendix A.5. We analyzed the errors in detail for the top 50 most-visited sites with CPs that contained errors. Because of the popularity of these sites, errors in their CPs have an impact on a large number of users. Table 2.4 shows the CPs and types of errors found. Only one of these websites, *microsoft.com,* displayed a full P3P policy.

The $facebook.com$ CP is invalid because it is missing *PURPOSE*, *RECIPIENT*, *RETENTION*, *CATEGORIES*

and *ACCESS* tokens. The included tokens, DSP and LAW, do not provide any information about the site's collection or use of data. This CP simply states that any privacy dispute will be resolved according to a law referenced in their privacy policy, and implies that the site collects no data associated with the cookie. When doing preliminary work for this study in 2009, the $facebook.com$ compact policy contained only the single invalid token $HONK$. Both of these CPs are useless for communicating with user agents and users. It is likely that $facebook.com$ is using their CP to avoid being blocked by IE.

Websites under the $msn.com$ domain exhibited a CP that includes the invalid $CUSo$ token. Two other Microsoft-owned sites, $microsoft.com$ and $windows.com$ use the same CP. These websites display the TRUSTe EU Safe Harbor Privacy seal. We believe that these websites are likely attempting to comply with P3P; however, they are not using P3P properly.

The $live.com$ CP does not include any $ACCESS$ tokens. This CP suggests collection of PII, but does not provide any information about whether users can access their personal information.

The $amazon.com$ and $imdb.com$ domains each contain a single invalid token and no other tokens, so they fall into the invalid-tokens and missing-tokens categories. It appears that these two websites use a CP only for the purpose of avoiding IE cookie filtering.

URL	Compact Policy	Errors found	Valid Full P3P Policy	IE Satisfactory cookie?	TRUSTe seal
facebook.com	DSP LAW	Missing tokens	NO	Yes	EU Safe Harbor
msn.com	BUS CUR CONo FIN IVDo ONL OUR PHY SAMo TELo	Missing tokens	NO	Yes	EU Safe Harbor
live.com	BUS CUR CONo FIN IVDo ONL OUR PHY SAMo TELo	Missing tokens	NO	Yes	EU Safe Harbor
amazon.com	*AMZN*	Invalid tokens, Missing tokens	NO	Yes	None
microsoft.com	ALL IND DSP COR ADM CONo CUR *CUSo* IVAo IVDo PSA PSD TAI TELo OUR SAMo CNT COM INT NAV ONL PHY PRE PUR UNI	Invalid tokens	YES	Yes	EU Safe Harbor
reference.aol.com	UNI INT STA NAV DEV CUR OUR	Missing tokens	NO	Yes	Web Privacy
atlas.mapquest.com	STA INT UNI CUR DEV NOI OUR	Missing tokens	NO	Yes	None
godaddy.com	IDC DSP COR LAW CUR ADM DEV TAI PSA PSD IVA IVD HIS OUR SAM PUB LEG UNI COM NAV STA	Conflict between IVA and IVD tokens	NO	Yes	EU Safe Harbor
imdb.com	*IMDB*	Invalid tokens, Missing tokens	NO	Yes	None
windows.com	ALL IND DSP COR ADM CONo CUR *CUSo* IVAo IVDo PSA PSD TAI TELo OUR SAMo CNT COM INT NAV ONL PHY PRE PUR UNI	Invalid tokens	NO	Yes	None
hulu.com	NOI DSP COR NID ADMa *OPTa* OUR NOR	Invalid tokens	NO	Yes	None

Table 2.4: CPs of 50 most-visited websites with errors, sorted by popularity. Invalid tokens are shown in italics.

The *aol.com* domain's CP is invalid since it is missing *ACCESS* and *RETENTION* tokens. The *mapquest.com* domain is missing a *RETENTION* token. The *godaddy.com* domain has conflicting *IVA* and *IVD* tokens. This CP is confusing since it claims the site is using identifiable information to perform individual analysis and decisions; however, it does not include any *CATEGORIES* token associated with the collection of identifiable information.

Finally, the *hulu.com* domain contains the invalid *OPTa* token which presumably is intended to be *OTPa* instead. This makes it syntactically incorrect. More importantly, the CP includes the *NID* token, claiming that no PII is associated with cookies. We read the human-readable privacy policy of this website and found that it explicitly mentions linking PII to cookies.

Host domain	Invalid CPs	Percent of total invalid CPs
tripod.com	2,575	23.0%
addresses.com	1,054	9.4%
msn.com	358	3.2%
cjb.net	247	2.2%
livedoor.biz	116	1.0%
ning.com	112	1.0%
Total invalid:	**4,462**	**39.92%**

Table 2.5: Domains accountable for most of the CP problems.

All but one of these top websites do not have a full P3P policy, and several of them have CPs that appear to be well-crafted to bypass IE filtering. Further analysis will be needed to determine if they actually follow the data practices they claim through their CPs; however, as detailed in Appendix A.4, there are inconsistencies that suggest they do not follow the practices they claim.

Domains such as *facebook.com*, *msn.com*, *live.com* and *aol.com* exhibited TRUSTe privacy seals, despite displaying invalid CPs. Indeed, we found that 391 of the 3,417 TRUSTe-certified websites have CPs, but 134 (34.3%) of these had at least one problem with their CPs, as detailed in Appendix A.7. 28 out of the 48 top websites with CPs appeared on the list of TRUSTe websites and 11 (39.3%) of these had invalid CPs. This suggests TRUSTe is not reviewing websites' CPs when issuing privacy seals.

Network advertisers tend to make heavy use of third-party cookies in order to provide targeted advertising. Therefore, the use of CPs among network advertisers is of particular importance. Without CPs, many network advertising cookies would be blocked by IE because they are used in a third-party context. In addition, users are generally not aware of what third-parties are setting cookies on the sites they visit or what their privacy practices are. If used properly, P3P could provide information about privacy practices that would otherwise be difficult for users to obtain. We collected a list of 75 network advertisers offering opt-out cookies that can be set using the Beef TACO (Targeted Advertising Cookie Opt-Out) Firefox extension. As detailed in Appendix A.6, We found that only 11 of them delivered CPs with their opt-out cookie. However, we found errors in only one of these CPs.

Some of the most-visited domains host many websites in their domain, and thus have many invalid CPs. Table 2.5 lists the 6 domains responsible for at least 100 invalid CPs each. These 6 domains are accountable for nearly 40% of CPs with errors. These include two of the top 100 most-visited web domains: *tripod.com* and *msn.com*.

2.7 Conclusions

In this paper we present data on errors commonly found in P3P compact policies that are detectable through automated analysis. We evaluated CPs collected from 33,139 websites on 19,820 domains and found *invalid tokens, missing tokens,* or *conflicting tokens* at 34% of these sites. We found CP errors on a wide range of sites, including some of the most popular websites on the Internet and TRUSTe seal holders. We also reviewed the opt-out cookies of 75 network advertisers, and found errors in one of the 11 CPs collected. We were surprised by the large number of errors we were able to detect in CPs through automated analysis alone. We expect that even more errors exist, but discovering them would require manual comparison with sites' human-readable privacy policies or first-hand knowledge of sites' actual privacy practices. The large number of CP errors is troubling and suggests that CPs cannot be relied on for accurate information about website privacy policies with respect to cookies.

We conducted a number of analyses to try to understand why such a large fraction of CPs contain errors. Our results suggest that while some errors are likely introduced through mistakes (e.g. typos or misunderstanding the P3P specification), most appear to result from web administrators writing CPs for the purpose of avoiding IE cookie filtering without considering the accuracy of their CPs. In addition, we found large numbers of websites sharing the same erroneous CPs, including groups of websites hosted on the same domain.

P3P is designed to provide website privacy policies in a computer-readable format that enables automated analysis and decision making. CPs provide a simple way for websites to offer a summary of their privacy practices with respect to cookies in a format that is easily processed by web browsers. The IE web browser uses CPs to make cookie blocking decisions. Thus, CP errors are likely to cause IE to allow cookies that should be blocked under a user's privacy settings to go unblocked, and users who rely on IE's cookie settings may be misled. This problem is exacerbated by the fact that the IE cookie-filtering implementation does not check for CP errors. Thus even the invalid- and missing-token errors, which are a clear violation of the P3P specification, go undetected by IE. Indeed, some websites appear to exploit this IE implementation loophole and publish CPs containing only bogus tokens or omitting tokens in the categories that would cause IE to filter their cookies. A number of online articles also suggest CPs that websites can use to avoid having their cookies blocked, and we found large numbers of sites that copied these suggested CPs verbatim.

CP errors would likely be reduced substantially if IE checked for these errors, and if the articles that informed web administrators about avoiding cookie blocking explained that CPs need to follow the P3P specification and accurately represent privacy practices with respect to cookies. In addition, if the administrators of domains that host large numbers of websites corrected the CPs for their domains, the number of errors would be significantly reduced.

The CP error data we report suggests that many websites are not taking P3P seriously and are behaving in ways that undermine the purpose of the P3P specification. Previous work suggests that errors in full P3P policies are also common [36] [32]. It appears that companies do not currently have sufficient incentives to provide accurate machine-readable privacy policies. Unless regulators use their authority to take action against companies that provide erroneous machine-readable policies, users will be unable to rely on these policies.

Chapter 3

A Large-Scale Evaluation of U.S. Financial Institutions' Standardized Privacy Notices

Although large-scale comparisons of privacy practices across an industry have the potential to illuminate the state of consumer privacy and to uncover egregious practices, the freeform legalese of most privacy policies makes comparisons time-consuming and expensive. Financial institutions in the United States are required by the Gramm-Leach-Bliley Act to provide annual privacy notices. In 2009, eight federal agencies jointly released a model privacy form for these disclosures. While use of the model privacy form is not required, it has been widely adopted. With so many institutions' policies available in a standard format, large-scale comparisons are now more readily achievable.

We built an automated web crawler and document parser for the model privacy form and automatically evaluated 6,191 U.S. financial institutions' privacy notices. We found large variance in stated practices, even among institutions of the same type. While thousands of financial institutions share personal information without providing the opportunity for consumers to opt out, some institutions' practices are more privacy-protective. Statistical analyses show that large institutions and those geographically located in northeastern regions share consumers' personal information at higher rates than all other institutions.

Furthermore, we uncovered institutions in apparent violation of data sharing opt-out requirements mandated by law, as well as institutions making self-contradictory statements. We discuss implications for privacy in the financial industry, issues with the design and use of the model privacy form, and future directions for standardized privacy notices.

3.1 Introduction

When the United States Congress was considering the Gramm-Leach-Bliley Act of 1999 (GLBA), allowing the consolidation of different types of financial institutions, privacy advocates argued

This chapter is largely based on "Are They Actually Any Different? Comparing Thousands of Financial Institutions' Privacy Practices" [46].

that it was important to notify consumers about these institutions' data practices and allow consumers to limit the use and sharing of their data [47]. The act passed with a provision mandating annual privacy notices. In the years that followed, these disclosures were widely criticized for being difficult to read and understand [48]. In response, eight federal agencies jointly released a *model privacy form* in 2009 [49]. This standardized format for enumerating privacy practices was designed to "make disclosure of institutions' information sharing practices and consumer choices more transparent" in an easy-to-read format [49].

Besides making it easier for consumers to find privacy information, standardized privacy notices also enable automated, large-scale comparisons of privacy practices. The idea of providing privacy notices in standardized formats has long held great potential for empowering consumers to compare companies' privacy practices. From standards for machine-readable privacy policies, such as the Platform for Privacy Preferences (P3P) [30], to recent attempts to have humans annotate websites' privacy policies and terms of service [50], much time and energy has gone into attempts to provide privacy information in a standardized format. Unfortunately, these initiatives generally do not reach fruition. For instance, in Chapter 2 we found that websites misuse machine-readable privacy disclosures, while attempts to have humans annotate privacy practices do not scale well.

Although financial institutions in the United States are not required to use the model privacy form to enumerate their privacy practices, the use of this form provides a safe harbor for privacy disclosures under GLBA [49]. As a result, many financial institutions have used this model privacy form to make their mandatory privacy disclosures (which we term *standardized notices* throughout the rest of this paper). This state of affairs provides a rare opportunity for analyzing privacy practices across an entire industry.

To this end, we collected lists of financial institutions in the United States and wrote a computer program that automatically queries Google in search of these companies' standardized notices. Upon finding such a notice, the program automatically parses the standardized notice and feeds the extracted information into a database, enabling a large-scale comparison of financial institutions' privacy practices. Starting from lists of financial institutions from the Federal Reserve (FED), the Federal Deposit Insurance Corporation (FDIC), and the National Credit Union Administration (NCUA), we searched for standardized notices from 19,329 financial institutions, finding standardized notices from 6,191 of these institutions.

We then compared these 6,191 institutions in terms of their data-sharing practices, consumers' ability to opt out of data-sharing, and the personal information the policies say may be collected. To investigate how different factors affect institutions' sharing practices, we further conducted statistical analyses using additional information included in the FDIC list regarding various institutions' characteristics. For additional insight into how competitors compare, we also analyzed the policies of institutions on a Forbes list of the 100 largest banks [51] and a J.D. Power survey of credit card satisfaction [52].

We found wide variance in financial institutions' privacy practices. These differences in privacy practices also distinguished institutions of the same type, suggesting that consumers might have the opportunity to pick a financial institution with more consumer-friendly privacy practices if information to help them find these institutions were more readily available. To that end, we made a website publicly available for consumers to compare thousands of institutions privacy practices. We further found that large institutions as well as those geographically located in the northeastern region of the U.S. are statistically more likely to share consumers' personal information for marketing purposes than all other institutions. Finally, we found deficiencies in both the specification and the use of the model privacy form that may counterintuitively limit consumers' access to information about financial institutions' privacy practices.

In Section 3.2, we summarize the relevant provisions of GLBA and prior work on standardized privacy notices. In Section 3.3, we describe the data set we collected and explain our methodology. We present our results in Section 3.4 and discuss in Section 3.5 our findings and their implications for financial institutions' privacy practices and standardized privacy notices. We include an Appendix with detailed results and screenshots of the model privacy form.

3.2 Background and Related Work

As a result of a combination of state laws, GLBA, and administrative rules, financial institutions issue privacy notices and privacy policies to consumers. In this section, we describe privacy provisions of GLBA, some criticisms of those provisions, and the regulatory development of an optional standardized format for financial institutions' privacy disclosures. We also discuss relevant financial state laws. Finally, we highlight efforts to improve privacy notices beyond just the financial industry, including the creation of formal specifications, standardized formats, and usable privacy notices.

3.2.1 Privacy provisions of GLBA

In this paper we examine financial institutions' annual privacy disclosures that are mandated by GLBA, which was signed into law on November 12, 1999 [53]. GLBA's primary purpose was to encourage competition in the financial services industry by removing barriers that prevented common ownership (affiliation) between commercial banks, investment banks, and insurance businesses [54, 55, 56].

Affiliation between different types of financial services companies presented an opportunity for newly affiliated companies to share information. In response to concerns about the privacy of consumer information, Congress included Title V, known as the Privacy Rule, in GLBA. This rule requires financial institutions to provide annual notices of their privacy policies and practices (15 U.S.C. §§ 6802–6803). The rule also mandates that customers have the right to opt out of data sharing with nonaffiliated third-party companies. However, the Privacy Rule does not mandate that consumers have the right to opt out of sharing between affiliated companies. Thus, joint marketing efforts are exempt from opt-out requirements [57].

Although GLBA's Privacy Rule does not give consumers a general right to opt out of all data sharing, the Fair Credit Reporting Act (FCRA) does give consumers that right for certain types of credit information. The FCRA exempts from its definition of a "consumer report" (the type of communication regulated by the Act) any communication between affiliates. However, this exemption only applies if the communication is "clearly and conspicuously disclosed to the consumer . . . and the consumer is given the opportunity, before the time that the information is initially communicated, to direct that such information not be communicated among such persons" (15 U.S.C. § 1681a(d)(2)(A)(iii)). In other words, consumers must be able to opt out of data sharing about their creditworthiness between affiliates of credit reporting agencies.

The Fair and Accurate Credit Transactions Act of 2003 (FACTA) [58] amended the FCRA to further restrict the use of information shared between affiliates. The rule, called the "Affiliate Marketing Rule," prohibits companies that receive information that would be considered a consumer report if not for § 1681a(d)(2)(A)(iii) from using that information for marketing unless the consumer is given notice and the opportunity to opt out (15 U.S.C. § 1681s-3(a)).

The provisions of GLBA, the FCRA, and FACTA combine to establish three contexts in which financial institutions must provide notice and the opportunity to opt out.[1] GLBA's Financial Privacy Rule applies to the sharing of consumer financial information with nonaffiliates, the FCRA restricts sharing consumer report information between affiliated companies, and FACTA limits when consumer report information shared between affiliates may be used for marketing [59].

3.2.2 Criticisms of GLBA's privacy provisions

The privacy protections offered by GLBA have prompted a range of criticisms. Some critics feel that GLBA offers incomplete or too few privacy protections. For instance, in an examination of GLBA privacy provisions, Janger et al. conclude that GLBA "leaves the burden of bargaining on the less informed party, the individual consumer" [60]. Schiller also argues that the notice provisions provided by GLBA do not go far enough toward providing privacy protections [61]. She recommends that GLBA further restrict information sharing among affiliates. Freeman similarly concludes that GLBA was a good start but "need[s] further refinement" [62], arguing that the "opt-out" provision has made it unlikely that many customers will take the active steps needed to protect their confidential data" [62]. Nojeim also argues that GLBA is incomplete because it does not prevent the flow of personal information among affiliates and uses an opt-out approach, failing to require consumers' active consent [63].

Other critics feel that the protections offered by GLBA are an impediment to the free market. Some economists have claimed that "efforts to protect privacy in the financial services industry (and elsewhere) are obstacles to the functioning of optimally efficient markets" [64]. Lacker, for example, argues that in a perfectly competitive market, financial privacy would be determined by economic forces regardless of the choice mechanisms offered [65]. Those who support open information sharing also often claim that it makes the market more efficient and benefits both financial institutions and consumers. They further claim that other laws, such as the Fair Credit Reporting Act, provide sufficient privacy protections for consumers [66]. In counterpoint, Swire argues that inappropriate disclosure of personal information can easily lead to a "misallocation of resources" [66].

A handful of researchers have examined financial institutions' privacy disclosures. However, all of these past investigations have taken place on a small scale, and many occurred soon after GLBA came into effect. Prior to GLBA, an evaluation of financial institutions' websites conducted by U.S. regulatory agencies found that only 40% of the websites posted a privacy policy [67]. Sheng et al. performed a longitudinal study of 50 financial institutions' privacy policies. They found that although privacy policies became more complete and contained more detailed information about sharing practices after GLBA, the amount of sharing among affiliates and nonaffiliates increased [68]. Antón et al. examined 40 online privacy policies under

[1] Other notice and opt-out requirements may exist for specific types of financial institutions

GLBA and found a lack of standardized vocabulary across the policies, counter to the mandate of GLBA [69].

3.2.3 Development of the model privacy form

A few years after GLBA was enacted, eight U.S. regulators[2] jointly noted wide variations in the privacy notices financial institutions were sending to consumers. They found these notices "difficult to compare, even among financial institutions with identical practices" and questioned "whether such notices comply with the requirement that they be clear and conspicuous" [49]. As a result, regulators started a process to create a standard model for privacy notices that "consumers could more easily use and understand" [49]. Financial institutions, researchers, and communications firms took part in this process.

The process began in the summer of 2004. The regulators hired a communications firm to develop a prototype of the standard notice. To do so, the firm conducted two focus groups and several individual interviews with 60 participants, releasing a report of their findings in February 2006 [70]. Notably, the main goal of the prototype notice was to help consumers understand financial institutions' sharing practices, not necessarily to provide a comprehensive list of the types of personal information that financial institutions collect [70]. In March 2007, the regulators issued the prototype as a proposed model form for public comment.

Following public comments on the proposed model form, the regulators commissioned a quantitative survey designed to evaluate the effectiveness of the revised model form. The survey, which was conducted in the spring of 2008, tested comprehension and usability of the model form as compared with three other styles of notice. Notices of three fictitious banks with different sharing practices were tested among 1,032 consumers recruited from five different US cities. The prototype outperformed the alternative styles tested [71].

In December 2008, Levy and Hastak submitted a report to the regulators analyzing the results of the usability testing [72]. Although participants who tested the proposed prototype better understood the differences in sharing practices, Levy and Hastak found that participants experienced problems understanding how to exercise opt-out rights. The report proposed improvements to reduce the length of the disclosure table and to increase the clarity of opt-out choices. The regulators revised the model form again based on both the Levy-Hastak report and public comments the regulators received after publishing the survey results.

The regulators commissioned the same communications firm that designed the original prototype to conduct validation testing. The firm conducted a 7-participant study and concluded in its February 2009 report that the improvements suggested by Levy and Hastak improved clarity with respect to opt-out choices without affecting understanding of sharing practices [73]. Garrison et al. give a more detailed account of the user testing behind the model forms [74].

In December of 2009, the regulators released the final model privacy form [49]. Figure 3.1 shows a detailed view of the four sections of the model privacy form that we deem most relevant to consumers. Although use of the model privacy form is voluntary, financial institutions may rely on this model privacy form as a safe harbor to provide privacy disclosures [49], potentially

[2] The Office of the Comptroller of the Currency, the Board of Governors of the Federal Reserve System, the Federal Deposit Insurance Corporation, the National Credit Union Administration, the Federal Trade Commission, the Securities and Exchange Commission, the Office of Thrift Supervision, and the Commodity Futures Trading Commission.

spurring the model privacy form's adoption. Notably, this model privacy form has led to one of the first large-scale uses of a standardized format for privacy disclosures, facilitating our large-scale analysis.

3.2.4 State laws

U.S. states have enacted a number of laws limiting financial institutions' ability to share financial data. GLBA includes a provision providing that it does not preempt state laws that are consistent with it. State laws that are inconsistent are invalid only to the extent of the inconsistency (15 U.S.C. § 6807) [75, 76]. A state law with stronger consumer protections is explicitly not inconsistent (and, thus, not preempted). Many states have laws that prohibit financial institutions from disclosing customer information unless that disclosure is authorized or required by law or court order (see Proskauer § 5:6.2 [77] for examples).

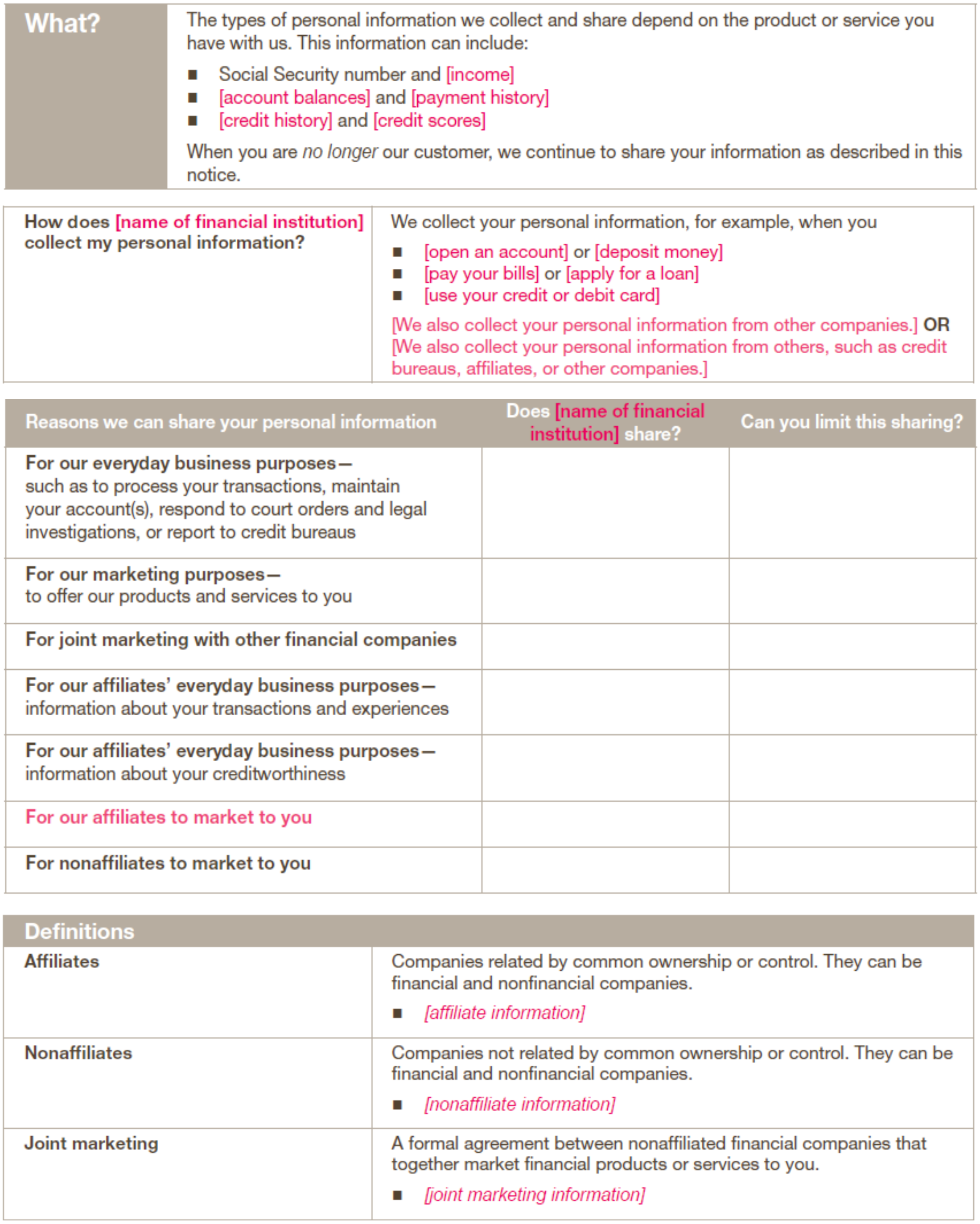
California's Financial Information Privacy Act (Cal. Fin. Code §§ 4050–60) is a notable example of a state law enacted in the wake of GLBA. It was enacted in 2004 with the intent to "afford persons greater privacy protections than those provided in . . . the federal Gramm-Leach-Bliley Act" (*Id.* §4051(b)). CalFIPA requires consumers to opt in before a financial institution may share "nonpublic personal information" with a nonaffiliated third party. It allows nonpublic personal information to be shared between most types of affiliates only after notice and the opportunity to opt out.

Although GLBA seems to explicitly allow state laws with stronger provisions, the affiliate-sharing rule has been held invalid due to preemption under FCRA. In *American Banker's Association v. Lockyer*, 541 F.3d 1214 (9th Cir. 2008), the Ninth Circuit Court of Appeals held that CalFIPA was preempted by the FCRA with regard to the opt-out requirement for the sharing of consumer report information between affiliates. Although GLBA allows state laws with stronger protections for consumers than are provided under GLBA, it does not "modify, limit, or supersede" the FCRA (15 U.S.C. § 6806). The FCRA preempts any state laws that contain provisions "with respect to the exchange of information among persons affiliated by common ownership or common corporate control" (15 U.S.C. § 1681t(b)(2)). Because CalFIPA purported to set different requirements than the FCRA for information sharing between affiliates, the Ninth Circuit ruled CalFIPA invalid with respect to consumer report information.

3.2.5 Privacy policies

The idea that consumers should receive clear notice about privacy is a core principle of many privacy frameworks, including the OECD's 1980 privacy guidelines [78] and the U.S. Federal Trade Commission's Fair Information Practice Principles (FIPPs) [79]. Privacy notice is often presented to consumers in the form of a privacy policy. Overall, privacy notice has been found to impact trust and promote social welfare. For instance, in a study of retail websites, Tang et al. found that the clarity and credibility of privacy notices were crucial for influencing consumer trust [80]. When information about privacy is made accessible to consumers, Tsai et al. found that consumers will pay a premium price to make purchases from more privacy-protective businesses [81].

Unfortunately, a number of issues negatively impact the usability of current privacy policies. Privacy policies are generally written at a very high reading level. For instance, in a study of



What?	The types of personal information we collect and share depend on the product or service you have with us. This information can include: ■ Social Security number and [income] ■ [account balances] and [payment history] ■ [credit history] and [credit scores] When you are *no longer* our customer, we continue to share your information as described in this notice.

How does [name of financial institution] collect my personal information?	We collect your personal information, for example, when you ■ [open an account] or [deposit money] ■ [pay your bills] or [apply for a loan] ■ [use your credit or debit card] [We also collect your personal information from other companies.] **OR** [We also collect your personal information from others, such as credit bureaus, affiliates, or other companies.]

Reasons we can share your personal information	Does [name of financial institution] share?	Can you limit this sharing?
For our everyday business purposes— such as to process your transactions, maintain your account(s), respond to court orders and legal investigations, or report to credit bureaus		
For our marketing purposes— to offer our products and services to you		
For joint marketing with other financial companies		
For our affiliates' everyday business purposes— information about your transactions and experiences		
For our affiliates' everyday business purposes— information about your creditworthiness		
For our affiliates to market to you		
For nonaffiliates to market to you		

Definitions	
Affiliates	Companies related by common ownership or control. They can be financial and nonfinancial companies. ■ *[affiliate information]*
Nonaffiliates	Companies not related by common ownership or control. They can be financial and nonfinancial companies. ■ *[nonaffiliate information]*
Joint marketing	A formal agreement between nonaffiliated financial companies that together market financial products or services to you. ■ *[joint marketing information]*

Figure 3.1: Four primary sections of the model privacy form. From top to bottom, these sections state what information is collected, explain how information is collected, state data-sharing practices, and identify partner companies. These screenshots are taken from the model notice [49]; institutions replace the pink text as they fill out the table. Appendix B.8 contains the full model privacy form.

health websites, Graber et al. found the average privacy policy to require two years of college education to comprehend [82]. Similarly, Jensen and Potts examined 64 privacy policies and found that many were difficult to find and read [83]. The reading level of privacy policies is not the only barrier to comprehension; Ur et al. found instances of privacy policies being unavailable in a user's language, in contrast to the rest of a website [84]. McDonald and Cranor examined the length of privacy policies, estimating that a user would need to spend hundreds of hours a year to read all of the privacy policies relevant to their browsing [85].

For privacy notices to be effective, they must be clear and comparable across websites. Standardized privacy notices—whether human-readable or machine readable—help facilitate large-scale comparison and evaluation [86]. For instance, the Platform for Privacy Preferences (P3P) is an XML-based W3C standard for machine-readable privacy policies that specifies what data will be collected and how it will be used [30]. Cranor et al. conducted a study of several hundred computer-readable privacy policies encoded using P3P. They used automated tools to analyze the data collection, use, and sharing practices encoded in each policy. [87]. Although adopted to some degree, P3P has not received support across different browsers, minimizing its usefulness. Cranor et al. found high rates of syntax errors among the P3P policies they examined [87]. Furthermore, in Chapter 2 we found a number of websites misrepresenting their privacy practices through erroneous or misleading P3P compact policies, which are short strings designed to summarize privacy practices associated with cookies.

Well-designed, standardized formats for privacy notice can overcome many of these obstacles. Furthermore, privacy notices can be compared easily if they are presented in a standardized format. Researchers have examined methods for presenting privacy policies in a standardized, usable manner. For example, Kelley et al. found that displaying privacy policy information in a tabular "nutrition label" format made it easier for users to find information [88]. Even when companies don't provide standardized notice about their privacy practices or terms of use, projects like "Terms of Service; Didn't Read" have used crowdsourcing to put this information into a standardized, usable format [50].

3.3 Methodology

To perform our evaluation of privacy notices, we first compiled a comprehensive list of financial institutions in the United States. Then, we automatically retrieved standardized notices of these institutions and parsed their contents. Finally, we performed quantitative analyses that allowed us identify some of the institutions' characteristics that impact their sharing practices. In this section, we detail these steps.

3.3.1 Obtaining lists of financial institutions

As the first step in searching for U.S. financial institutions' standardized notices based on the model privacy form, we needed a list of these institutions. Having a list of the names and geographic locations of these institutions enabled us to collect privacy disclosures in a systematic way and minimize confusion between banks with similar names (e.g., multiple, seemingly independent banks were called "First National Bank," "Liberty Bank," "Pinnacle Bank," etc.). To this end, we compiled two complementary lists encompassing a total of 19,329 financial institutions. The first list comprised a number of different types of financial institutions. The second list comprised only federal credit unions, which were absent from the first list.

We created our first list of 12,511 distinct financial institutions by merging lists from the Federal Reserve and the Federal Deposit Insurance Corporation (FDIC), two of the largest U.S. government agencies related to the financial industry. To obtain the Federal Reserve (FED) list of 6,588 financial institutions, we made a Freedom of Information Act (FOIA) request. The list of 6,781 financial institutions insured by the Federal Deposit Insurance Corporation is available online. The FDIC list also includes an institution's characteristics, location, assets, and contact information [89]. We merged these two lists based on each institution's "Research, Statistics, Supervision and Regulation, and Discount and Credit" (RSSD) ID number, removing duplicate entries. The RSSD ID uniquely identifies all institutions that have reporting obligations to the Federal Reserve. Although these two lists overlapped to an extent, we found that many institutions were present on only one of these lists. Following the merging process, our list contained 12,511 financial institutions.

We also made FOIA requests to obtain lists of financial institutions from the other main United States government agencies that regulate financial institutions, notably the Consumer Protection Financial Bureau (CFPB) and the Office of the Comptroller of the Currency (OCC). Although these lists together included 101 institutions absent from both the Federal Reserve and FDIC lists, they had much less metadata about the institutions' characteristics. Therefore, we chose to exclude these additional institutions.

Our second list comprised 6,818 credit unions supervised by the National Credit Union Administration (NCUA).[3] The NCUA regulates federal credit unions in the United States. In addition to the name of each credit union, the list contained each institution's full mailing address, as well as information on its peer group.

3.3.2 Retrieving standardized notices

Using Google's search engine, we conducted an automated web crawl to collect instances of the model privacy form. We used the header of the model privacy form, "What does *institution name* do with your personal information," as a search string, inserting the corresponding institution's name. To minimize the chance of accidentally retrieving another institution's standardized notice, particularly in light of the large number of financial institutions with similar names, we restricted each query to a financial institution's website domain using Google's *as_sitesearch* parameter. We had website URLs for most of the institutions in the FDIC list; however, the institutions in the FED and credit unions lists did not include website URLs. To determine the website domain for those institutions missing it, we performed an automated Google query of the string "*Institution name*, *City*, *State*" and took the domain of the first result to be that institution's domain. This heuristic is imperfect, yet we believe it conservatively minimizes false associations (incorrectly attributing a standardized notice to the wrong institution) at the expense of increasing the number of false negatives (not finding notices for institutions that have them available). We obtained up to ten candidate forms for each company and selected the most completed one for further analysis, setting a minimum threshold of elements included to con-

[3] National Credit Union Administration. 5300 Call Report Quarterly Data. `http://www.ncua.gov/DataApps/QCallRptData/Pages/CallRptData.aspx`

sider it valid. Appendix B.1 includes the technical details of our web crawler and standardized notices parser.

Across the 19,329 financial institutions in our two lists, we obtained standardized notices for 6,191 financial institutions. Of the 6,409 institutions whose website domain was known from the FDIC list, we obtained standardized notices for 3,594 institutions (56% of the institutions). Of the 6,102 institutions whose website domain was not listed, we obtained standardized notices for 787 institutions (13%). Finally, of the 6,818 credit unions, none of whose domains were known a priori, we downloaded 44,543 files, which included standardized notices for 1,810 credit unions (27%). The standardized notices from these 6,191 financial institutions make up the data set for all of our further analyses.

3.3.3 Parsing standardized notices

Having selected at most one standardized notice for each institution, our automated parsing program extracted data about each institution's privacy practices. The model privacy form has a strict document structure, including a number of subsections. As the first step in extracting data, we split the standardized notice's text into the sections specified in the model notice, primarily using the four subsections shown in Figure 3.1 in Section 3.2 of this paper.

We wrote regular expressions defining particular patterns based on the specification of the model privacy form [49] and wrote the extracted practices to a CSV spreadsheet.

During the development of our parsing program, we repeatedly tested our parser on small groups of standardized notices and manually checked for instances that were not matched. Based on these manual checks, we iteratively improved our parser to capture rewordings we commonly observed. For instance, we observed "use your credit or debit card" being replaced by the similar statements "use your credit/debit card," "use your credit card," "use your debit card," and "use your ATM card." We adjusted the parser to recognize all of these variants. Similarly, as we detail in Appendix B.2, we iteratively updated our parser to recognize many variants of revision dates.

We paid particular attention to parsing the *disclosure table* (the third table shown in Figure 3.1), which states an institution's data-sharing and opt-out practices across seven different purposes. We initially searched for "Yes," "No," and "We don't share," the values permitted in the specification of the model privacy form [49]. Based on our iterative process, we supported six additional case-insensitive variants ("we do not share," "we don't collect," "we do not collect," "we have no affiliates," "Y," and "N").

That said, it would be intractable to update the parser to recognize every corner case among the thousands of standardized notices. To estimate the accuracy of our automated parser, we manually verified the parser's accuracy on a random sample of 50 institutions' privacy disclosures. For each of the sections of the document we examined, our parser was accurate for between 90% and 100% of documents. We describe this verification process in detail in Appendix B.2.

3.3.4 Analysis

First we analyzed the general prevalence of different privacy practices as indicated in institutions' standardized notices. For instance, we examined the types of information institutions said they collected, the occasions on which institutions said they collected data, and the different sharing practices and opt-out mechanisms institutions presented to consumers.

As a secondary goal, we also investigated whether institutions complied with relevant portions of GLBA and the FCRA, as well as the degree to which institutions deviated from the specification of the model privacy form. We manually verified instances where our parser found idiosyncratic results or where automated analysis suggested violations of GLBA or the FCRA. As part of this analysis, we also visited the webpages of a random subset of 50 institutions to see how the model privacy form was used in practice.

In March 2013, we performed similar analyses on a smaller set of FDIC-insured financial institutions and published preliminary results [46]. In this earlier analysis, we identified 24 institutions whose stated practices in their standardized notice would violate GLBA, the FCRA, or both. In November 2013, we sent a letter on Carnegie Mellon letterhead to the 19 of these institutions for which we were able to find a postal address. This letter outlined the problematic statements in their institution's standardized notice. We discuss these institutions' responses to our letters in Section 3.4.3.

We further investigated whether the institution type, as reported by the Federal Reserve, was correlated with the institution's privacy practices. In addition to institution types reported by the Federal Reserve, we considered all federal credit unions to form an additional institution type, which we termed *credit union*.

Finally, using the subset of institutions for which we had additional information regarding institutions' characteristics, we investigated which of those characteristics were correlated with their sharing practices. We joined the data we parsed automatically from standardized notices with each institution's characteristics, as reported in the FDIC Institution Directory [89] and list of institutions from the Federal Reserve. In the FDIC list, these characteristics included an institution's geographic region, assets, and type of institution. We used these characteristics as independent variables and the different sharing practices as dependent variables to build logistic regression models. We build a regression model for six of the seven sharing practices in the disclosure table; we excluded the "for our everyday business purposes" row, for which nearly all institutions had identical practices.

3.4 Results

We first provide an overview of institutions' privacy practices, including the reasons for which they share data and the means through which consumers can opt out. We found substantial variation in practices across institutions, as well as dozens of companies that appear to be violating the law by not offering legally mandated opt-outs. To understand more fully whether competing companies' privacy practices differ, which would provide an opportunity for consumer choice, we also examined the data-sharing practices of companies that appear on lists of recommended banks and credit cards, again finding a wide range of practices. Then, we conducted statistical analyses to investigate how institutions' characteristics such as size, geographic location, and type affect their sharing practices.

Finally, we present observations about misuse of the model privacy form and discuss how the design of the model privacy form might impact institutions' transparency with respect to data-collection practices.

3.4.1 Data sharing practices

In this section, we describe financial institutions' stated data-sharing practices. We discuss with whom data is shared, reasons why data is shared, and the mechanisms institutions give consumers for opting out of data sharing when applicable. We also present institutions' disclosures of the information they collect and how they collect it. We argue that these final two disclosures are not particularly informative.

Overall, our results show that sharing and opt-out practices vary widely across financial institutions. This variety of practices suggests that helping consumers easily compare institutions' practices could empower them to select companies that better align with their privacy expectations.

With whom data is shared

The standardized notices present consumers with information about how a financial institution shares their data with other companies. These disclosures discuss affiliates, which are financial or nonfinancial companies that are "related by common ownership or control" to the institution making the disclosure. The disclosures also discuss nonaffiliates, which are third parties that are not affiliates or joint employees. In the definitions section of the model privacy form, institutions not only provide boilerplate definitions of the terms "affiliates," "nonaffiliates," and "joint marketing," but also list their partners in each category.

Institutions varied starkly in their practices, as shown in Table 3.1. On the question of sharing with affiliates, 28% of institutions said they have affiliates and share with them, 25% said that they do not share with their affiliates, and 43% said that they do not have any affiliates. In contrast, 12% of institutions said they share with nonaffiliates, 66% said they do not, and only 18% said they do not have nonaffiliates. Joint marketing practices also differed; 42% of institutions said that they engage in joint marketing whereas 55% said that they do not. This section of the model privacy form was missing entirely for 0.9% of institutions, and a handful of institutions defined the terms without providing information about their own practices (labeled *blank* in Table 3.1). The differences we noted suggest that financial institutions follow considerably different practices.

Reasons data is shared

The model privacy form's disclosure table lists seven reasons for which an institution might share data, along with the institution's own practices for each of these reasons. For each of these reasons, institutions vary from not sharing data at all to sharing data without offering an opt-out. Notably, a few institutions' policies state that they do not offer opt-outs for data sharing even when legally required to do so.

The disclosure table comprises seven rows, each representing a reason an institution might share data, such as the institution's everyday business purposes or joint marketing purposes. One row, "for our affiliates to market to you," is optional for institutions that do not have affiliates, whose affiliates do not use personal information, or whose affiliates have a separate notice [49]. Of the 6,191 institutions in our data set, 3,754 institutions (61%) omitted this row. Note that we did not check for consistency between the disclosure table and the definitions section of the model privacy form.

Practice	Number of institutions	Percentage of total
Affiliates		
Share with affiliates	1,726	28%
Do not share	1,543	25%
No affiliates	2,632	43%
Blank	237	4%
Nonaffiliates		
Share with nonaffiliates	730	12%
Do not share	4,038	66%
No nonaffiliates	1,085	18%
Blank	285	5%
Joint Marketing		
Jointly market	2,575	42%
Do not jointly market	3,356	55%
Blank	207	3%

Table 3.1: The data-sharing practices of the institutions in our primary data set. *Blank* indicates that the institution defined the term, yet provided no information about its own practices. We did not observe this section for 53 of the 6,191 institutions.

We grouped institutions' practices into three primary categories based on their responses to the questions, "Does [institution name] share?" and, "Can you limit this sharing?" We labeled institutions that answered "no" to the first question as *does not share*. Institutions that responded "yes" to the first question and "yes" to the second question provide an opt-out for this sharing, so we labeled those institutions *share, opt-out*. We assigned the label *share, no opt-out* to institutions that answered "yes" and "no," respectively. When a particular row of the table was not parsed, we labeled that value *missing*. As we discuss further in Section 3.4.5, we assign the label *illogical* when answers to these two questions are contradictory (e.g., an institution says it shares in the first column, but says it does not share in the second); this occurs for 13 to 42 institutions (0.2%–0.7%) per row.

Companies are required to provide opt-outs for some types of data-sharing, but are not required to do so in other cases. In particular, institutions that share information about creditworthiness with affiliates, or that share with either affiliates or nonaffiliates for marketing purposes, must provide an opt-out. In Section 3.2.1, we discussed the legal basis for these mandatory opt-outs in detail. Institutions that share for "our marketing purposes," "for joint marketing," or that share information about transactions and experiences with affiliates "may choose to provide an opt-out" [49].

Table 3.2 summarizes institutions' sharing practices. Where not required to provide an opt-out, most institutions chose not to provide one. Almost all institutions shared personal information for their everyday business purposes without offering an opt out. More than half of the institutions (61.9%) said they share "for our marketing purposes" without offering an opt-out, and a third (33.0%) said they share "for joint marketing" without an opt-out. Fewer (21.5%)

Reason for sharing personal information	Does not share		Offers opt-out		No opt-out		(Missing)	
For our everyday business purposes– such as to process your transactions, maintain your account(s), respond to court orders and legal investigations, or report to credit bureaus	45	0.7%	9	0.1%	6,016	97.2%	108	1.7%
For our marketing purposes– to offer our products and services to you	1,808	29.2%	410	6.6%	3,832	61.9%	127	2.1%
For joint marketing with other financial companies	3,434	55.5%	563	9.1%	2,044	33.0%	124	2.0%
For our affiliates' everyday business purposes– information about your transactions and experiences	4,492	72.6%	158	2.6%	1,331	21.5%	189	3.1%
For our affiliates' everyday business purposes– information about your creditworthiness *[Opt-out mandatory]*	5,317	85.9%	572	9.2%	80	1.3%	189	3.1%
For our affiliates to market to you *[Opt-out mandatory when sharing; row may be omitted in certain cases]*	1,682	27.2%	715	11.5%	21	0.3%	3,754	60.6%
For nonaffiliates to market to you *[Opt-out mandatory when sharing]*	5,459	88.2%	455	7.3%	31	0.5%	204	3.3%

Table 3.2: A summary of 6,191 financial institutions' practices for sharing consumers' personal information. Institutions self-reported these practices in the model privacy form's disclosure table. Values that are missing could be caused by an institution omitting that row of the table, or by an error in our parser.

said they share information about transactions and experiences "for affiliates' everyday business purpose" without an opt-out.

Although many institutions did not offer an opt-out if not required to do so, some institutions did not share data or voluntarily chose to offer opt-outs. If comparative privacy information were easily accessible, consumers could choose to do business with the more privacy-protective institutions.

Opt-out mechanisms

The mechanism for opting out of data sharing could impact consumers' likelihood to opt out. We parsed the contents of the "to limit our sharing" section of the model privacy form, searching for instructions on opting out via mail, email, web, and telephone. The opt-outs offered are shown in Table B.14. Overall, 20.5% of institutions offer at least one opt-out mechanism. We observed 627 institutions that provided exactly one mechanism, 491 institutions that provided two different mechanisms, and 152 institutions that provided at least three different mechanisms. did not provide any of these four opt-out mechanisms.

Non-computer-based opt-out mechanisms were more prevalent than computer-based methods. Of the institutions offering an opt-out, 59.9% allowed consumers to opt out over the phone,

Opt-out mechanism(s)	Institutions	Percentage
Only phone	391	30.8%
Phone and website	265	20.9%
Only postal mail	217	17.1%
Phone and postal mail	153	12.0%
Three or more mechanisms	152	12.0%
Phone and email	46	3.6%
Postal mail and website	25	2.0%
Only website	17	1.3%
Only email	2	0.2%
Postal mail and email	1	0.1%
Website and email	1	0.1%

Table 3.3: Institutions' opt-out mechanisms. Overall, 1,270 institutions offered an opt-out. The most common opt-out mechanisms were phone, website, and postal mail.

via postal mail, or using either mechanism. We counted institutions as providing a postal mail opt-out if they either instructed consumers to send mail to a particular address or, more popularly, provided a detachable, mail-in form to fill out. For 48.1% of institutions, we observed a full mail-in form. Computer-based opt-outs were relatively less popular; 28.2% of institutions let consumers opt out via email or a website.

What information is collected

The first section of the model privacy form discloses "the types of personal information that the institution collects and shares" based on a predefined list of 24 types of information financial institutions commonly collect. The model privacy form specifies that the term "Social Security number" be the first bullet, followed by exactly five of the following 23 terms: "income; account balances; payment history; transaction history; transaction or loss history; credit history; credit scores; assets; investment experience; credit-based insurance scores; insurance claim history; medical information; overdraft history; purchase history; account transactions; risk tolerance; medical-related debts; credit card or other debt; mortgage rates and payments; retirement assets; checking account information; employment information; wire transfer instructions" [49]. In total, exactly six terms should be arranged in three bullet points, as shown in Figure 3.1 in the background section of the paper. The main design objective of this section of the model privacy form was to familiarize customers with the concept of personal information, but not necessarily to provide a comprehensive list of the types of personal information that institutions collect [70]. Unfortunately, given that institutions are told to include exactly six out of 24 data types, the omission of a data type does not provide any meaningful information about whether or not the institution collects that type of data.

We parsed this section, searching for "Social Security number" and the aforementioned 23 terms, as well as close variants. The most common terms institutions chose to include were account balance (5,493 institutions), payment history (4,902), credit history (4,891), income (3,428), credit scores (2,752), and transaction history (2,165). Notably, these are the six terms listed in pink font (intended to be replaced by financial institutions) in the model privacy form. Furthermore, we expect that few consumers would be surprised if a financial institution collected any of the types of information an institution is permitted to list in this section.

As a result, the current requirements do not provide transparency of collection practices. Customers with access to different institutions' notices would not have a complete perspective of those institutions' data-collection practices. To provide more useful information to consumers, companies could be required to list all data they collect, or to disclose any types of data they collect that might surprise consumers.

In addition, while having a standardized language for data collection is necessary to enhance transparency and facilitate comparison of companies' practices, we found that some of the terms are redundant and potentially ambiguous. For example, it would be difficult for an average consumer to differentiate between "transaction history" and "transaction or loss history." Similarly, it is unclear whether "account balance," "payment history," and "transaction history" are all part of "checking account information." On the other hand some institutions listed additional types of data they collect outside of those specified for use in the model privacy form. Taken together, these results suggest the need to improve this section of the model privacy form to enhance transparency and account for all institutions' practices.

How information is collected

On the second page of the model privacy form, financial institutions are required to say how they collect consumers' information, again using phrases from a predefined list. The specification of the model privacy notice states that "institutions must use five (5) of the following terms to complete the bulleted list for this question," followed by a list of 34 occasions [49].

The five most frequent terms were simply the five listed in pink as examples in the model privacy form [49]: "open an account," "apply for a loan," "use your credit or debit card," "deposit money," and "pay your bills." On the opposite end of the spectrum, only one institution noted collecting information when consumers tell them about investment or retirement earnings, while no institutions said they collect information when consumers sell securities to them.

Given that institutions are permitted to include only five terms, the omission of a term does not provide any meaningful information about whether or not the institution collects data during that type of event. Such a limitation reduces institutions' transparency and does not benefit consumers.

Furthermore, many of the current terms may not be very informative because they are obvious. Some services requested by customers obviously necessitate collection of personal information. For example, it may not be necessary to tell people that their personal information will be collected when they open an account or apply for a loan, given the paperwork involved in doing either. It might be more useful to inform consumers about situations when it is less obvious that personal information will be collected.

The model privacy form also contains disclosures about other sources that provide data to an institution. Under the section titled, "How does *name* collect my personal information?" institutions must include either of the following statements if they apply to their practices: "We also collect your personal information from others, such as credit bureaus, affiliates, or other companies," or, "We also collect your personal information from other companies" [49]. We observed that 82.9% of institutions collect additional information from credit bureaus, 83.4% do so from "other companies," and 73.2% collect data from affiliates.

3.4.2 Comparing similar institutions

The previous analyses uncovered differences in sharing practices across all institutions, yet such a general analysis does not show the degree to which direct competitors or institutions providing comparable services have similar privacy practices. One might assume that differences in practices result from institutions offering different types of services. When similar institutions vary in privacy practices, a consumer armed with this information could choose where to do business, enabling privacy choice.

In this section, we compare the practices of similar institutions. First, we split the institutions into different types, as defined by the Federal Reserve. We also added all federal credit unions from the NCUA list as an additional type of financial institution. We eliminate institution types for which we obtained fewer than ten institutions' standardized notices. The different types of institutions are shown in Table 3.4.

Even among the same institution types, practices differed. Figure 3.2 shows a comparison of both institutions of the same and different types. In that figure, the presence of different colors in a horizontal bar indicates institutions of the same type that differ in their practices. We do not present a graph of sharing for an institution's own "everyday business purposes" because nearly all institutions shared data for that purpose without offering an opt-out.

In addition to widespread data sharing for "everyday business purposes" by all type of institutions, between 53.4% and 79.2% of institutions of each type shared data for their own marketing purposes. Whereas only 9.5% of credit unions chose not to share data for their marketing purposes, 44.0% of state commercial banks supervised by the FDIC did not share data for this purpose. Between 1.2% and 16.3% of institutions in each specialization shared data for this purpose, yet offered an opt-out.

Opt-outs were particularly common for sharing related to affiliates' marketing purposes. Note that institutions that shared data for this purpose were required to offer an opt-out. Between 22.0% (credit unions) and 65.6% (financial holding companies) of institutions shared data for affiliates' marketing purposes, yet said that consumers could limit this sharing by opting out. Opt-outs were comparatively less common for types of sharing for which institutions were not required to provide an opt-out; no more than 24.5% of institutions in a category voluntarily offered opt-outs.

The 126 financial holding companies whose standardized notices we obtained had less consumer-friendly sharing practices than all other types of institutions. While 62.4% of financial holding companies shared data about customers' transactions and experiences with affiliates without offering an opt-out, no more than 35.0% of the institutions in any other category did the same. Similarly, only 34.4% of financial holding companies did not share data for "affiliates to market to you," whereas 53.1%–75.9% of institutions in the other categories chose not to share data for this reason.

Institution Type	Description	Examples
Bank Holding Company (BHC)	Companies that own or control one or more U.S. banks and which are supervised by the FED	Pinnacle Bancorp Inc.
Commercial Bank - OCC (N)	Companies that engage in various lending activities and which are supervised by the OCC	Wells Fargo Financial National Bank
Commercial Bank - FED (SM)	Companies that engage in various lending activities and which are supervised by the FED	First State Bank of Colorado
Commercial Bank - FDIC (NM)	Companies that engage in various lending activities and which are supervised by the FDIC	Farmers State Bank
Credit Union	Institutions created and operated by its members, who share profits amongst them. Supervised by the NCUA	Lafayette Credit Union
Financial Holding Company (FHD)	Companies engaged in a broad range of banking-related activities, including insurance underwriting, securities dealing and underwriting, financial and investment advisory services, merchant banking, issuing or selling securitized interests in bank-eligible assets, and generally engaging in any non-banking activity authorized by the Bank Holding Company Act. They are supervised by the FED	Capital One Financial Corporation
Savings and Loan Holding Company (SLHC)	Companies that directly or indirectly control one or more savings association	AJS Bancorp Inc.
Savings Association - OTS (SA)	Companies that accept deposits primarily from individuals and channels its funds primarily into residential mortgage loans and which are supervised by the OTS	Century Savings and Loan Association
Savings Bank - FDIC (SB)	Companies organized to encourage thrift by paying interest dividends on savings and which are supervised by FDIC	Royal Savings Bank

Table 3.4: The 9 different institution types that we analyzed and compared. With the exception of credit unions, this classification is provided by the Federal Reserve [90].

Large banks and credit card companies

We also examined even more directly whether consumers might be able to choose between more and less privacy-protective competitors. To do so, we compared the institutions on a list compiled by Forbes [51] of the 100 largest banks, as well as the institutions on a list compiled by J.D. Power & Associates of consumer satisfaction with Credit Card Companies [52]. Even among companies in these lists, we found differences in privacy practices, suggesting that making privacy practices more salient could empower consumers to choose more privacy-protective institutions. In addition to the aforementioned categories of primary specialization, Figure 3.2 includes graphs for the *large banks* and *credit card companies* we discuss in this section.

From the Forbes list of the 100 largest banks in the U.S. [51], we obtained model privacy forms for 73 banks. Some of the remaining institutions used image files of the standardized notice, which we could not parse. Others did not appear to have a standardized notice. Since a consumer might choose from among these large banks, we investigated how their privacy practices compare. Table B.1 in the appendix provides a summary of large banks' practices, while Table B.2 in the Appendix notes each individual bank's practices. Relative to financial institutions overall, the large banks tended to be less privacy-protective. The proportion of large banks that shared data was larger than the proportion of institutions in each of the nine primary specializations that did the same for five of the six types of sharing shown in Figure 3.2. For example, 90.4% of large banks shared data for affiliates' marketing purposes, whereas only between 24.1% (credit unions) and 65.6% (financial holding companies) of institutions in each of the nine specializations did the same.

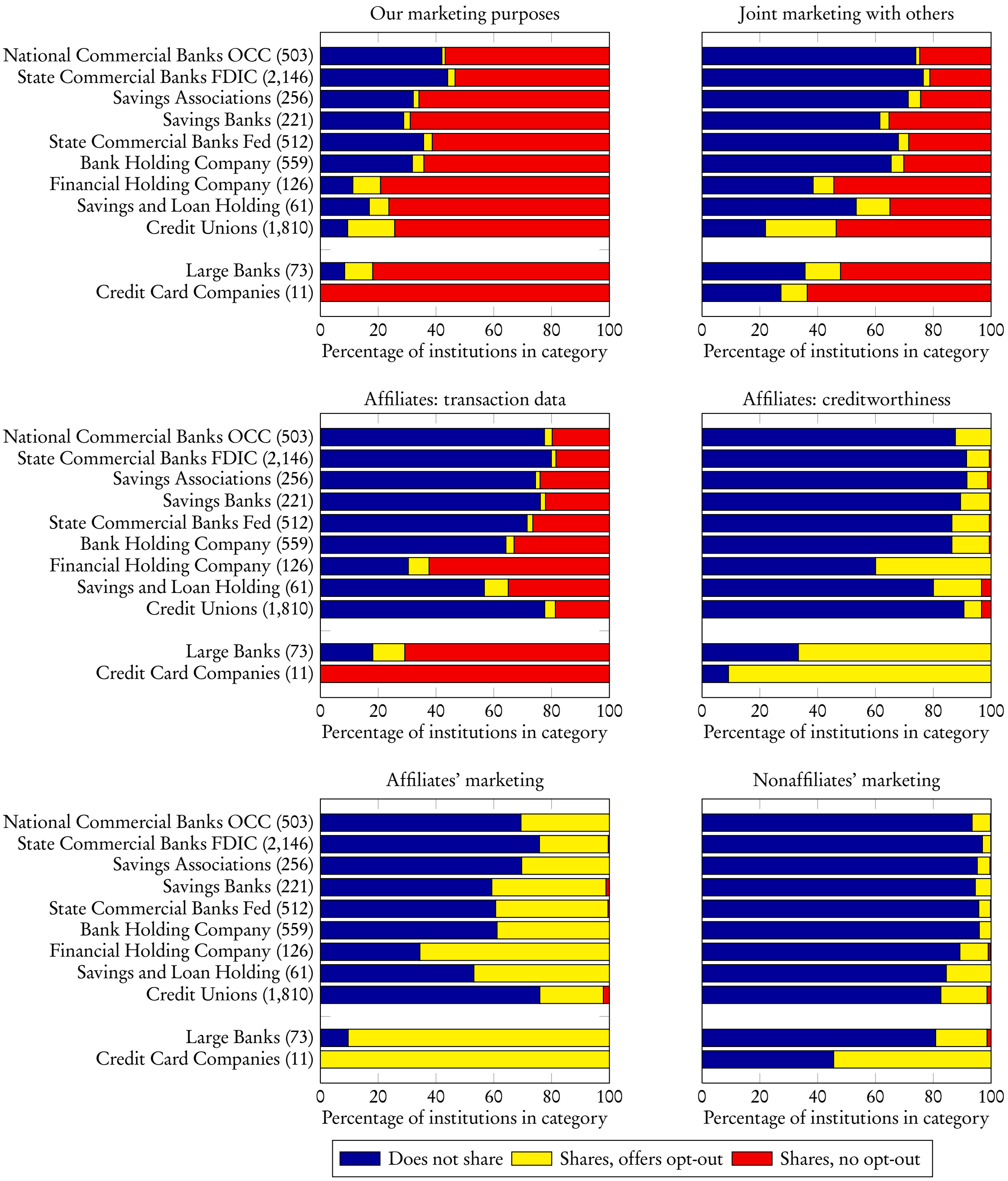


Figure 3.2: The prevalence of sharing practices from the disclosure table. We exclude missing data; notably, many institutions did not disclose a policy in the optional "for our affiliates to market to you" category.

Credit Card Companies' Practices

We also analyzed the sharing practices of the eleven credit-card companies listed in a consumer-satisfaction survey conducted by J.D. Power and Associates [52]. Most of these companies shared data for many reasons, yet a few had more privacy-protective practices for certain types of sharing. However, for the company's own marketing and for providing affiliates information about transactions and experiences, all eleven credit card companies shared data without offering an opt out. Similarly, for affiliates' marketing purposes, all eleven credit card companies shared data, though all did offer an opt-out.

Eight of the eleven companies said they share consumers' personal information without offering an opt-out for "our marketing purposes," "joint marketing," and "affiliates' everyday business purposes - transactions and experiences." Only GE Capital, U.S. Bank, and Wells Fargo said they do not share for joint marketing. Similarly, more than half of the companies said they share for "nonaffiliates to market to you." The practices of each credit-card company are listed in Table B.3 in the appendix.

3.4.3 Compliance with FCRA and GLB

As discussed in Section 3.2.1, GLBA prohibits financial institutions from sharing nonpublic personal information with nonaffiliated third parties unless the institution offers consumers the opportunity to opt out of that sharing. Similarly, the FCRA mandates the provision of an opt-out before information about consumers' creditworthiness may be shared with affiliates and, as amended by FACTA, mandates the provision of an opt-out before consumer report information may be shared with affiliates for marketing purposes. While FCRA and FACTA only apply to credit reporting agencies and not necessarily to other types of financial institutions, the opt-out requirements established in the privacy model are not restricted to credit reporting agencies [49].

We found 109 institutions that stated both that they share for one or more of these reasons and that consumers cannot limit this sharing. We manually verified that each institution's standardized notice was parsed correctly. If the policies these institutions state in their standardized notices represent their actual practices, these institutions are in violation of the the law. We list these institutions in Section B.3 of the appendix. A total of 66 institutions said they shared information about creditworthiness "for our affiliates' everyday business purposes" and said that consumers could not limit this sharing. Furthermore, 20 institutions did the same "for our affiliates to market to you," while 30 institutions followed the same practice "for nonaffiliates to market to you." Note that some institutions had more than one violation, therefore the total number of violations don't add up to the total number of non-compliant companies.

Responses to notification of non-compliance

In a previous analysis of 3,422 standardized notices we conducted in March 2013, we found 24 companies whose opt-out practices appeared to be in violation of the FCRA, FACTA, or GLBA [46]. In November 2013, we contacted the 19 companies for which we could find a mailing address. We mailed each company a letter on Carnegie Mellon letterhead to inform them about the problematic assertions in their standardized notice.

Five institutions formally responded to us. All five institutions agreed that their standardized notices contained mistakes. All five institutions updated their standardized notices to comply with the law. Furthermore, we observed that four companies that did not respond to us also updated and corrected their notices. The remaining 15 institutions' stated practices remain in apparent violation of the law.

In July 2014, we sent letters to 76 credit unions and 20 other financial institutions; which based on our updated analysis, their opt-outs appear to be in violations of U.S. We did not find contact information for the other 13 seemingly non-compliant institutions.

3.4.4 Factors correlated with sharing practices

For the subset of institutions in the FDIC list, which provided additional metadata for each institution listed, we investigated how different characteristics of financial institutions correlated with those institutions' privacy practices. The factors we investigated included the institution's size in terms of assets, the type of institution according to the FED classification, the geographic region where the institutions' headquarters were located, whether the institution had been granted any trust power to conduct fiduciary activities [91], and whether the institution was owned by shareholders. We list these factors and provide details in Table 3.5. We selected this subset of characteristics from a larger set in the Federal Deposit Insurance Corporation (FDIC) directory [89]. Our goal was to build the most parsimonious statistical models that included proxy variables that accounted for the most relevant characteristics of the institutions. We used the experience gained in our pilot paper to improve our statistical models [46].

For example, there were a number of variables in the FDIC directory that measure the size of the institutions in different manners, those included equity, income, number of offices, and whether the company was a Bank Holding Company (BHC). We decided to measure institution's size using the total assets' variable. Similarly, there were various variables indicating institutions' locations and we decided to use the OCC four districts to categorize them in four general regions. Also, other variables in the directory such as charter agency and regulator were part of the FED categories; which consider both specialization and regulator.

To evaluate the impact of these factors on institutions' sharing practices, we built logistic regression models. In total, we built six regression models corresponding to six of the seven practices listed in the disclosure table. We chose not to build a model for sharing associated with the institution's everyday business purposes because that practice varied minimally. We gradually increased the number of variables in our models, always starting with assets; which was clearly a strong predictor based on previously performed proportionality χ^2 tests. Then added location, institution type, and additional indicator variables. We also switched the order in which variables were added and checked the models fits. We finally selected the models with the lowest residual error.

When an institution did not share consumers' personal information for a particular purpose, the binary outcome variable in the regression took the value 0. When an institution shared information, regardless of whether it offered an opt-out, the outcome variable took the value 1. We also built models where the outcome variable had three levels: not sharing, sharing with an opt-out, and sharing without an opt-out. The results of these models were similar to the binary outcome models, and we report results from the binary model in this paper as those are easier to interpret.

Factor	Definition	Possible values	Control category
Assets bracket*	The sum of all assets owned by the institution. Includes cash, loans, securities, and bank premises, but not off-balance-sheet accounts	Very small, Small, Medium, Large, Very large	Very small
Institution Type	Classification of institutions according the the Federal Reserve	Commercial bank supervised by the OCC (N), commercial bank supervised by the Federal Reserve (SM), commercial bank supervised by the FDIC (NM), saving bank supervised by the FDIC (SB), savings association supervised buy the OTS (SA)	NM
OCC District	OCC District where the institution is physically located (see discussion in Section 3.4.4)	Northeastern, Southern, Central, Western	Western
Ownership type	Whether the institution is owned by shareholders (Stock) or not (Non-stock)	Stock, Non-stock	Stock
Trust Powers	Trust powers are defined on a per-state basis	Yes, No	No
Metro Statistical Area	Is the institution in a region with at least one urbanized area with population ≥50,000?	Yes, No	No

Table 3.5: Independent variables considered in our logistic regression models. *We created five assets brackets based on the percentiles of the assets variable distribution (Mean = 1.389 B, Min = 3.7 M, Max = 360 B). Very small ($x < 25\%$), Small ($25\% < x < 50\%$), Medium ($50\% < x < 75\%$), Large ($75\% < x < 90\%$), and Very large ($90\% < x$)

Our logistic regression models revealed a number of factors that were significantly correlated with institutions' privacy practices (Table 3.6). Chief among these factors were the institution size (measured in terms of assets) and the OCC District where the institution was geographically located. The type of institution was a significant factor only for own marketing, affiliate and non-affiliate marketing practices. We discuss the impact of each of these characteristics in the following section and present detailed results for each regression model in Section F.3 in the Appendix.

Institutions' size

We found consistently for all sharing purposes that the larger the institutions the more they share. Table B.5 in the Appendix shows the fraction of institutions in the different assets' brackets that don't share, share and offer opt-outs, and share without offering opt-outs. For example, only 10.5% of institutions in the 25th percentile shared for joint marketing purposes without offering opt-out choices, while 54.4% of institutions above the 90th did so. Similarly, only 1.4% of institutions in the 25th percentile shared with non-affiliates to market users and 9.1% of institutions above the 90th percentile did so. Our regression models shown in Table F.3 in

Factor	Control category	Own marketing	Joint marketing	Affiliates' (Trans.)	Affiliates' (Credit.)	Affiliates' marketing	Non-affiliates' marketing
Assets bracket	Very small	↑	↑	↑	↑	↑	↑
OCC District	Western	↓	↑	↓	N/A	↑	↑
Institution type	Commercial bank, FDIC	↑	N/A	N/A	N/A	↑	↑
Trust Powers	No powers	N/A	↑	↑	N/A	N/A	N/A
Ownership type	Stock	N/A	N/A	N/A	↓	N/A	N/A

Table 3.6: Summary of characteristics that significantly impact sharing practices. ↑ and ↓ respectively denote an increase and decrease in sharing with respect to the control category. N/A denotes that the variable was not included in the corresponding final model.

the Appendix show the specific effect of each assets' bracket on sharing. For example, when compared with a small institution, the odds that a very large institution would share for joint marketing purposes are increased in more than 10 times (i.e., $exp(2.39)$) and the odds that a very large institution would share with non-affiliates to marker users are increased in more than 6 times (i.e., $exp(2.39)$). It is important to give special attention to joint marketing practices as the principal reason why the GLBA exception to permit joint marketing with non-affiliates without requiring institutions to offer an opt-out was to allow small institutions to compete with large ones [92]. Nevertheless, we have found that large companies are more likely to share for this purpose than small companies.

Geographic location

We also found the geographical location of the institution to be significantly correlated with its sharing practices. Table B.6 in the appendix contains detailed results of how practices vary across OCC regions.[4] Only 30.3% of institutions in the Northeastern region didn't share customers' information for their own marketing purposes, while 66.3% shared without offering an opt-out. In contrast, the proportions of companies in the Southern region that shared and did not offer an opt-out (50.4%) and didn't share (47.2%) information are roughly equal. We also found differences in sharing for joint marketing. Whereas 32.9% of institutions in the Northeastern

[4] The states in each of the four OCC region are Northeastern: Connecticut, Delaware, DC, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Puerto Rico, Rhode Island, U.S. Virgin Islands, Vermont, Virginia, and West Virginia; Southern: Alabama, Arkansas, Florida, Georgia, Louisiana, Mississipi, Oklahoma, Tennessee and Texas; Central: Illinois, Indiana, Kentucky, Michigan, Minnesota, Ohio, and Wisconsin; and Western: Alaska, American Samoa, Arizona, California, Colorado, Guam, Hawaii, Idaho, Iowa, Kansas, Missouri, Montana, Nebraska, Nevada, New Mexico, Oregon, South Dakota, States of Micronesia, Utah, Washington, and Wyoming

region shared for joint marketing without offering an opt-out, less than 23% of institutions in the Southern and Central regions did so.

These results show that there are significant differences in sharing practices across geographical regions, and these differences ultimately impact customers in those regions. Our regression models allowed us to investigate the specific effect of geographic location for each of the sharing purposes. Institutions in the Northeastern OCC region shared at a higher rate than those in the Western region for both joint marketing ($p = 0.01$) and for affiliates to market users ($p < 0.001$). Similarly, institutions in the Central OCC region shared at a higher rate than those in the Western region for both joint marketing ($p = 0.05$) and for non-affiliates to market users ($p = 0.02$). In general, institutions in the Southern region appear to share at the same or lower rate than institutions in the Western region, and institutions in the Central or Northeastern regions appear to share at higher rate than institutions in the Western region. We looked closer at differences between states in each of the four OCC regions. We selected the state with the largest number of institutions in each of those regions. Table 3.7 shows these states' practices of sharing for joint marketing and for affiliates to market users, The per-state results were consistent with the OCC-region results. In particular, institutions in New York (Northeastern region) shared more than institutions in the other three states for both joint marketing without offering opt-out choices (30.9%) and affiliate marketing (47.6%). Conversely, institutions in California (Western region) shared less than institutions in the other three states for both joint marketing without offering opt-out choices (8.3%) and affiliate marketing (16.1%). It is also important to remember, as mentioned in the related work, that the California's Financial Information Privacy Act (CalFIPA) requires consumers to opt in before a financial institution may share "nonpublic personal information" with a nonaffiliated third party.

Sharing practice	**Texas**		**Illinois**		**California**		**New York**	
	(Southern)		**(Central)**		**(Western)**		**(Northeastern)**	
Joint marketing with other financial companies (N = 775)*								
Don't Share	213	78.0%	207	74.7%	126	87.5%	55	67.9%
Share & Opt-Out	6	2.2%	3	1.1%	6	4.2%	1	1.2%
Share & No Opt-Out	54	19.8%	67	24.2%	12	8.3%	25	30.9%
For our affiliates to market to you (N = 287)*								
Don't Share	58	73.4%	84	80.8%	52	83.9%	22	52.4%
Share & Opt-Out	21	26.6%	20	19.2%	10	16.1%	19	45.2%
Share & No Opt-Out	0	0.0%	0	0.0%	0	0.0%	1	2.4%

Table 3.7: Sharing practices of four representative states. Overall, institutions in California shared less than institutions from the other three states and institutions in New York shared more than institutions from the other three states. Differences were statistically significant at α=0.05 using a χ^2 proportionality test.

Institution type

The type of institution significantly impacted three of the six studied practices. Table B.7 in the Appendix shows that in comparison to other types of institutions, commercial banks supervised by the FDIC most frequently did not share data for their own marketing purposes, and for affiliates and non-affiliates to market users. Our regression models also show that savings association share significantly more than commercial banks supervised by the FDIC ($p = 0.03$). Other commercial banks also share at higher rate than FDIC commercial banks for both affiliates and non-affiliates to market users ($p < 0.05$). In general the type of institution impact differently sharing for own marketing practices and both sharing for affiliates and non-affiliates to market users; however, they don't impact either joint marketing or sharing for everyday business purposes.

Other factors

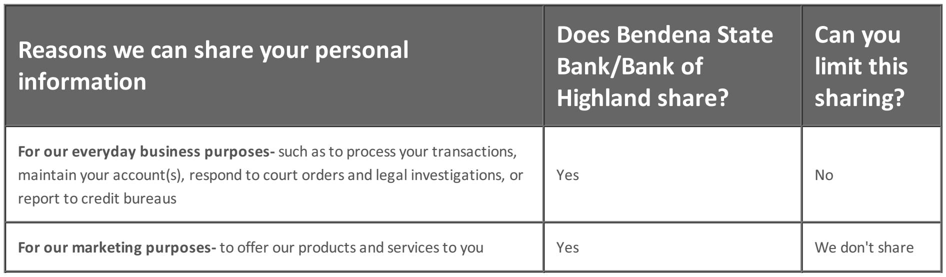
Two additional characteristics affected joint marketing sharing and sharing for everyday business purposes practices. In particular, banks with granted trust powers shared at a significantly higher rate for joint marketing and everyday business purposes (transactions and experiences). Trust powers are granted at the state level under criteria that vary by state [91] and are correlated with the institution's size. The larger the institution, the more likely it will have trust powers. Nevertheless, regardless of such correlation, our regression model shows that, even when controlling for institution' size, institutions with trust powers showed to have an increasing effect in sharing. We also found that companies owned by shareholders were more likely to share creditworthiness information for their affiliates' everyday's business practices.

3.4.5 Misuse of the model privacy form

During our manual analyses of standardized notices during the development and verification of our parser (described in Appendix B.2), we noticed deviations from both the letter and the goal of the model privacy form. In this section, we discuss ways in which financial institutions deviated from the specification [49] of the model privacy form.

Self-contradictory statements

As we iteratively improved our parser, we noticed self-contradictory statements in some institutions' standardized notices. One egregious, yet common, example was answering "Yes" to "Does *name* share" and answering "We do not share" to "Can you limit this sharing?" in a single row. As shown in Figure 3.3, Bendena State Bank (`bendenastatebank.com`) was among 15 different banks to do so. In a less confusing inconsistency, limiting sharing that does not occur does not make complete sense, yet the Monitor Bank (`monitorbank.com`) and many others answered "No" to "Does *name* share" and answered "Yes" to "Can you limit this sharing?" Other institutions used equally confusing wording to express this concept. For instance, in the "can you limit this sharing?" section of the disclosure table, Merrimac Bank (`merrimacbank.com`) stated "Yes, if we shared." These three kinds of logical inconsistencies can potentially confuse consumers.



Reasons we can share your personal information	Does Bendena State Bank/Bank of Highland share?	Can you limit this sharing?
For our everyday business purposes- such as to process your transactions, maintain your account(s), respond to court orders and legal investigations, or report to credit bureaus	Yes	No
For our marketing purposes- to offer our products and services to you	Yes	We don't share

Figure 3.3: Bendena State Bank was among 15 institutions to state that it shares a particular type of information in one column, yet to state contradictorily "we don't share" in the subsequent column.

Typos and Omissions

While we would argue that logical inconsistencies are a major issue in communicating with consumers, a number of more minor issues cropped up. For instance, we designed our parser to be robust to small differences in wording, such as by ignoring capitalization, considering most punctuation to be optional, and matching either "non-affiliates" or "nonaffiliates." Nevertheless, typos in standardized notices caused many of our parsing "errors." Most of these typos were small, yet caused problems since our regular expressions searched for particular wording. For instance, Bank of Glen Ullin (`bankofglenullin.com`) misspelled "open an account" as "open *and* account." Cape Ann Savings Bank (`capeannsavings.com`) replaced "for our everyday business purposes" with "for *your* everyday business purposes." West Texas State Bank (`ebanktexas.com`) and others used "credit *card* bureaus" in place of "credit bureaus."

Financial institutions also commonly omitted required sections of the model privacy form, again causing problems for our parser. Middlesex Savings Bank (`middlesexbank.com`), for instance, included the "definitions" section, yet left out definitions of the terms "affiliates," "nonaffiliates," and "joint marketing." In many cases, institutions used the model privacy form as their website's privacy policy, replacing the form's headers with the bank's logo and other branding.

Many institutions invented their own wording despite the model specifications [49]. For instance, Fisco (`fisco.com`) said that they collect information when customers "complete subscription documents" and "submit contributions or redemption requests," neither of which was among the 34 standardized terms. Similarly, Monitor Bank (`monitorbank.com`) said it collects "deposit account number(s)," "phone number," "address," "date of birth," and "loan number(s)." While it was not surprising that a financial institution might collect these data, none was listed in the specification [49]. Arguably, however, these institutions' more detailed disclosures might actually be more useful to consumers.

We also observed creative wording in the disclosure table. As a result of our iterative design process, our parser handled most of these variations. For instance, to communicate that one could not limit sharing since the insitution has no affiliates, different institutions wrote each of the following values in the relevant cell of the disclosure table: "*Name* has no affiliates"; "We have no affiliates"; "We don't share"; "We do not share"; "No"; and "N."

Confusingly, institutions sometimes entirely rewrote rows of the disclosure table. City Securities (`citysecurities.com`), for instance, combined three rows of the disclosure table into the single row "For our affiliates' everyday business purposes or for our affiliates to market to you." They also invented a new row for the disclosure table: "For departing Financial Advisors to take limited customer information pursuant to The Broker Protocol*."

Furthermore, institutions commonly ignored the formatting of the model notice and omitted elements. For instance, Hampden Bank (hampdenbank.com), like a handful of others, included most of the information that would be contained in the standard-format disclosure in their privacy policy, yet left out most of the section headers and table formatting. Rather than including a table with the words "Why?...What?...How?" in one column, they created replacement statements like "How do we use the information we collect?" While the semantic meaning is the same, either a human or a computer program would have more trouble comparing institutions' policies, losing some of the benefits of providing privacy notice in a standardized format.

3.5 Discussion

A major advantage of all standardized privacy disclosures is that they enable the direct comparison of companies' privacy practices. In this study, we put this theoretical advantage into action and compared privacy notices of 6,191 financial institutions in the United States, as well as the institutions on consumer-advice lists of 100 largest banks and 11 top credit card companies. Here we discuss the main implications of this work.

3.5.1 Users' Choices

We found differences in data-sharing practices across financial institutions, even within institutions of the same type. Some institutions were more privacy-protective and did not share consumers' personal information for purposes like marketing even when they were permitted to do so. Other institutions did share consumers' personal information, yet allowed consumers to opt out of this data-sharing even when they were not required to offer an opt-out. This suggests that informed consumers would have the opportunity to select institutions with data practices that match their privacy expectations. However, consumers looking for a credit card company would have very limited options since all these companies in our study share data for their own marketing purposes and share data on transactions and experiences with affiliates, without opt-out choices. Most of these companies also share data for joint marketing without opt-outs and more than half share with non-affiliates too.

Overall, our results showed that consumers do have the option to do business with more privacy-protective financial institutions, for some categories of financial services, if they so choose. An important consideration, however, is how consumers might identify institutions with better privacy practices. For small-scale comparisons, the standardized layout of the model privacy form has huge advantages over traditional, non-standardized privacy policies. Because the same information is located in the same place on each standardized notice, consumers can directly compare two or more institutions' privacy practices by placing these institutions' standardized notices next to each other.

While the possibility of consumers choosing financial institutions based in part on privacy practices seems promising, the lack of a simple mechanism for a consumer to make these comparisons on any sort of large scale is unfortunate. For instance, it would be good if a consumer could go to a website and have the ability to say, "I currently bank at Company X. Please tell me about competing banks in the same geographic area that are more privacy-protective." To this end, we built an interactive website (http://cups.cs.cmu.edu/bankprivacy) to help consumers search for or compare financial institutions. The predictable structure of the standardized notices enabled our construction of an automated parser, which was the first step in enabling such an online database.

One can imagine financial institutions with strong privacy practices using privacy practices as a competitive advantage. In past studies, consumers have even paid a premium price to purchase items from companies with more consumer-friendly privacy practices [81], and it stands to reason that they might similarly favor financial institutions with exemplary privacy practices. Both industry and policy makers could benefit from future research investigating consumers' privacy preferences in the financial domain. Results from such research can assist the shaping of companies' practices and mandated requirements.

3.5.2 The Role of Regulators

Our large-scale analysis enabled us to observe how financial regulations impact consumer privacy protections in practice. Many institutions did not provide opt-outs for the three types of data sharing for which they were not required to offer an opt-out. In these three cases, between 158 and 561 institutions provided an opt-out when sharing data, providing consumers choice even when not required to do so. Between 1,816 to 4,507 institutions did not share consumer data at all for each of these three purposes. In contrast, between 1,323 and 3,823 institutions shared data for these purposes without offering an opt-out. This practice is permitted, yet less consumer-friendly.

Limitations of Standardized Notices. We found some issues with the specification of the model privacy form itself. For instance, when specifying what personal information they collect, institutions were mandated to list "Social Security number" and exactly 5 other types of information chosen from a list of 23 possibilities. Similarly, they were required to choose exactly 5 events from a list of 34 possible occasions on which they collect personal information. A glaring issue with these two lists of possibilities is that the types of information and events on the lists were fairly obvious. Consumers probably would not be surprised if their bank collected all 23 types of information on all 34 occasions listed. Indeed, a greater cause for concern might be if, for example, a bank chose *not to* collect a customer's "account balance" when he or she "used his or her credit or debit card." This realization suggests that these particular parts of the model privacy form are not very informative to consumers, who would likely care more about unexpected or non-obvious collection practices.

Short, standardized notices have been suggested as the top layer in a "layered" privacy notice, which has been advocated by both industry groups and regulators [93]. Layered notices bring the most salient information to the forefront of a consumer's attention, yet allow the consumer to obtain additional information easily, such as with a single click. However, the model privacy form has not been designed as a layered notice. The form arbitrarily truncates some categories of information, yet no additional information is made available about an institution's data-collection practices.

This issue is compounded by the manner in which institutions appear to be using the model privacy form. Rather than presenting the model privacy form as a supplement highlighting important points of a full-length privacy policy, the model privacy form replaced full-length policies for many of the institutions we examined. Even though full-length privacy policies are too long for average consumers to read [85], the complete absence of a full-length policy means that institutions do not disclose many of their privacy practices should privacy advocates or other experts choose to inspect them. The specification of the model privacy form [49] notes that "financial institutions may rely on [the model privacy form] as a safe harbor to provide disclosures." It is possible that this safe-harbor provision substantially reduces consumer awareness of privacy practices since institutions are required only to disclose some, rather than all, of their privacy practices on this short-form notice. While we believe the availability of short-form notices to be a good thing for consumers, we also believe that traditional privacy policies should still be made available.

Compliance and Oversight. Standardized notices can also make oversight of privacy disclosures more efficient. Because the standardized notices provided under the Gramm-Leach-Bliley Act are now posted online by many financial institutions we were able to automate the process of collecting and evaluating them. We detected notices with stated sharing practices in apparent violation of United States law. For three of these data-sharing purposes listed in the disclosure table, institutions were required to provide consumers a way to limit sharing [49]. In violation of the law, dozens of institutions said they shared data for these purposes, yet reported that consumers could not limit sharing. When we contacted institutions for which this was the case, some of them explained that the sharing practices they were disclosing annually to their customers were not their actual practices. Although they amended their standardized notices accordingly, these cases make us question to what extend consumers could rely on privacy notices to evaluate companies' actual practices and to what extent stricter regulations and enforcement are necessary. These results also call into question current oversight mechanisms for financial institutions' privacy practices. We then suggest that oversight institutions like the Consumer Financial Protection Bureau (CFPB) could partner with academics with the skills to perform these evaluations at large scale (but with limited economic resources) to assist them with their oversight task.

Incentives to Use Standardized Notices. Given the benefits demonstrated through this work, we believe that regulators should continue incentivizing companies to use standardized notices online. In fact, the CFPB is currently seeking the amendment of GLBA to create such incentives. Companies may be incentivized to use online standardized notices if they can use those notices instead of delivering paper notices. Specifically, if there is an online communication mechanism already established with a customer, the company may not need to deliver a paper notice as long as the customer is provided with a conspicuous link to the online notice. A pointer to the online notice can be provided when monthly statements or other notices are delivered to the customer, either via postal mail or email. If a particular customer does not currently communicate electronically with his or her financial institution, or if the company does not have a website, the company would still be required to provide a paper notice. While it is important to make sure that customers without Internet access have the opportunity to learn about and opt out of sharing practices, requiring all financial institutions with websites to post a standardized notice online would benefit all parties. If the company already has an online presence, adding an online standardized notice does not represent significant additional overhead.

3.5.3 Online Notices and Implementation Issues

In addition to the benefits mentioned above, online notices can be personalized, enable online opt-out methods, and enable links to additional information. For example, users may be able to see a notice that applies to their particular state. We have found that institutions often use the "Other Important Information" section in the model privacy form to include sharing practices exceptions for residents of different states. An online notice can easily provide a drop-down menu allowing customers to select their state of residence to view the applicable privacy notice. Furthermore, an online privacy notice can show whether the opt-out right has already been exercised. We believe that customers' privacy can further be improved if in addition to traditional offline methods such as mail and phone, online opt-out methods are offered. Due to space limitations, the paper-based standardized format does not allow companies to list all the data types that they collect, all the methods that they use to collect information, and the names of the entities with whom they share customers' personal information. In an online notice, this additional and relevant information can be available just one click away from the baseline notice.

Through our large-scale analysis of financial institutions' standardized notices we found that many institutions deviate from the standard model requirements in various ways. For example, some companies use slightly different data types from what is required by the model form to refer to types of personal information that they collect. Some omit information such as the date when the notice was created, or the lists of their affiliates, non-affiliates and joint marketers. We also found inconsistencies in the sharing table entries, including companies listing a "Yes" under the sharing column, but then stating "we don't share," a self-contradiction, under the opt-out column. Also, some companies that claim to offer opt-outs don't offer any specific opt-out method under the "To limit our sharing" section. We believe that many of these problems could be mitigated if a government agency provided an interactive tool that companies could use to generate standardized notices for online posting. The PDF form builder currently available does not prevent these problems. Students at Carnegie Mellon have been developing prototype online form builders that we expect to demonstrate on our website shortly at http://cups.cs.cmu.edu/bankprivacy/.

We faced three problems during our analysis of financial institutions' privacy policies: lack of a comprehensive and publicly available database of financial institutions and their web addresses, lack of a consistent directory path where online standardized notices are located, and lack of consistency in the standardized format. We believe that requiring companies to provide their websites URLs (if they have one) to the CFPB or appropriate authority and making a centralized database with that information publicly available will enable the development of tools such as our bank privacy website. To further facilitate the collection and analysis of online notices at large scale, we suggest that companies should be required to store those notices in well-known and standard location such as INSTITUTIONNAME.COM/notices/privacy/. Finally, an online version of the standard notice could easily include a computer-readable section that would facilitate automated collection, comparison, and analysis.

3.5.4 Study Limitations

The automatic retrieval and parsing of standardized notices allowed us to perform a large-scale analysis of financial institutions' privacy notices, yet introduced some limitations. As we did not have access to the domain names of most of the financial institutions in our original list, we used the conservative heuristics described in Section 3.3 to first find institutions' domain names and then retrieve their corresponding notices if they had one. We were able to retrieve notices from about one third of companies in the original set. We randomly selected 100 companies from the set of those from which we could not automatically retrieve a standardized notice and manually attempted to retrieve domain names and notices from them. We manually found notices from 40 of those 100 companies, suggesting that our heuristics could be improved. However, finding those notices was a time consuming task and required several steps that may not be possible to fully automate. Crowd sourcing could be an alternative, but likely an expensive one as it is time consuming to find notices. A possibility is to use crowd sourcing to find companies' domain names, which is less time consuming, and then use those domain names to automatically attempt to retrieve notices. We also found that small companies (e.g., credit unions) were less likely to have both Internet presence and use standardized notices and that large companies (e.g., BHC) have multiple subsidiaries with different domains that we were unable to find automatically. However, most of these are not consumer facing and tend to have the same privacy policy as the parent company. In sum, our sample of notices may be slightly biased towards larger companies (as they are more likely to use standardized notices) and at the same time, we may have missed very large companies (e.g., BHC) that use different domain names for their subsidiaries. Nevertheless, our sample of notices was heterogeneous enough to allow us to statistically compare financial institutions of different types. Finally, we relied on privacy notices to evaluate and compare companies' practices; however we don't know whether or not those notices accurately reflect real practices. In fact, we have discussed anecdotal evidence that suggest that notices may not always reflect actual practices. Transparency through privacy notices can therefore only be improved if appropriate accountability mechanisms are in place.

Chapter 4

An In-depth Analysis of Online Advertising Companies' Privacy Policies

We analyzed the privacy policies of 75 online tracking companies with the goal of assessing whether they contain information relevant for users to make privacy decisions. We compared privacy policies from large companies, companies that are members of self-regulatory organizations, and non-member companies and found that many of them are silent with regard to important consumer-relevant practices including the collection and use of sensitive information and linkage of tracking data with personally-identifiable information. We evaluated these policies against self-regulatory guidelines and found that many policies are not fully compliant. Furthermore, the overly general requirements established in those guidelines allow companies to have compliant practices without providing transparency to users. Few companies disclose their data retention times or offer users the opportunity to access the information collected about them. The lack of consistent terminology to refer to affiliate and non-affiliate partners, and the mix of practices for first-party and third-party contexts make it challenging for users to clearly assess the risks and make meaningful decisions. We discuss options to improve the transparency and usability of online tracking companies' privacy practices.

4.1 Introduction

Online Behavioral Advertising (OBA) is the practice of tracking Internet users' online activities to deliver ads that are more likely to be relevant to them. While the advertising industry has attempted to self-regulate, Internet users, policy makers, and privacy scholars have raised concerns about the lack of transparency and user control.

In the current self-regulatory regime, OBA companies are directed to publish privacy policies to provide consumer notice and offer opt-out choices [95, 96]. Privacy polices have been shown to be ineffective from a users' perspective [21, 97]; however, they are important for pro-

This chapter is based on "Are They Worth Reading? An In-depth Analysis of Online Advertising Companies' Privacy Policies" [94].

viding transparency. Efforts are being made to interpret policies for users through natural language processing (NLP) tools [98, 99] and crowd sourcing [50]. These efforts will succeed only if privacy policies contain relevant information.

This project analyzed 75 online tracking companies' privacy policies, looking for 59 distinct practices relevant to users. We also gathered data about the proportion of members of ad industry self-regulatory programs and the prevalence of disclosures related to the most consumer-relevant practices and consumer choices.

We found that only 20% of online tracking companies in a public database listed affiliations with the Digital Advertising Alliance (DAA) or the Network Advertising Initiative (NAI), the two predominant advertising self-regulatory organizations in the US. We also found important differences among the evaluated policies, both with respect to disclosed practices and clarity. We identified companies with more privacy-respectful practices as well as companies with more privacy-concerning practices.

Information sharing is unsurprisingly common, but companies tend to conceal their sharing partners' usage of that information. Half of the evaluated companies do not specify their data retention period. Moreover, most companies do not provide options to stop data collection and less than a third provide opportunities to opt out of targeted ads directly in their privacy policies. Most companies do not provide any access to collected information. Further, most companies are unclear or silent about collection and use of non-PII considered sensitive such as income range or health conditions. Large companies and ad industry self-regulatory association members exhibit relatively more comprehensive privacy policies.

We show that the current state of online advertising self-regulation does not provide the level of transparency and control that users demand. In addition to unusable privacy policies, the combination of advertising companies functioning as third-parties (i.e., not user-facing), and the widespread sharing of information among tracking companies creates additional transparency challenges. We conclude by discussing policy and technology options to improve the transparency and usability of online tracking companies' privacy policies.

4.2 Background and related work

We first introduce current practices and concerns related to OBA and efforts to protect users' privacy. We then discuss previous investigations of privacy policies of first-party websites in different domains. Finally, we discuss users' expectations of OBA.

4.2.1 OBA Practices and Self Regulation

In an attempt to make advertising more effective, online advertising companies track Internet users' online activities to show them ads based on their inferred interests. However, the advertising industry has been criticized for targeting ads based on sensitive or personal information [100], discriminating against users [101], or even manipulating users' purchasing intentions [102]. Privacy scholars have argued that the lack of transparency about consumer scores that online tracking companies create can lead to problems of abuse and discrimination as the lack of transparency about credit scores did before the enactment of the Fair Credit Reporting Act [103]. Online tracking companies collect and share users' tracking data in a way that allows data aggregators to create accurate profiles of users' interests and behaviors [104]. Large data aggregators are able to combine interest data with users' personal information and then sell that information to marketers [105]. In March 2013 Facebook announced a partnership with data aggregators to match ads based on users' online and offline behaviors [106] and other offline companies are already tying users' identities with their online activities [107].

The U.S. Government has relied on industry self-regulation with special emphasis on the principles of notice and consent to protect users' privacy [4]. Advertising self-regulatory organizations require members to follow guidelines that include education, transparency, user control, use limitation and security practices [95, 96]. However, as we will discuss in Chapters 6 and 7, currently users are unable to make decisions using transparency and user control tools provided by the ad industry; and member companies do not always comply with self-regulation transparency requirements [17]. Recognizing the existent problems with self-regulation and aiming to protect online privacy beyond OBA, the White House has asked companies to develop enforceable codes of conduct [1] and the Federal Trade Commission (FTC) recommended legislation to provide greater transparency and control over the practices of information brokers [5]. Finally, the California Online Privacy Protection Act of 2003 (CalOPPA) was amended in 2013 to require websites to state how they respond to Do-Not-Track signals. Accordingly, the California's Attorney General has issued a set of recommendations to improve the usability of privacy policies [108].

4.2.2 Evaluation of Privacy Policies

There is a consensus that privacy policies have been ineffective at informing individuals about companies' privacy practices [97]. Cranor argues that privacy policies, and more generally notice and consent mechanisms, are meaningless unless users are empowered with usable and enforceable choice mechanisms [86]. An analysis of the usability of 64 privacy policies from both popular and health-related websites found that both types of websites had policies that suffered from the same usability problems: difficult to access and hard to understand by the average Internet user [83]. Research has also found that the content of health care websites' privacy policies does not match users' needs [109], and that in order to understand those privacy policies users would need reading skills levels that most Americans don't have [110]. A longitudinal evaluation of 312 popular websites found that the average number of words increased and their readability has decreased over time [111].

Research has further assessed the impact of government regulations on the content of privacy policies. An evaluation of health-related organizations' websites before and after the enactment of HIPAA found that transparency of practices increased, but policies became more difficult to understand and users' choices did not improve [112]. Similarly, a longitudinal study of 50 financial institutions' privacy policies found that although privacy policies contained more detailed information about sharing practices after the GLB Act, the amount of sharing among affiliates and non-affiliates increased [68].

In general, users don't like reading privacy policies, they don't understand them [20], and misunderstand their purposes [21]. Furthermore, it has been estimated that if Internet users read website privacy policies it would represent an annual cost of more then $700 billion dollars, which is higher than the cost of accessing the Internet itself [85].

To the best of our knowledge, no previous research has investigated privacy policies from online tracking companies with the level of detail that we present here. Furthermore, our work does not focus on readability of those policies, but their actual content. The reason is because we believe that privacy policies are not and should not be intended for users to read. However, privacy policies are important to provide transparency of practices. These practices can then be extracted by automatic tools or crowd sourcing and presented to users in a more usable manner and in the right context. Therefore, we attempt to provide evidence of the level of transparency of online tracking companies, which will affect the level of success of efforts currently attempting to extract information from these policies [50, 98, 99].

4.2.3 Users' Privacy Expectations

Surveys of Internet users have found high levels of concern about online tracking. Turow et al. found that 87% of telephone survey respondents would not allow advertisers to track them online if given a choice [113]. A more recent Pew telephone survey found that 68% of respondents did not like targeted ads because they didn't like their online behaviors to be tracked [114]. In Chapter 5 we discuss how users are not completely against targeted ads, but they are concerned about the lack of transparency and control that they have over the tracking that enables it. Apart from tracking, transparency, and choice concerns, users have also expressed concerns about the type of targeted ads that they might see, which might lead to embarrassment [115]. In Chapter 8 we show that users relied most on OBA companies' sharing and retention practices to decide what types of information they would disclose for the purpose of receiving targeted ads.

4.3 Methodology

In January 2014 we retrieved a comprehensive list of tracking companies from Evidon's online database.[1] This list had 2,750 companies under various non-mutually exclusive categories including, ad networks, ad servers, ad exchanges, analytics, optimizers, supply-side and demand-side platforms, data management platforms, publishers, among others. It also included the affiliations (if any) that these companies had with self-regulatory organizations. We also obtained a list of the 36 largest tracking companies according to the 2013 Evidon global report [116].

4.3.1 Selection of Companies

We began our analysis with three sets of 36 companies: The 36 largest companies; 36 member companies randomly selected from the set of companies that Evidon reported were affiliated with either of the two largest self-regulatory organizations (Network Advertising Initiative and Digital Advertising Alliance) in January 2014; and 36 companies randomly selected from the set of non-member companies. During the initial analysis process the size of the sets changed. The large set grew from 36 to 37 companies after we realized that one of the large companies, Adobe, had separate privacy policies for its analytics unit and its advertising unit. Therefore we decided to treat these units as separate companies. In addition, we eliminated three companies from the member set that were already included in the large set, thus reducing the size of the member set

[1] `http://www.evidon.com/consumers-privacy/company-database`

Information collected or inferred	**Entities with which info may be shared**	**Retention and Access**
C1: Computers information (e.g., device ID, IP address, OS, cookies, web beacons) C2: Non-sensitive non-PII (e.g., gender, age, non-sensitive interests) C3: Sensitive non-PII (e.g., race, religion, sexual orientation, health conditions, income bracket, credit score) C4: Personally identifiable information (PII) (e.g., name and contact information) C5: Sensitive PII (e.g., financial information, Government ID) C6: Geolocation data (e.g., GPS coordinates or WiFi approximate location	S1: Affiliates S2: Non-affiliates (in general) S3: Non-affiliates (web publishers) S4: Non-affiliates (ad companies) S5: Non-affiliates that can link received information with users' offline activities S6: Non-affiliates that can link received information with users' PII S7: Law enforcement S8: Other non-affiliates	R1: Retention of non-PII R2: Retention of PII A1: Access (e.g., authenticated or anonymous access) A2: Access format (profiles data, raw non-PII, and PII) A3: Access options (e.g., view, edit) A4: Data portability and deletion
Purposes	**Consent Model (Can users limit?)**	**Choice method**
P1: Targeted ads P2: Marketing (e.g., use contact information to offer products) P3: User analytics (e.g., understand how users interact with websites) P4: Ad analytics (e.g., measure performance of ad campaigns) P5: Website customization or optimization P6: Enforcement of terms of services P7: Other uses specified P8: Other uses unspecified	CS1: Use of non-PII for targeted ads CS2: Use of sensitive non-PII for targeted ads CS3: Use of PII for targeted ads CS4: Collection of non-PII CS5: Use of PII for other purposes CS6: Retrospective merging of PII and non-PII CS7: Prospective merging of PII and non-PII CS8: Online and offline information merging CS9: Merging of information across devices	CH1: DAA/NAI Home page link CH2: DAA/NAI Opt-out page link CH3: Opt-out button in policy CH4: Opt-out button elsewhere CH5: Other choice method
Security and other practices	**Contact, Mergers, and Policy Changes**	**Affiliates and Affiliations**
SO1: Mention EU provisions SO2: Mention children provisions SO3: Mask IP Address SO4: Store data encrypted SO5: Mention how tracking works SO6: Mention information sources SO7: Link to educational material SO8: Suggest browser settings	CT1: Contact address CT2: Contact recipient PC1: Policy change notices PC2: Policy update date M1: Mergers/Acquisitions notices and choices	AF1: Define affiliates AF2: Define non-affiliates AF3: DAA/NAI affiliations claimed AF4: Actual NAI/DAI Affiliation

Table 4.1: 59 practices we looked for in online tracking companies' privacy policies.

to 33 companies. Our final set was then comprised of 37 large, 33 member and 36 non-member companies.

In June 2014, after we completed the coding process, we found discrepancies between membership lists on the DAA and NAI websites and the affiliations listed by Evidon in January 2014. In particular, 20 companies with listed affiliations in Evidon's database were not included as members in the DAA or NAI websites. We also found that according to the DAA and NAI websites, 19 of the large companies were members. We decided to consider a company as a member only if it appeared in the DAA or NAI websites and to compare practices of member and non-member companies as well as practices of large and random companies. Therefore, we compared practices of companies in each of the following sets: large companies that were DAA or NAI members, hereafter referred as *large members*, non large companies that were DAA or NAI members, hereafter referred as *random members*, large companies that were not members, hereafter referred as *large non-members*, and random companies that were non-members. hereafter referred as *random non-members*.

In Section 4.4, we focus on comparing practices of members and non-members and we discuss specific differences between large and random companies if those differences exist.

4.3.2 Investigated Practices

We investigated 59 practices pertaining to collection, sharing, use, retention, user consent, access, contact, special provisions for children and European residents, security and user education. We selected these practices based on self-regulatory principles, FTC notice requirements, our knowledge of current practices in which advertising companies engage, as well as users' privacy expectations discussed in the research literature. Table 4.1 shows the specific practices that we attempted to extract from these privacy policies.

4.3.3 Policy Coding

Privacy policies are difficult to read and understand due to the use of legalistic and sometimes ambiguous language. To reduce the number of potential coding inaccuracies, we followed a collaborative and iterative process. There were two stages: development of codes and coding the policies. Three researchers were involved in the first stage and two of them in the second stage. To develop the appropriate set of codes for each evaluated practice, researcher 1 reviewed 10 policies from the set of large companies and proposed a preliminary set of answer choices for each practice. Then, researchers 2 and 3 analyzed the same subset of large companies and applied the proposed codes to extract these companies' practices. Third, the three researchers discussed the preliminary extraction results and identified an improved set of codes. Table B.8 in the Appendix lists the complete set of codes associated with the 59 practices shown in Table 4.1.

Next, researcher 2 coded all the policies. Following the same agreed criteria, researcher 1 coded a subset of 15 policies (20% of each set). We compared the coding of these 15 policies and discussed instances were codes were different. Disagreement occurred due to either factual or interpretation errors. After fixing the factual errors, we conducted an inter-rater reliability test achieving an agreement of at least 80% on each investigated aspect. Then, researcher 1 revisited the rest of the policies to correct similar factual errors.

Interpretation errors happened due to missing or unclear information. For example, if the policy did not mention choices to limit collection of non-PII tracking data, one researcher would select "User cannot limit this practice" (see Table B.8 in the Appendix), while another researcher would select "The policy doesn't mention this." We revised our coding criteria for user consent practices and decided to use "The policy doesn't mention this" unless it was explicitly stated in the policy that the user cannot limit the practice. Similarly, one researcher would select "Information is collected" if it was either explicitly mentioned or could be inferred that the company was collecting a given data type, while the other researcher would select "Information is inferred." We revised our coding criteria for collection practices and decided to reduce the granularity of the codes by grouping "Information is inferred," "Information is collected," and "Information is collected and inferred" in Table B.8 in the Appendix as "Information is collected." We further grouped "Unclear" and "Policy does not mention" choices as "Don't mention." After specifying the new coding criteria we achieved full coding agreement for the subset of 15 coded policies. Researcher 1 then revisited the rest of the policies and applied the new criteria.

4.3.4 Policy Retrieval

Evidon's database included a URL that was supposed to link to each company's privacy policy. However, sometimes Evidon's links did not take us to the company's privacy policy. For example, sometimes Evidon's links pointed to the company's home page when Evidon had determined that the company did not have a policy, while other times the links took us to nonexistent web pages. When the URL did not link to a company's privacy policy, we visited that company's home page and looked for the privacy policy link (usually found at the bottom of the page). On most occasions, when Evidon's link was not functional we found that the company did not have a privacy policy. The exceptions were when the company had changed its name, or was merged with or acquired by another company. In those few cases, we used the Google search engine to determine the name of the new company and find its website and then its privacy policy if it existed. Some of the companies' privacy policies, mainly from the large category, included several links to other related pages. When that happened, we followed all available links to try to extract the practices of interest.

4.3.5 Limitations

The results we present in the next section offer a somewhat representative snapshot of OBA privacy policies in the winter of 2014. We tried to ensure a diverse set of companies by selecting both large companies and a sampling of random companies. Due to discrepancies between the information from Evidon and from the self-regulatory organizations that we we were unaware of until after we completed coding the policies, we had to regroup our samples after we coded them. Thus our two random groups represent a mix of the two original random samples, and not a random sampling of the non-member and member groups.

While we observed that OBA companies do not change their privacy policies frequently, it is likely that a small number of companies changed their policies over the period of several weeks during which our coding took place, and more may have changed their policies since then.

Finally, while we attempted to code the policies as objectively as possible, privacy policies are often ambiguous, silent, and difficult to understand. Therefore, the codes selected for some of the stated practices are subject to researchers' interpretation.

4.4 Results

We attempted to analyze privacy polices from 106 online tracking companies. As shown in Table 4.2, we found that many non-member companies either did not have an online privacy policy, had a privacy policy that was not intended for tracked Internet users, or had websites written in a language other than English. Only 84 of the 106 companies we examined had a privacy policy written in English, and only 75 of those had a privacy policy that included relevant content for tracked users. Furthermore, the lack of privacy policies was more salient among random non-member companies, but there were also large non-member companies that did not have privacy policies written in English with relevant content for tracked users.

	Members		Non-members		
	Large (#, % of sample)	Random (#, % of sample)	Large (#, % of sample)	Random (#, % of sample)	Total
Initial sample size	19	10	18	59	106
Have an English-language privacy policy	19 (100%)	10 (100%)	16 (89%)	39 (66%)	84
Have an English-language and tracked user privacy policy	19 (100%)	10 (100%)	14 (78%)	32 (54%)	75

Table 4.2: Tracked user privacy policies written in English. All member companies have English-written policies with relevant content for tracked users, but many randomly selected non-member companies do not have user-relevant privacy policies.

There were important differences among the evaluated policies both with respect to disclosed practices and clarity.

We organize the remaining results as follows. First, we report self-regulation affiliation rates. Second, we discuss important practices that are not disclosed or unclear. Third, we present stated practices that we consider problematic as well as those that we deem more privacy respectful. Finally, we discuss hurdles that make privacy policies of OBA companies challenging to understand.

4.4.1 Low Self-Regulation Adoption

Only a small fraction (30%) of tracking companies in Evidon's online database listed affiliations with self-regulatory organizations, and a smaller fraction (20%) listed affiliations with any of the major self-regulatory organizations in the US. Furthermore, only 18 (49%) of 37 large companies in our sample were DAA or NAI members and only 10 (14%) of 69 randomly selected companies were DAA or NAI members.

Regardless of whether the company was listed as member in either the DAA or NAI websites, we looked for any mention of affiliations with self-regulatory organizations made in the privacy policies themselves. Table B.17 in the Appendix shows which companies claimed affiliations with any self-regulatory organization. All member companies included statements regarding their affiliations with self-regulatory organizations; however, we also found that one non-member company (`sojern.com`) claimed affiliations with self-regulatory organizations, but was not listed as member on either the DAA or NAI websites. We emailed the DAA on June 24 and June 30 of 2014 informing them about this situation, but we did not get any response.

4.4.2 Silent and Unclear Practices

In this section, we show that non-member companies were less transparent than member companies across all practices; however, a large fraction of member companies were also silent with respect to important practices including, data collection, sharing, purpose of use, retention, and user consent.

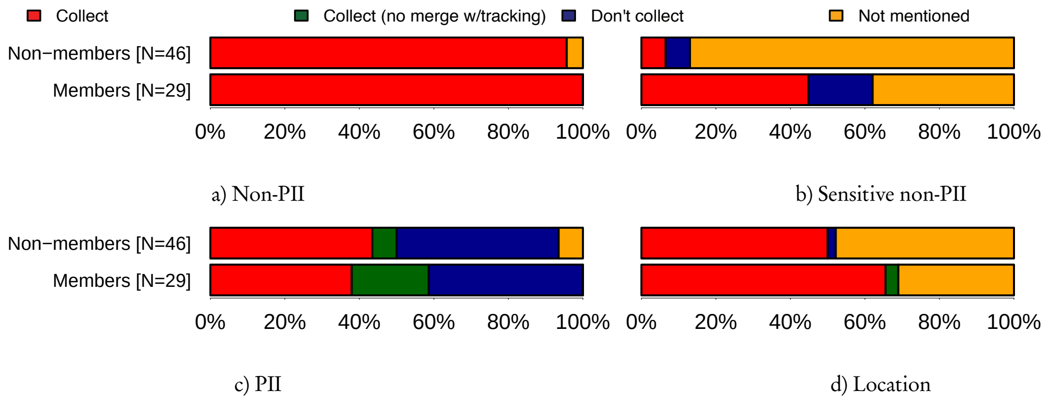
Collection

While most companies do not explicitly mention the collection of non-PII such as anonymous demographic or interest data, most of them mention the logging of page visits or inferring users' interests. Therefore, whenever a company mentioned anything related to logging page views or making inferences about users' interests, we coded that as collection of non-sensitive non-PII. Unsurprisingly, Figure 4.1a shows that most of the companies state they collect non-PII. In fact, as shown in Table B.9 in the Appendix, only two non-members (one large and one random) did not mention the collection of non-PII.

However, Figure 4.1b shows that, a very large fraction (87%) of non-member companies and more than a third (38%) of member companies don't explicitly disclose whether or not they collect sensitive non-PII ($p < 0.001$, Fisher's exact test).

While we could have assumed that the lack of disclosure meant "no collection," we decided to differentiate between those companies that explicitly state they do not collect such information and those that are silent about it. Making a clear statement about the collection of sensitive non-PII is particularly important as we will show in Chapter 8 that users are not comfortable disclosing sensitive information such as health or income related information, and many companies do not exhaustively list the information they collect, commonly stating that collection is "not limited to" a given list of data types.

As shown in Figure 4.1d, many of the companies were also silent about the collection of geolocation data, where a large fraction of both non-member (48%) and member (31%) companies did not include any statements regarding collection of this data type.



a) Non-PII

b) Sensitive non-PII

c) PII

d) Location

Figure 4.1: Collection Practices

Sharing

Sharing practices are particularly important because an uncontrolled transfer of information could lead to unclear, if not unintended, uses against users' expectations. We investigated sharing practices with both affiliates and non-affiliates. We considered as affiliates those companies under the same ownership, or those companies that receive information to provide a service to the company under analysis and that are contractually obliged to only use such information to provide the requested service. Here we discuss non-affiliate sharing. As shown in Figure 4.2a, most of the companies share only non-PII with non-affiliates. However, a considerable fraction of companies (17.3%) are silent about non-affiliate sharing.

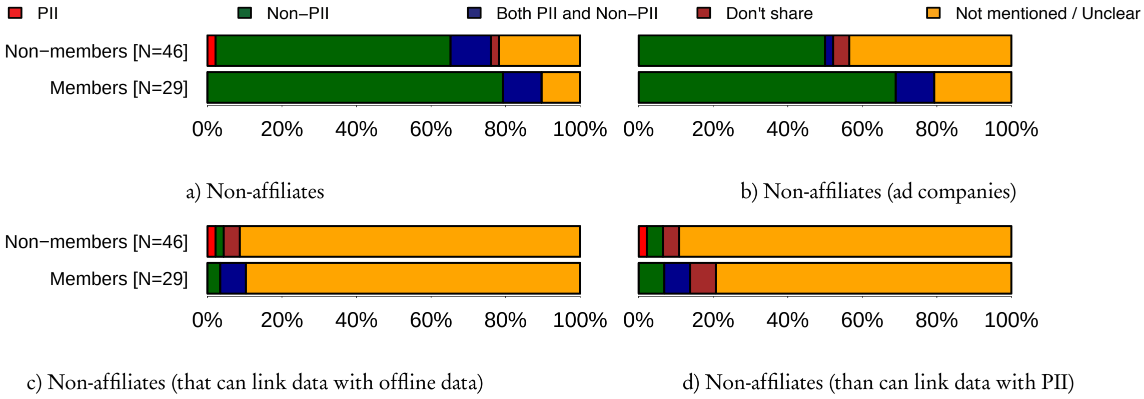
We further investigated whether companies disclose more specifically with whom they share. Unsurprisingly, as shown in Figure 4.2, companies were more silent as we looked into more specific types of sharing. Specifically, Figure 4.2b shows that non-member (44%) are more silent than members (21%) about sharing with other ad companies ($p = 0.05$, Fisher's exact test).

Particularly important is the sharing with non-affiliates that can link received data with users' offline behavior or otherwise with PII. However, as shown in Figures 4.2c and 4.2d, both member and non-member companies are silent about these situations. Again, we could have assumed that the silence regarding these practices meant that it does not happen. Nevertheless, merging tracking data with PII and offline data is not an uncommon practice. Data brokers, which are often recipients of information sold by online tracking companies, often merge individuals' PII with their interest data collected via other methods. In addition, companies do not assume responsibility for non-affiliate recipients' practices. Therefore, we considered it important for companies todisclose explicitly whether they share information under these circumstances.

The NAI code of conduct and DAA self-regulatory principles require member companies to provide a notice indicating how collected data will be used, "including transfer, if any, to a third party." This generic notice requirement makes it easy for companies to be compliant, however, it does not allow users to assess the risk of those data transfers.

Furthermore, while the NAI requires members who transfer non-PII to non-affiliates to require those recipients to "not attempt to merge such non-PII with PII" unless the user opts in [95], opt-in methods are also usually unclear and often users who voluntarily provide PII to other third-parties (usually in a different context) are implicitly opting it for such merging. Interestingly, the DAA principles also have a similar transfer limitation requirement, but that requirement only applies to service providers, not third-party trackers [96].

Finally, the NAI code of conduct only requires companies to offer an opt-out choice if they want to merge non-PII collected in the future (as opposed to previously) with PII [95].



a) Non-affiliates

b) Non-affiliates (ad companies)

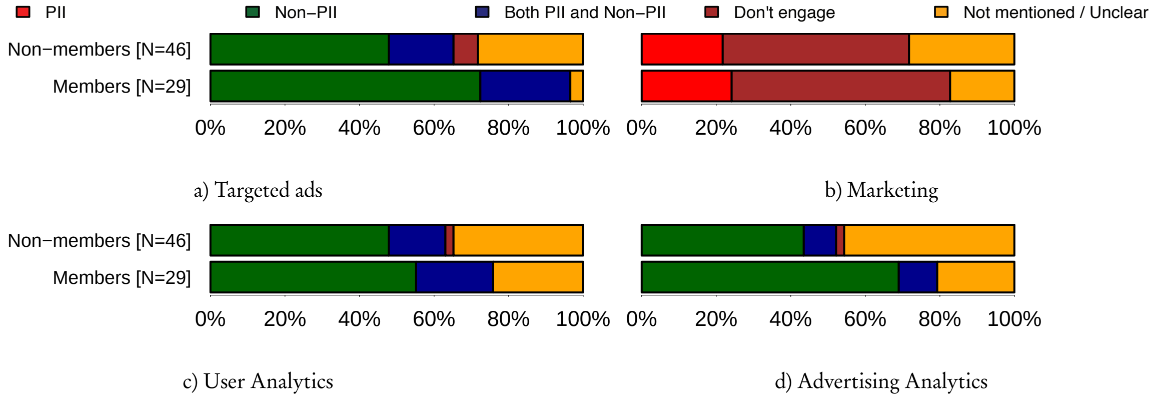
c) Non-affiliates (that can link data with offline data)

d) Non-affiliates (than can link data with PII)

Figure 4.2: Sharing Practices

Purposes

We attempted to extract statements related to various use practices including, ad targeting, marketing, user and ad analytics, website customization, enforcement of terms, and "other purposes." Here we limit our discussion to the former four. Table B.11 in the Appendix shows detailed use practices for each company. The types of information used for targeted ads are shown in Figure 4.3a.



a) Targeted ads

b) Marketing

c) User Analytics

d) Advertising Analytics

Figure 4.3: Purposes

Most companies (81%) explicitly state that they use either non-PII or both non-PII and PII for targeted advertising; however, there are "analytics" providers, "ad servers," and other ad related companies, which are not explicit about their engagement (or lack of) in targeted ads. Specifically, Figure 4.3a shows that non-member companies (28.3%) are more silent than member companies (3.4%) about this practice ($p = 0.006$, Fisher's exact test).

While we could have assumed that analytics providers would not engage in targeted ads and ad servers would, we found a handful of analytics companies that state that they engage in targeted ads and some ad servers that were silent about the practice. For example, Table B.11 in the Appendix shows that three non-member companies (`userreport.com`, `foreseeresults.com` and `twelvefold.com`), explicitly state that they do not engage in targeted ads. The former two are classified in Evidon's database as analytics providers, hence it is not surprising that they do not engage in targeted ads. However, `twelvefold.com` is categorized as ad server in addition to analytics provider, yet it states that it does not engage in delivering targeted ads. Furthermore, there were other companies categorized as analytics providers that state they engage in delivering targeted ads (e.g., `whos.amung.us`, `advanseads.com`). Therefore the categorization of a company cannot be used to infer its data use practices when the company does not explicitly state those practices.

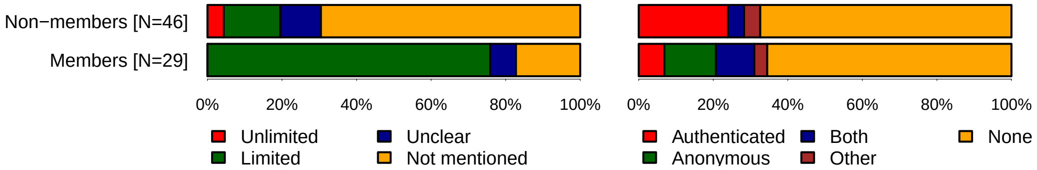
Figure 4.3b shows marketing (e.g., use of contact information for marketing purposes practices.) More than half (53%) of companies do not engage in marketing practices and (23%) explicitly state that they perform marketing. However, a considerable fraction of member (17%) and non-member (28%) companies who collect PII do not disclose whether or not they use this information for direct marketing purposes.

"User analytics" is defined as the practice of analyzing users' actions on first party websites and "ad analytics" is defined as the practice of evaluating the performance of advertisement everywhere they are shown. Both of these are common practices among online tracking companies; however, as shown in Figures 4.3c and 4.3d, a large fraction of companies do not disclose whether or not they engage in these practices.

Retention and Access

Both the DAA and NAI allow retention "as long as necessary to fulfill a legitimate business need, or as required by law" [95]. We found that many companies use similar language to obscure their retention periods. While it is reasonable that companies need to keep information to fulfill their business needs, this vague requirement should not prevent them from establishing a retention period. We are also unaware of any laws that require these companies to keep tracking data and believe that adding the phrase "as required by law" in this context is misleading. Figure 4.4a shows that a large faction of non-member companies (80%) and a smaller fraction of member companies (24%) do not disclose the retention period of collected non-PII ($p < 0.001$, Fisher's exact test).

Figure 4.4b shows that many companies (68%) do not mention any opportunity for users to access information they collect about or infer from users' online activities. Only a quarter (24%) of member and a small fraction (4%) of non-member companies offer "anonymous" or both "anonymous" and "authenticated" access. Therefore, in general very few companies provide access to this information. Table B.12 in the Appendix shows detailed retention and access practices of each company.



a) Retention period for Non-PII

b) Access

Figure 4.4: Retention and Access Practices

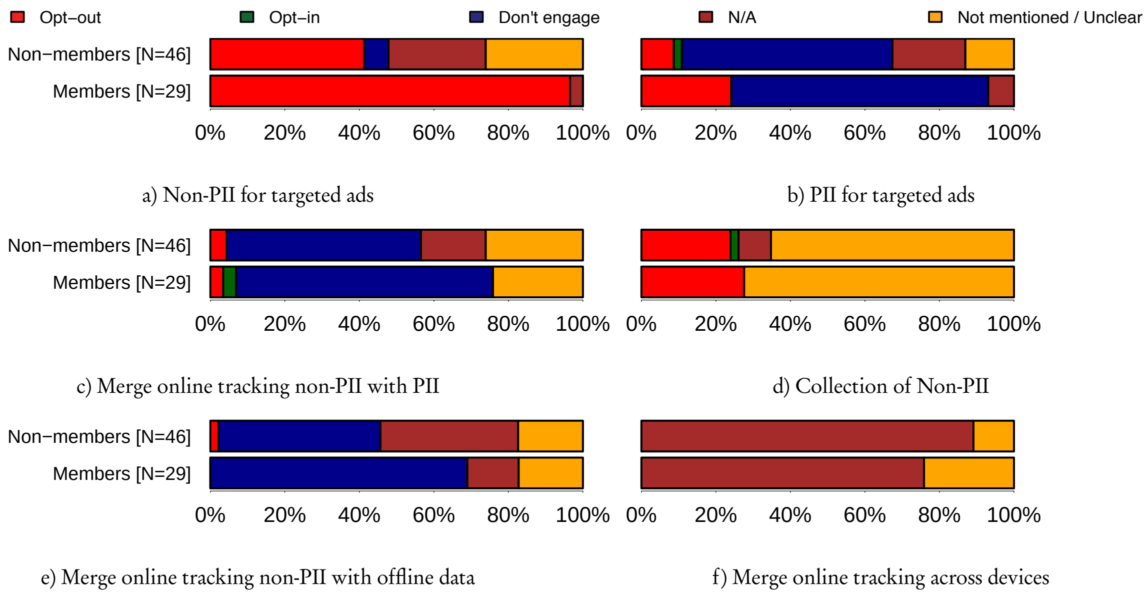
Consent Mechanisms

We investigated consent mechanisms to both determine the extent to which companies comply with NAI and DAA requirements and assess the salience of the choices offered. The NAI code of conduct establishes various user consent practices. It requires collection of users' opt-in consent before 1) merging PII with previously collected non-PII, a practice the NAI calls "retrospective merger," 2) use of precise geolocation data for targeted ads, and 3) use of sensitive data for targeted ads. It further requires offering of opt-out choices for collection of information for targeted ads (but not collection for other purposes) [95]. The DAA establish more lax consent requirements as it only requires companies to offer the opportunity to opt out of collection and use of data for targeted ads (but not collection for other purposes) [96].

Many companies offer opportunities to opt out of targeted ads (see Figure 4.5a), however the opportunities to stop the collection of information for other purposes are often not mentioned (see Figure 4.5c). Also, while most companies do not engage in merging non-PII with PII (59%)

or with off-line (53%) data, the majority that can engage do not specify consent options for any of those practices (see Figures 4.5d and 4.5e). Specifically, a quarter of member (24%) and non-member (26%) companies do not mention any choices to limit merging of PII and non-PII, although their polices suggest that such merging is possible.

Furthermore, Figure 4.5f shows that none of the companies that mention tracking across devices offer any options for users that limit it. Overall, while many companies offer opt-out choices for targeted ads, only very few offer choices for data collection, and almost none offer explicit choices to prevent merging of PII with non-PII.



a) Non-PII for targeted ads

b) PII for targeted ads

c) Merge online tracking non-PII with PII

d) Collection of Non-PII

e) Merge online tracking non-PII with offline data

f) Merge online tracking across devices

Figure 4.5: User Consent Practices. "N/A" denotes many companies that were not clear or explicit about engaging in the given practice and hence they don't offer related choice options. "Don't engage" denotes companies that explicitly stated that they don't do the given practice.

4.4.3 Disclosed Practices

There were several companies with more transparent and explicit practices. We first discuss companies with more privacy-respectful practices and then those with more questionable practices.

Privacy-friendly practices

Five (17%) member and three (7%) non-member companies explicitly mention that they do not collect sensitive non-PII (see Table B.9 in the Appendix for details). Furthermore, a large fraction of both member (41%) and non-member (44%) companies state that they do not collect information that personally identifies users.

Remarkably, one random member (`rocketfuel.com`), one large member (`adadvisor.net`), and two random non-member (`foreseeresults.com`, `visbrands.com`) companies explicitly state that they do not share with entities that can link received data with PII. Moreover, the two non-member companies also state that they do not share with entities that can link received data with offline data.

A handful of both member and non-member companies state specific and limited retention periods for tracking data, which range from 20 days to 2 years.

In addition, while many companies only offered the opportunity to opt-out of targeted ads, but not the opportunity to opt out of being tracked, we found 8 (24%) members and 11 (28%) non-member companies (see Table B.13 in the Appendix) using language that suggests that users can actually limit online tracking when they opt out.

Finally, as shown in Table B.17 in the Appendix, one large member and two large non-members indicate that they take measures to anonymize IP addresses. The large member (`quantcast.com`) indicates, "we do not store full IP addresses." One non-member (`histats.com`) states, "In order to ensure better privacy protection, Histats anonymize all IP addresses: the last three digits of the IPv4 are deleted immediately, and last 64 bits on IPv6." The second non-member company (`gemius.com`) refers to location information as "geographic location on the basis of anonymized IP address."

Privacy-concerning practices

A large fraction of members (45%) and a small fraction of non-members (6.5%) collect or infer sensitive non-PII ($p < 0.001$, Fisher's exact test). Similarly, a large fraction of both members (38%) and non-members (44%) collect PII without mentioning any use restrictions, and both member and non-member companies were silent about user choices to limit merging of non-PII with PII.

Moreover, a small fraction of member (10%) and non-member (13%) companies share PII or both PII and non-PII with non affiliates. Similarly, a small fraction of both member (14%) and non-member (7%) companies also state that they can share with non-affiliate companies that can link non-PII with PII.

While many companies do not disclose or are unclear about their retention period for online tracking data, one large non-member (`optimizely.com`) discloses unlimited retention period. It states that "Non-personally identifiable information may be stored indefinitely."

4.4.4 Opt-Out implementation

All member companies that engage in targeted ads offer opt-outs and, interestingly, a large fraction (41%) of non-member companies also claim to offer the opportunity to opt out of targeted ads.

The most popular opt-out methods among member companies are either a link to the DAA/NAI opt-out pages (76%) or DAA/NAI home pages (79%). Surprisingly, we found that a considerable fraction of non-member companies also include links to the DAA/NAI opt-out pages (22%) or DAA/NAI home pages (6.5%), even though those pages are only useful for opting out of targeted ads from members.

A large fraction of member companies (52%) compared with non-member (20%) companies use opt-out pages, where companies explain with somewhat more detail how targeted ads work, and provide an opt-out button as well as links to the DAA and NAI websites. Less than half of member companies (41%) and a relatively small fraction of non-member companies (17%) include an opt-out button directly in the privacy policy.

As shown in Figure 4.6, other choice methods include the opportunity to access and edit anonymous profiles (e.g., `bluekai.com/registry`), edit personal profiles (`adobe.com/`), opt out from participating in research surveys (`voicefive.com`), opt out from other companies (`optimizely.com`), establish preferences to receive text alerts for ads based on location (`att.com`), adjust account settings (`digg.com`), among many other specify ones. Overall, we found that many companies offer opt-out choices for targeted ads and marketing communications. However, user choices for other purposes such as collection of tracking data, merging of tracking data with PII, or tracking across devices are rather limited.

4.4.5 Other disclosures

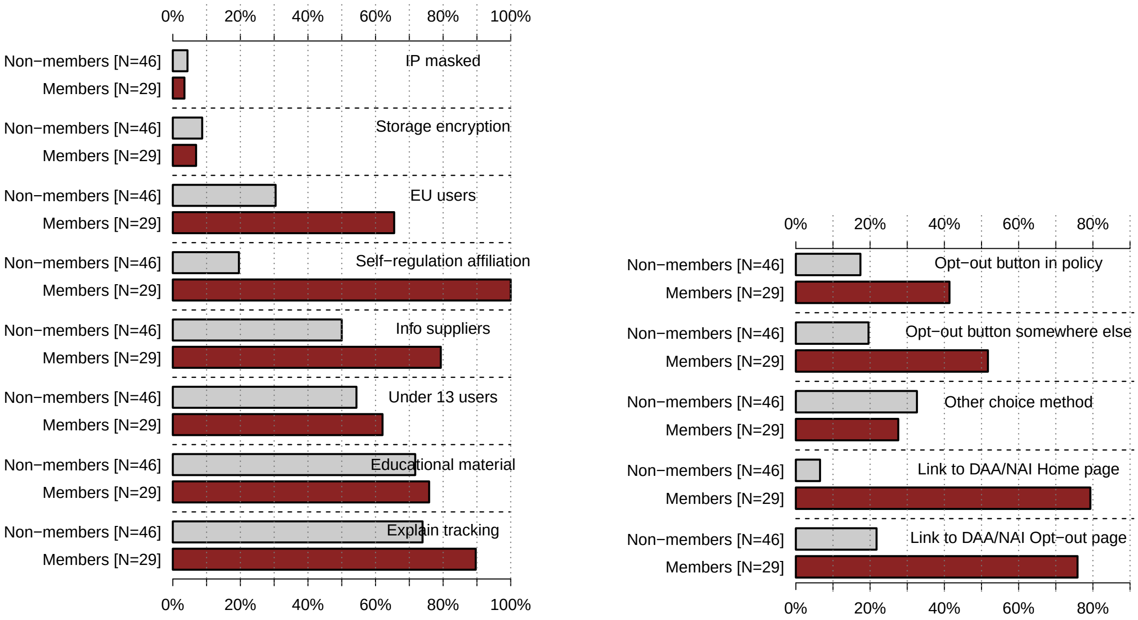
We investigated several other types of disclosures made in OBA privacy policies, including educational material, companies' contact information, policy changes and mergers/acquisitions notifications, whether or not special provisions for European residents and children are mentioned, as well as data security practices. Tables B.15 through B.17 in the Appendix show the details for each company.

Educational material

Both the NAI and DAA establish requirements to educate users. A large fraction of companies refer to cookies, web beacons, tags, pixels, or "pieces of code" to describe how they track users' online activities. However, describing how tracking works is arguably not very educational as users often do not understand the technology jargon used to describe it. Therefore, we searched for other educational material (or pointers to it) in the privacy policy. Figure 4.6a shows the fraction of companies making statements to describe online tracking and providing educational statements or links. We found two main types of educational material: suggestions to configure cookie browsing settings and pointers to the website `http://www.allaboutcookies.org/`. A few companies also provided a link to the DAA consumers' page `http://www.aboutads.info/consumers`. However, neither of these two websites provide useful recommendations to protect online privacy, but mostly talk about the benefits of cookies and online advertising. A large fraction of both member (76%) and non-member (72%) companies include these kinds of educational material in their privacy policies.

Information providers

The NAI requires companies to be diligent about receiving data for OBA purposes "from reliable sources that provide users with appropriate levels of notice and choice" [95]. Nevertheless, we found that while 79% of member companies mention that they receive information from third-parties, they do not indicate that those sources provide "appropriate levels of notice and choice," being reliable or otherwise accountable for handling user information responsibly. Examples of statements used include, "at times may also use Non-PII data from third parties," or "we may combine Non-Personal Information with data collected from other sources." Notably, the remaining 21% of member companies do not even mention whether or not they receive information from other entities.



a) Other practices

b) Opt-out implementation

Figure 4.6: Opt-out implementation and other practices. Only 63% of large and (53%) of member companies mention affiliations to the DAA or NAI in their privacy policies. Only a small fraction of large (33%), member (23%) and non-member (17%) companies provide an opt-out button directly in their policies. Only 63% of large and (53%) of member companies mention affiliations to the DAA or NAI in their privacy policies.

Europeans and children's provisions

We also looked at whether privacy policies included any particular statements for children or Europeans. As shown in Figure 4.6, a large fraction of member (65%) and a smaller fraction of non-member (30%) companies include statements for Europeans. These statements were shown more often when the company collected PII and they usually cited the US-EU and US-Swiss Safe Harbor Frameworks. Some companies also cited European regulations or European self-regulation organizations such as `youronlinechoices.com/uk`. Similarly, more than half of member (62%) and non-member (54%) companies include statements regarding children under 13. However, we did not find any company mentioning the self-regulatory program for children's advertising [117].

Self-regulation affiliation claims

All member and 20% of non-member companies mentioned their affiliations with self-regulatory organizations. However, not all of these mention affiliations to the NAI or DAA. In particular, one large member (`facebook.com`), one large non-member (`disqus.com`), and four random non-member (`tapjoy.com`, `apple.com`, `att.com`, and `verizon.com`) companies mention affiliations with TRUSTe. Furthermore, one large non-member (`gemius.com`) and one random non-member (userreport.com) companies mention adherence to ESOMAR (esomar.org), an European organization.

Security provisions

We found that most of the companies include boilerplate security statements, which we did not code. Instead, we looked at whether the companies stated that they encrypted the collected data. Notably, one large member named Neustar (adadvisor.net) states that "the contents of AdAdvisor Cookies are encrypted, and can't be read without the encryption key." We also found that one large member (tapjoy.com) and one random member (addthis.com) use exactly the same sentence to indicate that they use encryption, "We take reasonable security measures to protect against unauthorized access to or unauthorized alteration, disclosure or destruction of data. These include firewalls and encryption." Other companies also mention encryption, but were not specific about which data was encrypted, for example a random non-member company named SET Media (www.set.tv) mentions, "to maintain the security of its network and the data we collect. We use various technologies, including, in certain instances, encryption."

Policy changes and updates

We found that a large fraction of companies do not include a statement explaining how users will be informed if the privacy policies changed. Many non-member (44%) and member (24%) companies do not provide policy-change notifications to users ($p = 0.05$, Fisher's exact test). However, there were also companies (41%) across both sets that explicitly state that a notice would be provided in the policy when it changed. Some of the companies who collect contact information further indicate that they would both provide a notice in the policy and email customers if their policies changed.

Mergers and Acquisitions

During our evaluation period, we noticed that mergers and acquisitions among tracking companies are common. Notably, one large member company (bluekai.com) was acquired by Oracle, and a few other small companies were merged with larger or other small companies. Therefore, we looked into provisions related to how users would be informed and what options would be offered to them in case of mergers or acquisitions. Unsurprisingly, given our previous results, many companies (28%) across both sets were silent about this practice. Furthermore, a large fraction of companies (61.3%) across both sets mention that they may share users' information in case of mergers, yet do not mention any notification for users or any user choices. However, we also found four member (14%) and two non-member (4%) companies mentioning that some form of notice would be provided, two of them (one member and one non-member) indicating that users would be able to opt out of the sharing of their personal information.

4.4.6 Understandability Hurdles

Here we discuss identified aspects that make these privacy policies difficult to understand and act upon.

Mixed Practices

Online tracking companies normally have many "partners," which may include advertisers, publishers, other advertising or tracking companies, etc. We found that often privacy policies are unclear about who the intended audience is, often mixing practices that apply to their partners, their websites' visitors, and tracked Internet users. In very rare cases privacy policies are designed to exclusively inform tracked users and more often policies include paragraphs or sentences that could apply to both partners and tracked users, making it very difficult to disentangle the practices that apply exclusively to tracked users.

Among both member and non-member companies we observed several companies that are both service providers in first-party contexts as well as online tracking companies. These include both large (e.g., Adobe, Verizon, CBS, etc.) and smaller (e.g., Tapjoy, WildTangent Games, Traffiq, etc.) companies. Although large companies are clear about some of the different practices that apply to direct customers and general audience of tracked users, smaller companies are often less clear. There are often situations were it is impossible to determine whether a given practice applies to direct customers, tracked users, or both. A typical example of this situation is when a company collects personal information from a first-party relationship as well as tracking data. In this case, many companies are not explicit about linking or not tracking data with personal information. The situation is worse with other practices such as uses, sharing, access, and retention period, where it is often impossible to differentiate between practices that apply to information collected in first-party and third-party contexts.

Terminology

Given that sharing practices are common among advertising companies, we investigated how these companies define the affiliates and non-affiliates with whom users' information is shared. Many companies do not mention affiliates or non-affiliates, and those who do mention them, do not provide a clear definition, mentioning them vaguely. For example, privacy policies include sentences like, "may use or share the information we collect with our affiliates and third parties, such as our service providers, data processors, business partners and other third parties," "may share with advertisers and their service providers and partners," "may share with interested third parties," or "may use or share the information we collect with our affiliates and third parties, such as our service providers, data processors, business partners and other third parties," "may share with our partners like publishers, advertisers or connected sites."

While it is understandable that tracking companies may have different partnerships, from a users' perspective, it is very difficult to accurately determine which of those may or may not follow the same practices as the company under scrutiny. A consistent definition of affiliates and non-affiliates that tracking companies can use to refer to companies that follow or not their same practices would help users to better understand sharing and other practices and then be in a better position to assess the associated risks.

Companies also have different definitions of sensitive data. While for some companies income bracket is considered sensitive, for many others it is not. Similarly, for some companies over-the-counter medications are not sensitive data while others do not specify whether or not such data is sensitive. Also, geo-location is considered sensitive information by a small number, but not by many others. Without a clear definition of what constitutes sensitive data as well as a clear separation between sensitive and non-sensitive tracking data, Internet users cannot be certain whether advertising companies' practices infringe their privacy.

4.5 Discussion

OBA self-regulation is not providing effective privacy protections. Participation in self-regulation is voluntary and we found that only 20% of 2,750 companies in a public database of online tracking companies listed affiliations with the DAA or NAI, the two main online advertising self-regulation organization in the U.S. The discrepancies between affiliations included in Evidon's database as of January of 2014 and members listed in the DAA and NAI websites as of June 2014 suggest that membership may be dynamic and companies might join and leave at will. Interestingly, we also found that a handful of non-member companies suggested that users could opt out from OBA by visiting the DAA or NAI opt-out pages, which offer opt-outs only from their members.

We also found that the NAI code of conduct and DAA self-regulatory principles allow member companies to be compliant without offering significantly better protections than non-member companies. The NAI limited definition of sensitive data allows member companies to collect or infer information that research has shown users are not willing to share with online advertisers. Also, while member companies are more likely to have a privacy policy, both member and non-member companies have privacy policies that are silent about practices that impact users' privacy.

The DAA and NAI limitations for sharing with third-parties and merging PII and non-PII are not protective. Tracking companies that collect PII in first-party contexts can freely merge it with tracking data. Member companies who share with third parties are not required to mention the purpose of sharing. The end result is that information about users' online activities is often freely shared and such information can be linked with PII.

4.5.1 Improving notices for users

Transparency and usable users' choices are necessary for a self-regulated market to function. However, we have found that online tracking companies are not transparent and do not offer meaningful choices to users. User consent is often implied when the user visits a website with tracking. The NAI code of conduct requires companies to collect opt-in consent before using sensitive data or location for targeted ads, but it is unclear how to obtain opt-in consent in third-party contexts. The third-party nature of tracking in combination with the lack of transparency makes user consent meaningless.

Efforts are being made to use natural language processing (NLP) techniques to interpret privacy policies [98, 99]; however, if the problems identified here are not fixed, those efforts will be fruitless. For example, if companies are silent or have mixed practices, neither humans nor automatic algorithms will be able to make good use of them. We have compiled a list of 59 aspects that online tracking companies could use as a guide to assess the content of their privacy policies.

We found that many companies have more privacy-respectful practices; however, the current status of notices don't allow them to stand out from less protective companies or enable users to use that information to make privacy choices. We believe that finding ways to standardize terminology and structure of policies will benefit both users and those companies with more privacy-respectful practices.

We identified several factors that make online tracking companies' privacy policies very hard to evaluate and understand. The lack of, affiliates and non-affiliates definitions, agreement about sensitive and non-sensitive data, clarity about practices that apply for information collected in first- and third-party contexts, and clarity about the merging of non-PII with PII, makes it challenging to differentiate what kinds of information are shared with whom and assess privacy risks for users. Including a policy section that consistently defines affiliates and non-affiliates, collected or inferred data types, and data uses can improve these policies. We then could imagine a tabular section similar to a privacy nutrition label [20] that summarizes privacy policies in a more understandable manner. The fact that OBA practices have become complex should not mean that Internet users should bear the burden of this complexity.

4.6 Conclusion

We used Evidon's public list of 2,750 online tracking companies and Evidon's 2013 global report to draw a sample of 106 of these companies, including large companies, companies that are members of self-regulatory organizations, and non-member companies. Only 75 of these companies had English-language privacy policies with content relevant for tracked users, which we analyzed thoroughly. We found that most of these companies are silent with regard to important consumer-relevant practices including the collection and use of sensitive information and linkage of tracking data with personally-identifiable information. Policies lacked a clear and consistent definition of non-affiliates with whom online tracking companies share user information. Policies also mixed practices that apply to information collected in first- and third-party context, and they are rarely intended only for tracked users, but more often intended for different audiences simultaneously (e.g., partners, website visitors, and tracked users). These facts would make it very difficult and sometimes impossible for users to determine what practices apply to them and be able to properly assess the associated privacy risks. Unless these problems are fixed, ongoing efforts to use natural language processing (NLP) techniques and crowd sourcing to interpret privacy policies will not be able to improve transparency and empower users to protect their privacy in the context of OBA.

We also evaluated these policies against self-regulatory guidelines and found that many policies are not fully compliant. Furthermore, while member companies are more likely to offer the opportunity to opt out of targeted ads, previous research has shown that users are concerned about online tracking and interested in controlling data collection, an option that companies are not offering. We have provided recommendations to improve clarity and usability of online tracking companies' privacy policies.

Chapter 5

Perceptions of Online Behavioral Advertising

We report results of 48 semi-structured interviews about online behavioral advertising (OBA). We investigated non-technical users' attitudes about and understanding of OBA, using participants' expectations and beliefs to explain their attitudes. Participants found OBA to be simultaneously useful and privacy invasive. They were surprised to learn that browsing history is currently used to tailor advertisements, yet they were aware of contextual targeting.

Our results identify mismatches between participants' mental models and current approaches for providing users with notice and choice about OBA. Participants misinterpreted icons intended to notify them about behavioral targeting and expected that they could turn to their browser or antivirus software to control OBA. Participants had strong concerns about data collection, and the majority of participants believed that advertisers collect personally identifiable information. They also misunderstood the role of advertising networks, basing their opinions of an advertising network on that company's non-advertising activities. Participants' attitudes towards OBA were complex and context-dependent. While many participants felt tailored advertising could benefit them, existing notice and choice mechanisms are not effectively reaching users.

5.1 Introduction

In recent years, Internet advertising has become increasingly tailored to individual users. In the simplest case, *contextual advertising*, advertising networks choose which ads to display on a webpage based on the contents of that page. In the more complex technique of *online behavioral advertising* (OBA), advertising networks profile a user based on his or her online activities, such as the websites he or she visits over time. Using this profile, advertising networks show ads that are more likely to be of interest to a particular user, charging a premium price to do so [119]. Targeting advertising based on web searches has been shown to increase ad click-through rates considerably when compared with untargeted advertisements [120].

This chapter is based on "Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising" [118].

OBA presents both benefits and downsides to users. If their interests have been accurately profiled, users will receive more relevant advertising. However, collecting data about users' online activities can potentially violate their privacy. Previous research has found that users have substantial privacy concerns about OBA [?,113,114], while marketing surveys have found that consumers like OBA and that discomfort with OBA is reduced when users are properly informed that non-personally identifiable information is used for OBA [121]. This past work has employed surveys, which can sample a large number of individuals but are not conducive to open-ended questions or follow-up questions to explore attitudes and motivations. In contrast, we conducted interviews to learn how past experiences, knowledge, and understanding factor into users' attitudes towards online behavioral advertising.

Gaining a deeper understanding of consumers' attitudes towards OBA is particularly timely since initiatives are currently in progress to bolster consumer privacy concerning OBA. In March 2012, the U.S. Federal Trade Commission released a report on consumer privacy that discussed OBA at length. In this report, the FTC called on the advertising industry to "[give] consumers greater control over the collection and use of their personal data through simplified choices and increased transparency" [5]. Having a thorough understanding of how consumers perceive behavioral advertising and how they make choices about their privacy enables both technologists and policymakers to provide options and interfaces that better support consumers' privacy expectations.

In this paper, we report results of 48 semi-structured interviews that unpack the factors fueling users' attitudes about OBA. Beyond asking participants their opinions, we investigated their knowledge of the current practice of OBA, their understanding of how profiles can be created, and the extent to which the circumstances of data collection and the identity of the advertising network influence their attitudes.

Participants found OBA to be useful, yet they also expressed strong concerns about its privacy implications. In particular, the majority of participants believed that advertisers could access personally identifiable information. Furthermore, participants were surprised that OBA occurs; while a number of participants believed that browsing history could theoretically be used to target advertising, few were aware that this technique is currently used.

Participants' responses suggested that current approaches for providing notice about OBA are ineffective. Only a handful participants understood the meaning of icons intended to notify consumers about OBA. Participants could not accurately determine what information is collected for OBA purposes, or by whom, and they assumed the worst, leading them to oppose a practice they expected would involve the collection of personally-identifiable and financial information.

Our results also identify disconnects between participants' mental models and current approaches for giving consumers control over OBA. Existing privacy tools ranging from opt-out pages to browser plugins expect consumers to express OBA preferences on a per-company basis. However, participants misunderstood the role of advertising networks in the OBA ecosystem, evaluating companies based solely on activities unrelated to advertising. Participants were unsure where to turn to control OBA, and they most frequently expected they could use settings built into their browsers or antivirus software. Participants expressed complex OBA preferences that depended on the context of their browsing, an approach that is unsupported by current mechanisms.

We discuss related work in Section 5.2 and our methodology in Section 5.3. We report participants' background knowledge of Internet advertising and understanding of OBA icons in Section 5.4. In Section 5.5, we report participants' reactions to learning about OBA, in addition to factors that potentially influence participants' attitudes towards targeted advertising. We unpack the preceding results and discuss directions for better aliging notice and choice mechanisms with users' mental models of OBA in Section 5.6.

5.2 Background and related work

Online advertisers track users as they traverse the Internet, constructing profiles of individuals to enable targeted advertising based on each user's interests. Targeting advertisements can provide benefit to advertisers, helping advertisers find users who are more likely to be interested in the advertised product [122]. Furthermore, since advertising networks can charge higher prices to serve targeted advertising rather than general ads, OBA is "a way to support the websites and products you care about" and may reduce the number of ads consumers see that are not relevant to their interests [123].

In order to provide a baseline for examining misconceptions our participants held, we begin by explaining how OBA works from a technical perspective. We then discuss notice and choice mechanisms for OBA in the United States. Finally, we discuss prior surveys of consumer sentiment towards advertising and OBA. These surveys, conducted by both academics and the advertising industry, have found a range of positive and negative consumer attitudes about OBA. Our interview results provide deeper insight into the genesis and interrelationship between the attitudes reported in these surveys.

5.2.1 The Mechanics of OBA

Since the details of data collection and usage by specific advertising networks for the purpose of OBA can be considered trade secrets, the exact mechanics of how OBA works are generally not public. Nevertheless, basic mechanisms for enabling tracking, as well as methods currently used "in the wild," have been examined in the literature.

In general, the goal of online behavioral advertising is to create a profile of a user's Internet activities, such as the websites he or she visits. This profile can later be used to target advertisements. When a user visits a web page, that page's content can come from both a first party (the page that the user is explicitly visiting) and third parties (companies that have a relationship with the first party allowing them to place content, visible or not, on that page). Third parties include advertising networks, analytics companies, and social networks that contract with first-party websites. These third parties can set a unique identifier on a user's computer. Then, as the user visits different websites that include content from the same third party, that third party can associate these visits with the same computer. In recent years, a small number of third parties have increasingly served content on a larger number of pages, enabling these companies to track a user's browsing across large portions of the Internet [124].

On a technical level, this tracking can be accomplished in many ways. In one of the simplest cases, an advertiser can set a cookie with a unique identifier on a user's computer, correlating browsing activity with that unique identifier [124]. In a study proposing a method for measuring behavioral advertising, Balebako et al. found that blocking third-party cookies in a web browser achieved a reduction in behavioral targeting similar to opting out of behavioral advertising using industry opt-out websites, albeit only testing behavioral targeting in text ads from Google [125]. However, there exist myriad means of uniquely profiling a particular computer, ranging from browser fingerprinting [126] to using Flash Local Shared Objects (LSOs), HTML5 local storage, or other methods of maintaining a unique identifier on a user's computer over time [127]. Many of these techniques aim to uniquely associate browsing activity with a particular computer, rather than with an individual's real-life identity. The Network Advertising Initiative, a U.S. trade group for advertisers, notes, "It is possible to merge PII [personally identifiable information] and Non-PII for OBA and other uses. However, no NAI member ad networks currently engage in this practice" [128].

5.2.2 Notice and Choice

Behavioral advertising has led a number of parties to voice privacy concerns. For instance, the U.S. Federal Trade Commission has noted that data collection can be invisible, privacy notices may be difficult to understand, consumer profiles are sometimes very detailed, and that there is a "risk that data collected for behavioral advertising – including sensitive data regarding health, finances, or children – could fall into the wrong hands or be used for unanticipated purposes" [4].

The most visible attempts in the United States to provide consumers notice and choice about OBA have come about as a result of advertising industry self-regulation by groups such as the Digital Advertising Alliance (DAA) and Network Advertising Initiative (NAI). Policies from both groups have far-reaching implications; for instance, the DAA notes that its members "comprise 85% of the OBA marketplace" [129].

Both groups' principles include the idea of providing consumers notice and choice about behavioral advertising [96, 130]. For instance, the DAA's *consumer control* principle requires that users be able to opt out of receiving targeted advertisements, although it does not require that users be able to opt out of being tracked online [96]. These opt-outs can be enabled via the DAA's opt-out website,[1] on which consumers can opt out of OBA on a per-company basis. The DAA's *transparency* principle requires that consumers receiving OBA be given "enhanced notice," providing "the ability to exercise choice regarding the collection and use of data for online behavioral advertising" via "common wording and a link/icon that consumers will come to recognize" [96]. In 2010, the industry selected the "Advertising Option Icon" to indicate a link to enhanced notice for behavioral ads [131], as shown in the top half of Figure 7.2. However, some advertisements still display an older icon, shown in the bottom half of Figure 7.2, or none at all [17]. Since our interviews took place with U.S. residents, we included questions about the icons and taglines currently used in the U.S.

5.2.3 Previous User Studies

Studies from the past decade have examined different facets of user sentiment towards online advertising in general. In a 2002 paper, Rodgers described two studies with 106 student and 38 non-student participants that looked at interactions between user motivation for using the Internet and the effectiveness of certain types of banner ads, finding that for at least some users, "ads

[1] `http://www.aboutads.info/choices/`

that complement the user's motive may have more success at being noticed and clicked on than ads that do not" [132]. In a 2003 paper, Rettie et al. described a survey with 100 UK student participants, finding that only 13% enjoy Internet advertising. Fewer than 20% of participants found Internet ads informative or useful. Although 62% indicated that they prefer that websites not have ads, 69% agreed with accepting "ads as pay for content" [133]. In 2007, McCoy et al. described a study with 536 participants and found that online advertising caused users to report being both less likely to return to a website and less able to recall features of that website [134]. However, Campbell and Wright conducted a survey study with 97 participants and a laboratory study with 118 participants in 2008, finding that the personal relevance of ads increased users' positive attitudes toward repetitive online advertisements [135]. Taken together, this prior work suggests that users find online advertising annoying, yet targeted ad selection may reduce annoyance.

Other studies have looked specifically at user perception of OBA and online tracking, finding significant privacy concerns about the practice. Turow et al. conducted a 2009 survey of 1,000 US adult Internet-users and discovered that 68% of Americans "definitely would not" and 19% "probably would not" allow advertisers to track them online, even anonymously [113]. In a study published in 2010 that included 14 in-person interviews and an online survey of 314 participants, McDonald and Cranor found that just one-fifth of their online respondents preferred targeted ads to random ads, and 64% thought targeted ads were "invasive." The study found that "people understand ads support free content, but do not believe data are part of the deal" [136]. In a 2009 online study of 2,604 participants, Hastak and Culnan found that 46% of respondents were uncomfortable with the identities of the websites they visit being used to target ads, although this number decreased to 30% of participants when the practice was transparent and offered participants the choice not to receive targeted ads [137]. A 2012 Pew telephone survey of 2,253 participants found that 68% of respondents were "not okay with targeted advertising because [they] don't like having [their] online behavior tracked and analyzed" [114].

Stakeholders from both the privacy-services and advertising industries have also surveyed consumers about OBA. TRUSTe conducted a 2011 survey with 1,004 United States residents, asking about perceptions of OBA. 53% of participants agreed that online privacy is "a really important issue that I think about often," and another 41% agreed that it is "a somewhat important issue that I think about sometimes." Over a third of participants agreed with the statement: "I know how to protect my personal information online and consistently take the necessary steps to do so." Over half of participants indicated that they definitely or probably would not share their browsing behavior with advertisers, and only 15% indicated being willing or probably willing to consent to being tracked online for relevant ads. Only 8% of participants indicated liking OBA, and only 5% showed awareness of the Advertising Option Icon [138]. A 2011 marketing survey of 9,600 individuals across 31 countries found that 90% of respondents expressed concerns about the privacy of their personally identifiable information, yet 62% were willing to allow online advertisers to track their web usage "under the right circumstances" [139]. Our work fills in the gap of understanding how consumers are simultaneously privacy-concerned and willing to have their information collected. Furthermore, we explore these previously nebulous "right circumstances," identifying situations that cause particular concern for consumers.

Much of this past work has employed surveys to gauge the attitudes of a large number of participants. However, surveys are inherently limited in that they don't provide a way for consumers to discuss ideas and thoughts outside the questions asked, or for followup questions to be asked. As a result, surveys alone cannot fully explain how different nuances of attitude are connected to each other. Our research thus delves deeper than prior studies, employing 48 in-depth, in-person interviews. We gain an understanding of why users are hesitant to be tracked for OBA: privacy concerns and misunderstandings of OBA both appear to play roles. We also provide insight into how users make choices about tracking for specific companies and in specific scenarios, as well as how they understand industry-established OBA icons.

5.3 Methodology

In August 2011, we recruited 48 participants for a combination interview and usability study of privacy-enhancing tools. This study was approved by the Carnegie Mellon University IRB. All participants were recruited from the Pittsburgh region of the United States using Craigslist, flyers, and a university electronic message board. Recruitment material directed prospective participants to a screening survey. For the purposes of the usability study, we required participants to be familiar with either Internet Explorer 9 or Firefox 5, be willing to test privacy tools, and have no previous experience with the tools tested. We screened out individuals who had a degree or job in computer science or information technology.

The study lasted approximately 90 minutes. In this paper, we report on the results of a semi-structured interview that took place in the first 30 minutes of each session. The second part of the study was a usability test, which is discussed in detail in Chapter 6.

Interviews took place at the CyLab Usable Privacy and Security Laboratory on the Carnegie Mellon University campus. Each of the 48 interviews was moderated by one of two researchers who had jointly moderated 11 pilot interviews. We used audio recording to document each session. Participants were compensated $30 for participating in the combination interview and usability study.

5.3.1 Semi-structured Interview

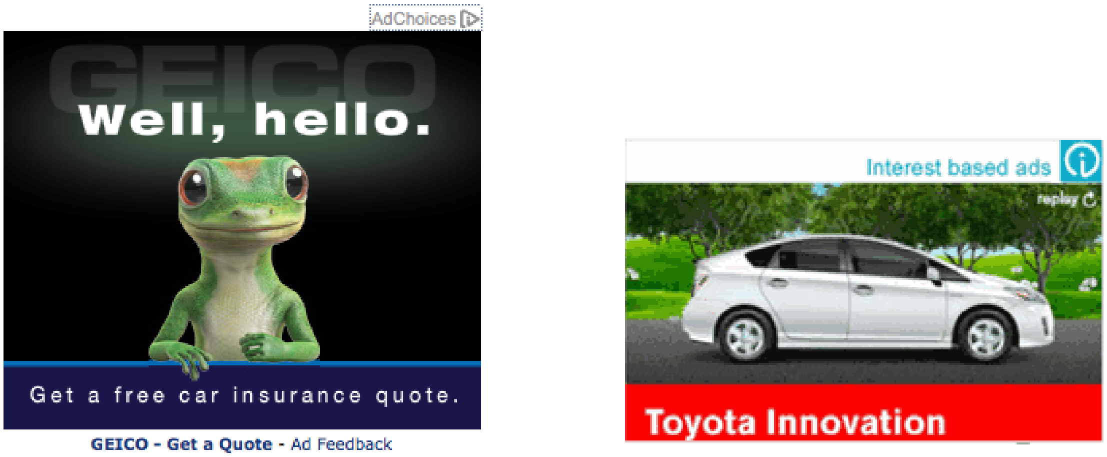
Our semi-structured interviews consisted of two parts separated by an informational video about OBA. The appendix contains the full interview script. The first part of the interview gauged participants' opinions and knowledge of Internet advertising. After participants watched the video, we asked more detailed questions about participants' impressions and understanding of online behavioral advertising.

Interview part one: We began with general questions to explore participants' attitudes about Internet advertising. Then, we asked questions about tailored advertising and interviewees' knowledge of online tracking mechanisms. To evaluate participants' knowledge and perception of Internet icons, we showed two disclosure icons, both of which are depicted in Figure 7.2. These icons were the Advertising Option Icon, which the DAA has standardized [140], as well as the older "Power-I" icon, which was still in use as of August 2011 [17].

The icons and accompanying taglines were first shown alone, and then "in context" on an advertisement. The DAA specifies that the Advertising Option icon should be displayed with one of three approved taglines [141], of which "AdChoices" is commonly used, while the "Power I" icon is usually displayed with the tagline "Interest based ads." We spent between five and ten minutes on this first portion of the study.

Informational video: When piloting the interview, we noticed that participants were generally unfamiliar with OBA. To give participants a baseline understanding of OBA for the remainder of the questions, we showed participants an informational video produced by the Wall Street Journal for their "What They Know" series.[2] The video lasted approximately 7 minutes. We selected this video because it clearly explains what behavioral targeting is and how cookies are used in the process of tracking online activities for the purpose of delivering tailored ads.

Interview part two: Following the video, we evaluated participants' understanding of behavioral advertising. Then, we asked questions about the benefits they perceived for users and other stakeholders. We also asked about any negative aspects they perceived in OBA activities. Next, we presented six hypothetical browsing scenarios, asking whether participants would be willing to have information collected about their browsing for the purpose of OBA in each situation. We further asked participants about their familiarity with advertising companies and willingness to allow these companies to collect information about their web browsing to tailor ads. Finally, we asked participants how they believed they could stop receiving targeted ads if they wanted to do so.



a) AdChoices icon

b) Power I

Figure 5.1: OBA disclosure icons "in context" on an ad.

5.3.2 Analysis

The moderators audio recorded the interviews and took notes during the interview sessions. Once all interviews had been completed, we collaboratively developed a codebook of salient themes we identified in the moderators' written notes. We then coded the audio recording of each interview, transcribing quotes that were especially representative of participants' attitudes. If a response was not captured by existing codes, we added a new code to the codebook. Our coding strategy allowed us to identify both common and uncommon concepts that arose during each interview. We further investigated how concepts correlated between questions to capture mental models underlying participants' attitudes towards OBA.

[2] `http://online.wsj.com/video/92E525EB-9E4A-4399-817D-8C4E6EF68F93.html`

Although our results are purely qualitative, we report the number of participants who fell into different coding categories. These numbers are intended to provide a sense of how frequently participants mentioned these concepts and do not imply statistical significance. Throughout the paper, we also report representative quotes with the goal of illuminating the thought process captured by important codes.

At some points in the interview, participants' responses to a particular question varied widely, with a number of views each held by only a handful of participants. When a question elicited a number of divergent viewpoints, we report an exhaustive list of all points mentioned by any participant. In some cases, these views were reported by as few as a single participant, and we don't necessarily expect that these views would generalize. However, we believe it is valuable for the reader to see the range of misconceptions.

5.3.3 Participants

Our 48 participants were fairly well-educated. They included 15 males and 33 females between the ages of 19 and 57 (mean age 29); eight were undergraduate students, 16 were graduate students, 2 were unemployed, and 22 were employed in a variety of occupations. As a result of our screening process, none had a background in computer science or web development. We refer to our participants using codes representing the order in which they were interviewed (P-1 through P-48).

Due to the limited recruitment area, our participants are not representative of the general Internet population. We make no effort to draw statistically significant conclusions, but instead focus on collecting rich qualitative data that allow us to understand the mental models of laypeople, unpacking the rationale behind their attitudes and behaviors.

5.4 Awareness of OBA

Overall, participants believed that online behavioral advertising provided benefits to consumers, yet posed privacy risks. Participants exhibited a lack of trust of advertising companies, most commonly associating the phrase "Internet advertising" with "pop-ups." Few participants understood industry-standard icons intended to communicate to them that OBA is occurring, and the majority of participants believed that cookies could be used by advertisers to collect their personally identifiable information. Furthermore, participants did not understand the role advertising networks play in OBA and felt that their attitude towards OBA would change based on the context of their browsing. However, they were unsure how to control OBA, expecting that they could turn to their web browser and antivirus software to do so.

For the remainder of this section, we report detailed results on participants' opinions of and beliefs about Internet advertising, tailored ads, and potential mechanisms for tailoring advertisements. All of these responses were given before participants watched the informational video about behavioral advertising. In Section 5.5, we present additional results from after participants watched the information video, focusing on participants' reactions to OBA.

5.4.1 Impressions of Internet Advertising

Participants had negative impressions of Internet advertising as a whole, often associating ads with annoying pop-up windows. When asked, “What is the first thing that comes to your mind when you hear ‘Internet Advertising’?” the most common response was “pop-ups” (21 of 48 participants). P-23 was representative in saying, “I think about those really annoying pop-up ads that always kind of pop up out of nowhere, and I just wanna get rid of them and block them, and they don’t go away.” The second most common response was “annoying” (7 participants). According to P-40, Internet advertising is “bothersome, not needed, distracting, potentially harmful or dangerous.” The same participant explained that “it could be a rogue site trying to obtain information in a less than forthright manner.” A number of participants recalled advertising techniques that have been discontinued. For instance, P-7 noted “a few bad experiences with late-90’s spyware,” mentioning both BonziBUDDY and Comet Cursor.

While many participants had negative impressions of Internet advertising, 25 of the 48 participants said “yes” when asked, “Is Internet advertising useful?” Six participants noted that it helps them find new products. Five participants specifically stated that Internet advertising is useful because it pays for free online services, though most participants expressed a desire for it to be less obtrusive. For example, P-33 expressed, “I am OK with it as long as it does not interfere with what I am trying to do… I understand that a lot of sites are free for my use because of advertising.” Five participants volunteered that they generally ignore advertising. For instance, P-40 explained, “I rarely pay attention to [Internet advertising].”

5.4.2 Tailored Advertising

Before watching the OBA informational video, more than half of participants stated that at least some of the advertisements they see on the Internet are tailored to their interests. However, they did not differentiate between contextual and behavioral ads. 34 of 48 participants responded affirmatively when asked, “Do you think that the ads you see when browsing the Internet are tailored to your personal interests?” Of these 34, 7 said that tailoring happens only on Facebook, 5 said only on Gmail, one referred to Amazon, and 24 did not mention a specific website. Although most participants appeared comfortable with Gmail and Facebook customizing ads based on the contents of their emails or their Facebook profile, a few did express discomfort. For example, P-46 said, “It kind of bothers me that the program they use is monitoring my email…It makes you wonder how much access someone else might have to your emails if a program’s monitoring it.” P-34 commented, “Just when I’m on Gmail, for instance, I notice that when I look at an email, the ad at the top seems to cater to what I’m looking at, and I just think that might be an invasion of privacy.”

Overall, participants found it useful to receive tailored ads; 31 of 48 participants responded in the affirmative to our question, “Is it useful for you to see ads that are tailored to your interests?” Non-obtrusive contextual ads were deemed particularly useful.

5.4.3 Beliefs About How Ads Are Tailored

To gauge participants’ awareness of how information could be collected for OBA purposes, we asked, “How do you think online advertising companies decide which ads are more suitable for you?” Participants provided an array of responses, sometimes mentioning several mechanisms for profiling users. The two most popular responses stated that ads could be customized based on a user's browsing history (14 of 48 participants) and web searches (13 participants). P-46 explained, "If a website tracks your history, which I'm not comfortable with, it might know what website you constantly go to." Some of the participants who mentioned browsing history specifically noted that they thought this was a hypothetical technique. For instance, P-22 conjectured, "I guess if they were monitoring what I did on the Internet...But I'd hope they weren't doing that."

Other common beliefs of how advertisements were tailored included a user's Facebook account (10 participants) and "using cookies" (10 participants). However, none of the participants who mentioned cookies could explain how they were used, often assuming that cookies were repositories of information on a user's computer that advertisers could access. P-34 explained, "I guess they can get into the cookies. I don't know all the details or understand all the technical details about it." Other common responses were the contents or subject of emails (8 participants) and previous online purchases (5 participants). Furthermore, 4 participants volunteered that websites sell or share customers' information. In the words of P-32, "I imagine that if I bought something from a website, that information may be bought/sold/shared with other websites as well."

5.4.4 Interpreting OBA Icons

The Digital Advertising Alliance states that "a prominent feature" of self-regulation is to "clearly inform consumers about data collection and use practices through enhanced notice provided via an icon" [140]. We asked participants about their familiarity with both the DAA's Advertising Option Icon (with the text "AdChoices") and older "Power I" icon (with the text "Interest based ads"). We further investigated what purpose participants believed these icons served, as well as what participants thought would happen if they clicked on the icon.

To gauge participants' familiarity with the icons themselves, we first presented enlarged icons and their corresponding taglines on a white background, lacking context. 41 of 48 participants responded that they had never seen either icon. One participant recalled having seen both icons, one participant recognized only the AdChoices Icon, and four participants stated they had seen the "Power I" icon. The remaining participant was unsure.

When shown these icons in context, next to advertisements, 25 participants still stated that they had never before seen either icon, while 8 were unsure. A total of 15 participants said they had seen the AdChoices icon and tagline, while 13 participants recognized the "Power I" icon and tagline. However, three participants misread the "Power I" icon's "Interest based ads" tagline as "*Internet* based ads," which P-45 felt represented "really great deals *online*."

The purpose of these icons, to provide information to consumers, eluded participants, even when the icons were shown in context on an advertisement. Only five participants thought either icon was intended to provide information about OBA. All five of these participants said the icon informed them that the advertisement was tailored to their interests. P-38 was representative of these participants, explaining that the icons "say that maybe these ads are chosen for you specifically based on your interests." However, no participants thought they were being informed about data collection, even though the DAA considers informing consumers about data collection a "prominent feature" of self regulation [140].

There was a wide range of participant misconceptions about the meaning of the icons and taglines, which we report exhaustively in order to illuminate this divergence. One of the most common expectations was that the AdChoices icon would let users choose the categories of ads they'd like to receive. Ten participants expected that clicking on the icon would let users inform advertisers of their interests. P-21 was representative of these users, thinking that "maybe a list would come up of the topics or subjects. You could choose to either add or remove subjects from that list." P-11 expected he could refine advertisers' existing profile of his interests. He thought that after visiting a site unrelated to his interests, he could "[choose] that those kinds of ads aren't targeted to me since I have zero interest in them." On the other hand, P-22 expected that "you could click on something that would say what your interests were or what you were currently looking for, and then a certain number of ads would pop up that were relevant to that."

Ten other participants believed that the icons served to solicit marketers to advertise in that spot. For example, P-21 said, "it looks like an icon advertising advertisements... A 'place ad here' kind of thing." Similarly, P-14 explained, "I imagine the purpose would be to offer you the option to be able to advertise yourself on webpages."

An additional common misconception was that clicking the icon would provide additional information about the advertised product. Eight participants expected to be taken to the website of the product being advertised if they clicked the icon, and seven additional participants thought they would see additional, related ads after clicking on the icon. For instance, P-18 expected to "see some advertising" while P-41 anticipated being presented with "a list of ads." Two other participants believed they would receive more information about the advertised product.

Other participants expected that clicking the icons would lead to negative consequences or pop-ups. Five participants believed that clicking on the icon would let advertisers track users, three participants believed that pop-ups would appear, and P-41 thought that the advertisement would expand. P-20 believed the icons "might be some kind of a scam, or different pop-ups would come up for the ad," while P-23 thought the icons were intended to "get more information about you somehow."

Some participants expressed surprise at the question itself when we asked what would happen if they clicked on the icon; they believed the icon couldn't be clicked. For instance, P-48 said, "I wouldn't imagine that you could click on it. I would just think it would be like part of the ad." In contrast, three participants thought the icon was intended to differentiate the ad from the content of the page and two thought it was intended to legitimize the ad. Four participants said they had no idea what the icons indicated, while many others were uncertain about their answers. For instance, P-06 thought that "because people are usually annoyed by ads, they want to let people know that they have choices. But I am not sure which kinds of choices."

5.5 Reactions to OBA

After showing participants the informational video, we investigated participants' reactions to and knowledge of the OBA ecosystem, which we report in this section. Many participants found value in behavioral advertising, yet the majority of participants noted that the practice negatively impacted their privacy; taken as a whole, participants found OBA smart, useful, scary, and creepy at the same time. They had many different impressions of the types of information collected during OBA, potentially influencing their attitudes towards OBA. Participants also varied in the situations in which they would like data to be collected for OBA purposes, as well as their opinions of companies that conduct OBA.

5.5.1 Opinions of OBA

Participants found both pros and cons to online behavioral advertising, weighing benefits such as helping consumers find products they might be interested in against privacy fears.

Participants felt that OBA's main benefits for users were helping them to find things they were interested in (19 of 48 participants) and seeing more relevant ads (18 participants). Four participants mentioned that OBA could provide a better Internet experience, while four other participants thought OBA could help them get better deals. For example, P-18 thought that while a consumer is shopping for books and "a competitor suggests a cheaper price, it can help you to save money."

Participants were also cognizant of the economic benefits advertisers can reap from OBA. 26 participants pointed out that advertisers could better target the right person, while 18 participants specifically mentioned that advertisers could make more money using OBA. Only five participants made a distinction between the company displaying an advertisement and the company whose product is advertised, noting that more merchants will advertise with a particular advertising network if it can better target potential customers. In addition, nine participants noted that OBA can command more money for websites on which ads are placed.

In contrast, the majority of participants noted privacy concerns as the main negative aspect of OBA. 41 of the 48 participants expressed concerns related to privacy when asked, "Are there any negative aspects of behavioral advertising?" In particular, participants disliked the idea of being monitored (8 participants) and complained about lacking control (6 participants) over this practice. P-12 explained, "I don't really like the idea of someone looking at what I am looking at, and that kind of freaks me out. Also, I do not like the idea of them putting stuff on my computer without me knowing about it." P-20 was especially concerned with the lack of control, expressing, "The user should be able to decide what kind of ads the user wants to see. The user needs to be in control." Similarly, P-31 said, "It is a little creepy... because I feel that I should get to decide what is going in and out of my computer."

While many participants were concerned about their information being used to create accurate profiles of them, five participants also feared the creation of inaccurate profiles. P-5 was annoyed because she felt the ads she sees are supposed to be tailored to her, yet do not match her needs. She explained, "I feel that sometimes advertisers stereotype me. I find this to be offensive." Similarly, P-45 commented, "Sometimes you click things by accident and it takes you there... it's collecting all this information about you that doesn't even describe who you are, or it could be someone else using your computer." P-27 noted that after making Google searches for her job investigating sexual homicides, she started receiving explicit advertisements that were unrepresentative of her actual interests.

Participants' overall feelings about OBA were mixed, weighing the perceived usefulness of OBA against privacy concerns. After participants spoke about the benefits and downsides of OBA, we asked, "Overall, how do you feel about online behavioral advertising?" On balance, 8 participants had primarily positive feelings about OBA, 19 viewed the practice negatively, and the remaining 21 were mixed. P-16 was representative of those who felt positively, stating, "I don't really see it as anything harmful unless I'm unaware of companies getting more personal information." She did note that she expects OBA to continue regardless of consumer sentiment, explaining that "all the companies are out to make money, so I don't see it stopping."

Many participants liked that OBA would show them more useful ads, yet they were concerned about privacy. P-41 was representative of this attitude when she said, "It seems like it can be helpful for the users, but at the same time it is also dangerous for the privacy problems." P-38 also had mixed feelings. She said, "I think the idea's good, but I don't like the fact that I feel like it's an invasion of your privacy. It makes me feel very insecure. Like if this is what people can figure out about me, then what else can they get off my computer?" P-43 was one of several participants who commented that OBA is a "smart strategy," while P-47 called OBA a "good advertising technique." P-31 synthesized positive and negative aspects of OBA, stating, "It is creepy but clever."

A number of participants suggested ways to make them more comfortable with OBA. P-20 suggested that he would "be more comfortable if the websites or the advertisers ask you directly what are your interests and what are the kinds of things that you like," as opposed to collecting data on user behavior. P-38 added, "I guess I would be more willing to do it if I had like a firmer understanding of how everything worked."

Many of the participants with negative opinions viewed OBA as scary or creepy, though they noted not being very familiar with how it works. For instance, P-14 said, "I don't think I really noticed it...but it definitely is kind of creepy when you think about it." P-45 concurred, relating a story about how she was searching for furniture the previous night and was confused when her advertisements started to feature those items. She stated, "It's scary. It makes me nervous. I was thinking about it last night when I was searching for stuff. Like I thought how do they know all this, how do they keep track of this, how do they do this?"

P-34 was unabashedly angry when she learned about OBA from the informational video. She said, "It makes me want to go home and delete all my cookies, but then I know that's not gonna help much. It makes me mad."

5.5.2 Concerns About Data Collection

Participants believed that advertisers could access a large range of information about them, including personally identifiable information such as their mailing address, name, and potentially even financial information. 26 of the 48 participants stated that advertisers could collect their name and address. P-46 was concerned about "people using it for more malicious means, stealing your credit card information, identity, SSN." Participants also expressed concern about advertising networks collecting and sharing information without telling users. In P-1's words, "They are gathering information...without you knowing it, maybe even giving that data to another party."

Participants commonly said they were scared about being tracked and monitored. P-32 expressed, "It is kind of a creepy thought that you are being followed and monitored." While discussing data collection, P-22 said, "People shouldn't be able to do that. And I think if everybody knew that everything you were doing was being tracked, they wouldn't do half the things that they did." P-40 said she was so concerned with online tracking that she deletes cookies on a daily basis. In a closing thought, P-17 stated, "Following me around, that turns me off."

A number of participants believed that their personal information was stored in cookies on their computers by third-party advertisers. When asked to describe a third-party cookie, P-41 said it was "another cookie that's accessible to my computer history of the web browsing." After learning about OBA from the informational video, P-21 stated, "I've obviously heard of cookies, but I just thought they were temporary Internet files. I didn't know what it was that they were holding, so that's kind of surprising." P-34 hypothesized that to target advertisements, "I guess they can get into the cookies." P-18 described cookies as "little pieces of software that collect certain information about you." However, third-party advertisers most commonly store only a unique identifier in a cookie in order to correlate visits to different websites as coming from a particular browser on a particular machine.

5.5.3 Attitudes Depend on Situation

We presented participants with six different browsing scenarios and asked for each, "Would you like online advertising companies to collect information about your web browsing in order to deliver tailored ads?" Participants were nearly evenly divided about whether to allow or disallow data collection in scenarios about planning a vacation, shopping for a car and car loan, looking for a job, and shopping online for food and household goods. Most participants said they would allow data collection while they were reading the news, while only a few participants expressed a willingness to permit data collection while they searched for STD treatments for a friend.

Participants' preferences were complex. None of the participants said they would allow data collection in all six scenarios, and only five participants said they would not allow data collection in any scenario.

For many situations, participants said they were willing to allow data collection because it would be harmless and might result in cheaper prices for them. For instance, P-22 was willing to allow data collection while she shopped for food and household goods since "there may be a sale on something I wanted anyway."

Privacy concerns drove participants' unwillingness to allow the collection of information. For instance, when deciding whether she wanted her information collected while planning a vacation, P-14 explained, "I'm always looking for...cheaper flights." However, she considered the privacy risk that "you'll know when I'm not at home," before exclaming, "That's tricky!" Health records were a common source of privacy concern. When asked about a hypothetical STD-treatment search scenario, P-45 said she was unwilling to permit data collection. She explained, "That's really personal. The other stuff, it's just material things. That's your health, it's really private."

However, participants sometimes said they didn't want their data collected because they perceived no utility in receiving related advertisements. P-41 declined data collection while shopping for food and household goods, saying, "I know what I'm buying, and I don't want any other distraction to spend money." In other cases, participants sensed a disutility in giving any data to marketers in certain industries. For instance, P-18 felt that "in the travel business, there's a lot of spam," while P-23 didn't want "to be bombarded with car ads for the rest of my life."

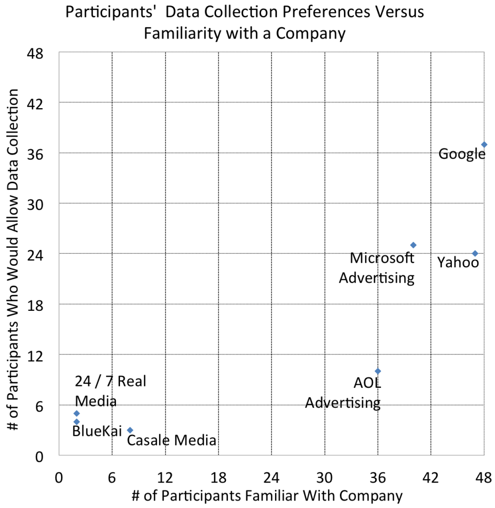


Figure 5.2: Participant familiarity with seven advertising companies versus their willingness to allow each company to collect their browsing information for OBA. The cluster in the bottom left indicates companies that were unfamiliar to participants and with whom they did not want to share information.

5.5.4 Attitudes Depend Partially on Company

While participants were very willing to allow certain companies to collect information about their browsing for the purpose of tailoring advertisements, they were hesitant to allow others. For seven different advertising networks, we first asked, “Are you familiar with [name of company]?” Over 75% of participants were familiar with AOL Advertising, Google, Microsoft Advertising, and Yahoo, whereas fewer than 20% of participants had heard of BlueKai, Casale Media, or 24/7 Real Media. We then asked, “Would you permit that company to collect information about your web browsing to show tailored ads?” As shown in Figure 5.2, 77% of participants would permit Google to collect information about their browsing, while very few would permit the unfamiliar companies to do so. Participants were mixed about whether they wanted Yahoo and Microsoft Advertising to collect their information, while the majority of participants did not want AOL Advertising to do so.

Many of the 37 participants who were willing to grant Google access to their information explained their decision in terms of trust. P-11 was representative in stating, “Their motto is to not be evil, and so far they’ve shown that they’re not.” Most participants mentioned using Google products, explaining that they trust Google because of their positive experiences with these services.

Google’s size and its preponderance of stored data factored into some participants’ decisions. For example, P-21 laughed when asked about Google, saying he was willing to let them collect his information since “they have a lot already.” P-41 recognized both benefits and downsides to Google’s size, stating, “In good ways it’s a really huge company that has a lot of information and it can be helpful. But at the same time, since they’re a really big company, I don’t know what they’re gonna do with my information.” P-31 was among those who felt the scope of Google’s services was a disadvantage, stating, “Google is a bit worse because it is like your doctor has also been your drug dealer. Google is supposed to be my secure email provider and protect my documents...Where do they draw the line?"

Apart from trust, a common explanation for allowing Google to collect browsing information was that it would help in search. For instance, P-16 felt that Google collecting her information "would probably help if I put in a search, if they could tailor it even more towards my interests." P-20 seemed not to recognize Google's advertising activities, stating that "Google is a very good and a safe search engine." Similarly, P-23 was surprised that she was asked about Google advertising since, to her, Google is "not a company that I really associate with advertisements."

As with Google, a number of participants were willing to let Microsoft collect information about their browsing, making this judgment based on the company's non-advertising activities. For example, P-22 was willing to permit data collection, saying, "I know Microsoft has to do a lot of things on your computer if it's your operating system, and I assume that they would collect information that would help them update your operating system." P-44 had the opposite reaction, saying he would not permit data collection from Microsoft because "I am a Mac guy." Many other participants also did not seem to distinguish between Microsoft as an advertising network and Microsoft as an advertiser. For instance, P-16 didn't want Microsoft to collect data "just because I really couldn't see myself buying Microsoft products on a regular basis."

In contrast to Google, Yahoo and AOL were viewed negatively by many participants. P-11 was concerned about Yahoo's viability as a corporation when he said he would not permit data collection, stating, "They're financially not so hot and I wouldn't trust what they would do if they got into a real pinch." P-18 felt that "Yahoo historically has had too many incidents where it made the media that their files were hacked into." When asked about AOL Advertising, P-34 began by stating, "I hate AOL." When asked if this attitude was the result of bad experiences, she continued, "That was a long time ago, and they're still on my list." P-23, who had also been unfamiliar with Google's advertising activities, stated, "I've heard of AOL, but I don't know that I knew that they had advertising."

Few participants wished to permit data collection by BlueKai, 24/7 Real Media, and Casale Media, the three companies that were most unfamiliar. Participants' responses were potentially biased since we asked about their familiarity with each company before asking if they would allow that company to collect their information. However, P-45 was representative of interviewees in stating, "If I don't know the name...I don't trust them, just like you wouldn't trust a friend or doctor you don't know too well." Multiple participants falsely concluded that 24/7 Real Media is a music vendor, potentially confusing it with Real Media. For instance, P-46 said, "Sounds like if you could buy songs from them, I'd be uncomfortable with it, because that means credit card and all that."

5.5.5 Exercising Choice

Near the end of the interview, we asked participants, "Are you aware of any ways that can help you stop receiving targeted ads?" "Deleting cookies" was the most common response by far, mentioned by 25 participants. However, a number of these participants mentioned that they learned about this technique from the informational video they watched earlier in the interview, suggesting that deleting cookies is overrepresented in our data. Three of these participants suggested clearing the browser's cache, in addition to deleting cookies, in order to stop tailored advertising.

Beyond the deleting cookies, participants' responses were quite divergent. In order to show the range of these responses, we report all responses participants gave, grouped into thematic categories.

A number of participants expected that general computer security tools would limit behavioral advertising. For instance, ten participants thought that antivirus and anti-spyware programs, such as Norton or McAfee, would have options for blocking behavioral advertising. To limit OBA, three participants also mentioned firewalls, one suggested using a proxy, and another participant suggested using Linux.

Seven participants expected there to be an option built into their web browser for controlling OBA. However, these participants expressed uncertainty. For example, P-20 thought there "should be an option for that in the web browser... There should be a privacy section." Similarly, P-41 explained, "I think it's already embedded in the computer program, like the Microsoft one. And does the Google Chrome also block the ads?" Three participants specifically mentioned browsers' "Private Browsing" modes as a way to stop receiving targeted advertising. No participants were aware of any specific software or browser plug-ins for managing OBA preferences. In addition, P-17 believed that changing Facebook's privacy settings would limit OBA.

Some participants thought ignoring advertisements was the best strategy for not receiving targeted advertising. Four participants suggested using ad-blocking software, while three others suggested never clicking on ads to control OBA. They believed that companies would be unable to track them as long as they didn't click on advertisements. On the other hand, P-14 thought she should unsubscribe from email lists to limit OBA.

Three participants expected there to be some sort of website on which they could choose to stop receiving targeted advertisements. However, two of these three participants were uncertain if such a website existed. For instance, P-34 "thought there were websites that could help you [stop receiving targeted ads], but I'm not sure." When asked to describe those websites, she said, "Same kind, I would assume, like you could choose your catalogs you want shipped at home. I have no idea, no idea, I'm just guessing." P-37 assumed that such a website might be part of the Norton Antivirus site. P-18 was the only participant who felt certain of the existence of such a website. She explained, "There's supposed to be [a national] agency that oversees marketing... There's a Do Not Call list. There's also, I've been told, a Do Not Email list."

No participants mentioned industry self-regulatory websites or opt-out programs at any point during the interviews. Similarly, no participants mentioned "Do Not Track" or "Tracking Protection List" efforts that are part of popular web browsers. In contrast, twelve participants felt they had no options at all for controlling targeted advertising.

5.6 Discussion

In our interviews, we recorded not only participants' attitudes about OBA, but also their knowledge of its practice. In this section, we discuss how participants' understanding and misunderstanding of OBA may have influenced their attitudes, filling a gap left by prior work employing surveys. We further discuss why many participants seemed unaware that their browsing activities are currently used for OBA purposes and felt unable to control OBA in accordance with their preferences. We conclude with suggested directions for improving notice and choice mechanisms.

5.6.1 Explaining Participants' Attitudes

Participants recognized in behavioral advertising both benefits for consumers and economic advantages for advertising networks. The most commonly articulated benefit of OBA was that it would help participants see advertisements targeted towards and relevant to their personal interests. Some participants also expected that marketers would use OBA to target special offers to consumers who were interested in a particular product.

However, privacy in the face of OBA was the most common concern for users, preventing them from wholeheartedly embracing behavioral advertising. Although some participants had specific privacy fears, such as being monitored or being profiled inaccurately, it was mostly a general, abstract notion of privacy violation that participants articulated. Despite recognizing the potential benefits of OBA, the majority of participants said they were opposed overall to the concept because of general fears of their privacy being violated.

Participants were reluctant to accept OBA, and they exhibited a distrust of advertising born from poor past experiences. Participants recalled aggressive advertising on the Internet to an extent that their most common free-association with "Internet advertising" was "pop-up." The association participants made between advertising and the words "pop-up" and "annoying" suggests that they lacked trust in the advertising industry, which may have led them to set a high bar to acceptance of behavioral advertising.

Although participants were generally aware that some advertisements were somehow targeted to them, many believed this tailoring occurred contextually on Gmail or based on their activities on Facebook. Participants' familiarity with contextual advertising is not surprising since the data used to select the advertisement is in clear view on the same page as the ad. Although many participants believed that advertisers could hypothetically choose advertisements based on which websites they had visited, they were less aware that this technique is currently practiced.

By its nature, behavioral profiling occurs over the long-term, temporally separating the presentation of an advertisement from the data that influenced its selection. Profiling based on browsing or search history doesn't inherently provide consumers with obvious context clues. To combat this structural lack of notice, a core tenet of industry self-regulation in the U.S. is to inform consumers when an ad has been tailored based on browsing activities. However, from our interviews, the icons intended to provide this notice seem ineffective. Many participants saw these icons, which are intended to inform them about data collection and use, and assumed they could click the icon and place an advertisement themselves, or receive more information about the advertised product. Considering the failure of these icons, which are the primary vehicle for notifying consumers about OBA, it is not surprising that participants were unaware that behavioral profiling currently occurs.

When participants learned that behavioral advertising currently occurs, even some of the participants who believed it to be theoretically possible expressed surprise. OBA is not visible on its own, and participants did not glean the intended meaning from industry icons for notifying consumers about OBA and providing a gateway to additional information and the possibility of opting out. This surprise, combined with participants' close association between the advertising industry and annoying pop-ups, may have led them to assume the worst about potential privacy violations due to OBA.

Had participants understood the profiling technologies that underpin most behavioral advertising strategies, pairing a browsing history with a unique identifier, they would have realized that some of their worst fears about the collection of personal information were likely unfounded. However, they did not understand these mechanisms, nor should they. A consumer should not need to be a technologist to be empowered to control the use of his or her information. Participants said that they had heard from the news or from the informational video we presented that cookies enable advertisers to track their browsing history, but that cookies also must be enabled for e-commerce shopping carts to function. Therefore, they seem to have constructed a mental model in which cookies on the computer store both their browsing history and their financial records from past online transactions, which would explain why a number of participants incorrectly believed that advertisers could look at their existing cookies to gather personal information. Following this incorrect mental model, when advertisers or hackers scoop up these cookies from the consumer's computer, they have access to a wide range of personal and financial information previously used for online commerce.

While participants recognized the benefits of OBA, they were upset that it currently occurs without their knowledge. They assumed that the same untrusted advertising companies that bombard them with annoying pop-ups are likely violating their privacy in other ways. Participants were unaware of the types of data that are used for OBA profiling. They were also unaware that long-standing industry guidelines explicitly prohibit the merging of personally identifiable information with previously collected non-personally identifiable information for OBA purposes without explicit user consent [130]. On the other hand, participants might rightly have been concerned if they had heard media reports about web sites accidentally leaking data to advertisers or about companies exploiting technical mechanisms to circumvent privacy protections [124, 142].

5.6.2 Notice and Choice

Because the icons intended to provide notice and choice to consumers did not convey this information to participants, we unpack how participants parsed these icons. Given the large number of participants who thought the icon solicited potential advertisers to "place their ads here," the icon and its text could more clearly indicate that they are intended for consumers. Considering participants' divergent expectations of the message it was communicating, the "AdChoices" tagline seems particularly ineffective.

Furthermore, the icon's location and presentation are a potential area for improvement. The icons are displayed near or within the advertisement itself, which we speculate may have caused participants to think the icon was part of the advertisement. If this hypothesis is true, it would help explain why many participants expected that clicking the icon's "i" symbol would provide more information about the product advertised. Some participants did not expect to receive information from clicking the icon, and some were simply scared to click it at all. Participants' hesitance may have stemmed from their past experience with advertisements, in concert with the icon's location. Pop-up and pop-out advertisements often have a box in the corner of the ad intended to expand or collapse the ad, so this concern may have been consistent with participants' aversion to advertisements popping out or expanding. To provide effective notice, it might be possible to distinguish more clearly both the provenance and purpose of the icon from the advertisement itself. The icon, its accompanying text, and its location are all potential areas for future investigation.

The taglines that accompany these icons contributed to this lack of notice. Participants' common misunderstanding that they could click on the AdChoices icon to select interesting advertising categories is not surprising. They expected to be able to make *choices* about *ads*. While some participants had no idea what choices they could make, others reasonably assumed that they would be able to choose which types of ads they would receive. A small number of advertising networks, including Google,[3] Microsoft,[4] and Yahoo,[5] already allow consumers to view and edit the categories of advertisements targeted to them. To align better with participants' expectations for the icon, the Advertising Option Icon could provide consumers similar functionality. Giving consumers a reasonable set of choices that they can customize in a small amount of time could benefit both advertisers and consumers. Such a system would empower consumers to control the types of ads they receive and correct inaccurate profiles, as well as provide advertisers with potentially more accurate and actionable information about particular consumers' interests. However, such a system could also be used to leak private information about a user.

5.6.3 Supporting Consumer Control

The reactions, misbeliefs, and mental models of our 48 participants suggest potential directions for improved notice and choice mechanisms. In particular, users could be empowered in ways that more closely match their existing expectations and understanding of OBA.

In addition to their responses to the icons and taglines, participants' expectations about how to stop behavioral advertising bode poorly for existing opt-out mechanisms. Deleting cookies, the most common response to our question about how a consumer could stop receiving targeted ads, would actually nullify existing cookie-based opt-out mechanisms. Although a number of participants stated that they were inspired by the informational video to delete their cookies, others stated that deleting cookies was common knowledge, suggesting that the informational video was not the only driver of this sentiment. When a participant states that learning about behavioral advertising "makes me want to go home and delete all my cookies," the idea of using cookies to record opt-out preferences seems problematic.

Furthermore, few participants thought about going to a website to manage their OBA preferences, and none of the participants mentioned clicking on the icons to limit OBA. However, these are currently the two main opt-out vectors. Participants receive advertisements on websites, so it makes sense that they don't expect to go to other websites to exercise preferences. While it seems reasonable for an opt-out mechanism to be located near behaviorally-targeted advertisements, unless consumers are clearly notified that this icon is not part of the advertisement and that they could control OBA by clicking on it, they will not click it.

While it may seem that more effective opt-out mechanisms would benefit consumers to the detriment of advertisers, participants who suggested never clicking advertisements in order to limit OBA provide a counterexample. Participants didn't understand how their browsing is recorded, so some assumed they could be tracked only if they interacted with an advertisement. In the absence of effective notice and choice about OBA, consumers may choose to disengage

[3] `http://www.google.com/ads/preferences/`

[4] `http://choice.live.com/AdvertisementChoice/Default.aspx`

[5] `http://info.yahoo.com/privacy/us/yahoo/opt_out/targeting/details.html`

entirely from advertising. As a result, more effective notice and choice mechanisms may provide benefits to principals throughout the OBA ecosystem.

Whereas few participants thought they could stop OBA by visiting websites and none were aware of existing cookie-based opt-out mechanisms, participants commonly believed that their existing tools for maintaining security and privacy could stop tracking. Although no participants stated familiarity with any specific browser features related to OBA or tracking (e.g., Do Not Track), participants' mental model of turning to their browsers' privacy settings to stop tracking lends support for tools built into browsers.

While the expectation held by ten participants that the Norton or McAfee antivirus suites contained a mechanism for limiting OBA might appear strange at first blush since OBA is unrelated to computer viruses, it seems to indicate that they expected the software that already protects their privacy and security to extend to OBA. As with browsers' built-in tools, security suite software is a one-stop shop for computer protection. Users did not want to take separate action for each new threat to their privacy. The Spring 2012 inclusion of privacy enhancing tools for OBA in the AVG antivirus suite thus seems to align with participants' expectations [143]. In contrast, opt-out mechanisms that would impose additional burdens on users or introduce new paradigms for protecting their privacy may be misaligned with users' expectations.

Users Shouldn't Need to Evaluate Companies

While most currently deployed mechanisms for consumer choice focus on asking users to make decisions about which advertising companies can serve them OBA, participants' difficulty reasoning about companies reveals flaws in this approach. When we asked participants whether they'd permit particular companies to collect information about their browsing, their reasoning differed for companies that engage only in advertising and companies that have a variety of consumer-facing divisions. Participants were unfamiliar with companies that work primarily in advertising, and many participants said they would prohibit data collection from these unfamiliar companies since they felt unqualified to judge their trustworthiness.

For companies such as Google, Microsoft, and Yahoo, which engage in a variety of activities familiar to consumers, participants misunderstood the activities of these companies' advertising divisions. For instance, some participants assumed that Microsoft Advertising collected data primarily for advertising or updating the Windows operating system. In this mindset, participants made decisions using information unconnected to these companies' OBA activities, and it is possible that these decisions would correlate only weakly with fully-informed opinions about those companies' OBA efforts. For Google and Yahoo, many participants assumed that these companies only advertised on Gmail or Yahoo, misunderstanding the role both companies take as advertising networks across many disparate web pages.

Overall, participants seemed to have great difficulty understanding the role of advertising networks within the OBA ecosystem. They felt uncomfortable allowing unfamiliar companies to collect their information, and they judged familiar companies on activities unrelated to advertising. While a user's perception of a company's overall trustworthiness does seem a reasonable metric with which to make decisions regarding the collection of information, the extent to which users' decisions were based on information unrelated to OBA suggests that their decisions were not fully informed.

Situational Choices

When asked whether they would permit information about their browsing to be collected for the purpose of OBA in different scenarios, participants displayed complex preferences. We speculate that participants' variegated preferences suggest that some users who would not want their browsing tracked in general would be willing to have information collected in certain scenarios. Participants differed in which situations aroused concerns about privacy, and both privacy and utility guided their decisions about whether data could be collected. The idea that privacy is sensitive to the norms of a particular context underlies Nissenbaum's idea of privacy as contextual integrity [144], yet the divergence in attitudes expressed by participants in this study suggests that norms about OBA differ from person to person.

The confluence of privacy and utility in participants' decision-making might suggest an approach in which users specify interest categories, although this idea was not investigated in our study. Such an approach could serve a dual purpose: users could prohibit the collection of data on particular topics for either privacy or utility reasons, while advertisers would have self-reported, potentially more accurate, data on a user's interests.

Investigating how a user might situationally control data collection is a potential direction for future work. Would such a system better empower users, or would even the most usable solution unnecessarily burden users? Would such a mechanism be built into browsers, and if so, what interface design would best help users express their potentially complex preferences?

5.7 Conclusion

Participants found behavioral advertising both useful and privacy-invasive. The majority of participants were either fully or partially opposed to OBA, finding the idea smart but creepy. However, this attitude seemed to be influenced in part by beliefs that more data is collected than actually is. Participants understood neither the roles of different companies involved in OBA, nor the technologies used to profile users, contributing to their misunderstandings.

Given effective notice about the practice of tailoring ads based on users' browsing activities, participants wouldn't need to understand the underlying technologies and business models. However, current notice and choice mechanisms are ineffective. Furthermore, current mechanisms focus on opting out of targeting by particular companies, yet participants displayed faulty reasoning in evaluating companies. In contrast, participants displayed complex preferences about the situations in which their browsing data could be collected, yet they currently cannot exercise these preferences.

Participants were unaware of existing ways to control OBA. To exercise consumer choice, participants expected that they could turn to familiar tools, such as their web browser or deleting their cookies. Deleting cookies, participants' most common response in this study, would nullify consumers' opt-outs. A Do Not Track header has been designed to allow users to set a preference in their browser that does not disappear when cookies are deleted. A handful of companies, including Yahoo [145], have announced plans to implement this header, although efforts to define fully the meaning of Do Not Track are ongoing in the W3C Tracking Protection Working Group.[6] Regardless, it remains to be seen whether this mechanism will provide

[6] `http://www.w3.org/2011/tracking-protection/`

effective choice for consumers. Future investigation is needed to test notice and choice mechanisms that better align with users' understanding of OBA, particularly by taking users' mental models of the process into consideration.

Chapter 6

A Usability Evaluation of Tools to Limit Online Behavioral Advertising

We present results of a 45-participant laboratory study investigating the usability of tools to limit online behavioral advertising (OBA). We tested nine tools, including tools that block access to advertising websites, tools that set cookies indicating a user's preference to opt out of OBA, and privacy tools that are built directly into web browsers. We interviewed participants about OBA, observed their behavior as they installed and used a privacy tool, and recorded their perceptions and attitudes about that tool. We found serious usability flaws in all nine tools we examined. The online opt-out tools were challenging for users to understand and configure. Users tend to be unfamiliar with most advertising companies, and therefore are unable to make meaningful choices. Users liked the fact that the browsers we tested had built-in Do-Not-Track features, but were wary of whether advertising companies would respect this preference. Users struggled to install and configure blocking lists to make effective use of blocking tools. They often erroneously concluded the tool was blocking OBA when they had not properly configured it to do so.

6.1 Introduction

The United States Federal Trade Commission (FTC) and other government regulators have voiced concern about online behavioral advertising (OBA) for over a decade [?]. The FTC defines *online behavioral advertising* as "the practice of tracking an individual's online activities in order to deliver advertising tailored to the individual's interests" [4]. Industry organizations have developed self-regulatory principles and frameworks that call for companies to offer consumers the ability to control targeted advertising. [1] [2]

Consumers may control OBA using a number of tools. However, successful use of these tools requires that the user is able to install a tool, configure it to match his or her preferences,

This chapter is based on "Why Johnny Can'T Opt out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising" [146].

[1] http://www.networkadvertising.org/networks/principles_comments.asp

[2] http://www.aboutads.info/principles/

and use the tool effectively. While these tools have the potential to satisfy the concerns of consumers and regulators, there has been little rigorous evaluation of the usability and effectiveness of these tools.

In this paper, we present results of an in-depth study investigating the usability of tools to limit OBA. We also provide a high-level discussion of usability problems associated with these tools.

We tested nine tools, including tools that block access to advertising websites, tools that set cookies indicating a user's preference to opt out of OBA, and privacy tools that are built directly into web browsers. We conducted a 45-participant, between-subjects laboratory study in which we interviewed participants about OBA, observed their behavior as they installed and used a privacy tool, and recorded their perceptions and attitudes about that tool.

We found serious usability flaws in all nine tools we examined. The online opt-out tools were challenging for users to understand and configure. Users mistakenly believed that opt-out tools were protecting them against tracking when those tools do not provide that functionality. Moreover, the current opt-out approach, which is based on users opting out from specific companies, is ineffective because users tend to be unfamiliar with most advertising companies, and therefore are unable to make meaningful choices. Further, since opting out depends on cookies, privacy-minded users who delete their cookies may unwittingly cancel their opt-out. Users liked the fact that the browsers we tested had built-in Do Not Track features, but were wary of whether advertising companies would respect this preference. Users were confused by technical jargon and complicated settings in some tools. Users also struggled to install and configure Tracking Protection Lists (TPLs) and other blocking lists to make effective use of blocking tools. They often erroneously concluded the tool was blocking OBA when they had not properly configured it to do so.

In the next section we present background and related work. We then introduce the privacy tools that we tested, present our testing methods, and discuss our results. We conclude with a summary of our high-level findings and a discussion of implications for online privacy today. We provide an appendix with more detailed results and screenshots of the tools tested.

6.2 Background and related work

Online advertisers track users as they navigate the Internet, constructing a profile for the purpose of delivering targeted advertisements. Third-party HTTP cookies are the main mechanism used for online tracking. Unlike first-party cookies, which are placed by the domain a user is visiting, third-party cookies are placed by another domain, such as an advertising network. Other tracking mechanisms, such as Flash Local Shared Objects (LSOs) and HTML 5 local storage, enable tracking even when the user clears cookies or switches browsers [147, 148].

6.2.1 User concerns about behavioral advertising

According to a 2009 study [113], if given a choice, 68% of Americans "definitely would not" and 19% "probably would not" allow advertisers to track them online even if their online activities would remain anonymous. McDonald and Cranor found that only 20% of their respondents prefer targeted ads to random ads, and 64% find the idea of targeted ads invasive [136].

6.2.2 Industry self-regulation

The Network Advertising Initiative (NAI) and Digital Advertising Alliance (DAA) are industry organizations that have published self-regulatory principles that mandate that users be able to opt out of ad targeting. Both organizations maintain websites where users can set advertising network opt-out cookies that signal that users do not wish to receive interest-based advertising from companies. However, Komanduri et al. found many instances of non-compliance with the NAI and DAA requirements [17]. A 2010 FTC staff report stated that "industry efforts to address privacy through self-regulation have been too slow, and up to now have failed to provide adequate and meaningful protection" [5].

Another example of attempted industry self-regulation is the Platform for Privacy Preferences (P3P), a standard for computer-readable privacy policies published by the World Wide Web Consortium (W3C) in 2002. P3P compact policies (CPs) are a set of tokens that summarize a website's privacy policy regarding cookies. IE9 uses CPs to evaluate websites' data practices and can reject cookies based on user preference [37]. In Chapter 2 we found that more than 20 of the 100 most-visited sites have inaccurate or erroneous CPs and discovered "thousands of sites using identical invalid CPs that had been recommended as workarounds for IE cookie blocking."

Two recent concepts for controlling OBA are Do Not Track (DNT) and Tracking Protection Lists (TPLs). Users can configure their web browser to send a DNT header with HTTP requests, signaling that they do not want to be tracked. However, there is not yet a consensus on how to define tracking or what websites should do upon receiving a DNT header. In IE9, Microsoft introduced TPLs, which are filter rules that allow users to block all content and scripts from specified websites.

6.2.3 Usability of privacy tools

Prior studies have examined the usability of privacy tools. Cranor et al. designed and conducted user evaluations of a privacy agent that examined websites' P3P policies and notified the user when they were inconsistent with his or her stated preferences [149]. Ha et al. conducted focus groups to examine users' awareness and management of cookies, and asked participants to evaluate two cookie-management tools [150]. In a series of interviews and surveys, McDonald and Cranor found that users were confused by the interface of built-in browser cookie-management tools [136].

A number of authors have offered guidance for the developers of privacy tools. Lederer et al. described five pitfalls in the design of privacy tools and offered suggestions for avoiding them. For example, they caution against designs that "require excessive configuration to manage privacy" [151]. Brunk offers recommendations for developers of privacy software including giving "the user feedback that preventative features are operational" [152]. Cranor advises privacy software developers to avoid privacy jargon, ease configuration, educate users, and use persistent indicators to convey information about the tool's capabilities and current state [153].

6.3 Privacy Tools Tested

We tested the usability of nine tools from three broad categories for controlling behavioral advertising. This list includes three *opt-out tools*, two *built-in browser settings*, and four *blocking tools*. The tools we selected are representative of the range of tools currently available to control behavioral advertising. Where we were aware of multiple similar tools, we selected those that appeared most comprehensive or easiest to use based on the authors' assessments. Tests of Internet Explorer settings were conducted using IE 9 on Windows 7. All other tools were tested using Mozilla Firefox 5.0.1 on either Windows 7 or Mac OS X Leopard.

6.3.1 Opt-out tools

Opt-out tools allow users to set opt-out cookies for one or more advertising networks. If a user sets an opt-out cookie for a particular advertising network, that network should not show a user advertising based on his or her browsing behavior, but may continue to track and profile that user. A separate opt-out cookie must be set for each advertising network. To simplify this process, opt-out tools provide a mechanism for users to opt out of dozens or hundreds of advertising networks all in one place.

DAA Consumer Choice is a web-based opt-out tool hosted by the Digital Advertising Alliance, an industry group. Consumers can go to the DAA website's "Consumer Choice" page,[3] select some or all of the participating companies, and click a button to set opt-out cookies. At the time of our testing, there were 79 participating companies.

Evidon Global Opt-Out is an opt-out tool hosted by Evidon, a company that provides technology to help advertisers comply with industry self-regulatory programs.[4] Similar to the DAA opt-out site, Evidon's opt-out page allows consumers to select companies from which to opt out of OBA. In addition, Evidon provides links to other companies from which a consumer may opt out through other means. At the time of testing, Evidon provided direct opt-out for 184 companies and links to opt-out information for 118 others.

PrivacyMark is a bookmark tool containing JavaScript that sets opt-out cookies whenever it is clicked. PrivacyMark[5] is offered by Privacy Choice, a company that sells privacy-related services to companies and provides free privacy tools for consumers. At the time of our testing, the tool set opt-out cookies for over 160 companies.

6.3.2 Browsers' built-in settings

Web browsers generally include privacy options among their built-in settings. These settings, while less comprehensive than add-ons or tools designed specifically for protecting privacy, are currently available to users of all major browsers. We tested the privacy settings on Internet Explorer and Firefox, the browsers that currently have the highest market share.[6] These browsers offer the ability to block cookies selectively based on a variety of factors, including whether they are first-party or third-party cookies.

[3] http://www.aboutads.info/choices/

[4] http://www.evidon.com/consumers/profile_manager#tab3

[5] http://www.privacychoice.org/privacymark

[6] http://gs.statcounter.com/

Mozilla Firefox 5 includes a privacy panel with a check box to "Tell web sites I do not want to be tracked" by sending a DNT header to each website a user visits. In addition, the privacy panel allows users to select options to delete browsing history automatically or choose to accept no cookies, accept cookies except from third-parties, or accept all cookies, including the option to set exceptions on a per-site basis.

Internet Explorer 9 (IE9) includes an Internet options panel with a privacy tab that displays a six-level privacy slider. These levels restrict or block cookies based on a website's P3P CP. A user can also choose advanced settings that block all first-party or third-party cookies, and set exceptions on a per-site basis. IE9 offers additional privacy features, which we discuss with the *blocking tools*.

6.3.3 Blocking tools

We tested four blocking tools, which allow users to choose domains or patterns to block. When using a blocking tool, users rely on the scope of a list of blocking rules rather than on the good faith of the advertising networks. When a site is blocked, the browser will not communicate with that site, completely preventing that site from tracking the user.

Ghostery 2.5.3 is a browser plugin available for all major web browsers. When a user visits a website, Ghostery[7] finds and disables cookies, scripts, and pixels that are used for tracking. It notifies users about which companies have been blocked and allows users the option of selectively unblocking these companies. Ghostery is now owned by Evidon.

TACO 4.0 blocks trackers and also provides a mechanism for setting opt-out cookies for a number of ad networks, as well as the ability to delete LSOs. In addition, TACO[8] offers features designed to help users protect their online privacy by creating disposable email addresses, protecting the data entered into forms on the Internet, and creating alternate Internet identities for the user. TACO is owned by Abine, a privacy services company.

Adblock Plus 1.3.9 is an open-source tool that relies on subscription lists to determine what to block. When a user installs Adblock Plus,[9] he or she chooses one or more filter subscriptions maintained by third parties.

IE9 Tracking Protection is a mechanism built into IE9 that blocks websites based on Tracking Protection Lists (TPLs). Users may install TPL subscriptions curated by third parties.

6.4 Methodology

We conducted a 45-participant, between-subjects laboratory study in which each participant tested one of nine tools that control OBA. The study took place on Carnegie Mellon University's Pittsburgh campus during August 2011.

6.4.1 Recruitment

We sought nontechnical participants who were not knowledgable about privacy enhancing tools, but who were interested in trying them. Since we were using IE9 on Windows 7 and

[7] http://www.ghostery.com/

[8] http://abine.com/preview/taco.php

[9] http://adblockplus.org/en/

Firefox 5 on Windows 7 and Mac OS X as our testing platforms, we recruited participants who had experience using one of these operating system and browser combinations. Participants, who received $30 Amazon gift cards, were recruited from the Pittsburgh region using Craigslist, flyers, and a university electronic message board. Recruitment material directed prospective participants to a screening survey. We recruited five participants for each of the nine tools we tested, for a total of 45 participants. Prior research has shown that many usability problems that are likely to occur in a given population can be identified with only five participants [154].

6.4.2 Testing protocol

Each 90-minute individual session was moderated by one of two researchers who had jointly moderated 11 pilot sessions. We used audio recording and screen capture to document each session. Participants were randomly assigned to the tools considering their browser and OS preferences. We began each session with a semi-structured interview to gather perceptions, knowledge, and attitude about online advertising. We then showed the participant an informational *Wall Street Journal* video about OBA.[10] Following the video, we probed the participant's attitudes and perceptions about behavioral advertising. Next, we asked participants to perform three types of tasks using a computer in our laboratory configured with their preferred Internet browser and operating system. We reset the browser settings both between participants and between tasks. We asked participants to think aloud as they performed each task and to work as though they were using their own computer.

Installation and Initial Configuration. We provided a simulated email from a friend recommending the assigned tool. The email linked to a website from the tool provider where the participant could download, use, or learn about the tool. After the participant installed and configured the tool to match his or her personal preferences, we asked an After Scenario Questionnaire (ASQ) [154] and open-ended questions to measure his or her perceptions and understanding of the tool.

Configuration of Specified Settings. We next asked participants to configure the tools to match specifications we provided. Tools in the same category had similar specifications. Evidon and DAA participants were asked to opt out of 13 specific companies. Ghostery and TACO participants were asked to block the same 13 companies, which were selected from the pool of companies common to these tools. Participants also chose specific settings for the tool's notification messages. AdBlock Plus participants were asked to subscribe to a specific filtering list and add a specific filtering rule. IE-TPL participants installed a specific TPL and also blocked a specific domain. IE and Firefox participants blocked third-party cookies, allowed first-party cookies, and added two exceptions. Participants using PrivacyMark did not perform this task since that tool cannot be configured. Participants then answered another ASQ survey and verbal questions.

Fine Tuning Settings to Resolve Problems. We then configured the tool to a fairly protective setting and asked the participant to perform five typical browsing tasks with the tool installed and active. Three of these tasks required third-party content, cookies, or scripts to function properly, and thus could not be completed when some of the tools were set to block tracking. We advised the participant to change the tool's settings if he or she faced difficulty completing

[10] http://online.wsj.com/video/92E525EB-9E4A-4399-817D-8C4E6EF68F93.html

these tasks. In one task, we asked participants to watch a video on nytimes.com. Participants testing AdBlock Plus or Ghostery could only see the video after unblocking brightcove.com, disabling the tool on nytimes.com, or completely disabling or uninstalling the tool. Similarly, we asked participants to shop for a laptop on dell.com. When participants testing Ghostery or TACO clicked a button to proceed to the checkout page, nothing happened unless they unblocked omniture.com, disabled the tool on dell.com, or uninstalled the tool. Finally, we asked participants to log into Facebook, using an account we provided, and invite a friend to play Farmville. Participants testing Ghostery and TACO saw whitespace where the game should have been. Participants then answered questions and filled out a System Usability Scale (SUS) questionnaire [155].

6.4.3 Limitations

Due to the limited recruitment area, our participants are not representative of the general Internet population. We make no effort to draw statistically significant conclusions, but instead focus on understanding the underlying problems faced by each participant. As with any laboratory study, participants were not in their usual working environments. Participants only used their assigned tools for about an hour; an experiment over an extended time period might reveal further insights about how users interact with the tools over time and might reveal changes in behavior as users become more familiar with the tools. However, we note that a user who is dissatisfied with a tool within the first hour may opt not to continue using it. Furthermore, because most of these tools offer few visual indicators of what they are doing and do not require ongoing interaction with the user interface, users may not gain much additional familiarity through continued use.

6.5 Results

We first describe our participants' demographics. Then, we present usability results for all three categories of evaluated tools. We summarize our results in Table 6.1.

Tool	Capabilities	Strengths	Weaknesses
Blocking			
TACO	Blocks tracking, sets permanent opt-out cookies and blocks third-party cookies	Sets opt-out cookies by default and prevents deletion. Facilitates awareness of trackers when users click icon or enable alert. Suggests workarounds for broken website elements. Provides diverse privacy features.	Large number of privacy features overwhelmed participants. Configuration interface confusing, includes jargon. Initial configuration took a long time. Difficult for participants to find specific trackers to unblock. Participants unaware that default settings don't block trackers. Participants didn't notice workaround suggestions.

Continued on next page . . .

Tool	Capabilities	Strengths	Weaknesses
Ghostery	Blocks tracking	Facilitates awareness of trackers through on-screen alerts. Alerts helped resolve broken website elements. Easy installation.	Configuration interface includes jargon. Participants unaware that default settings don't block trackers. Multiple steps required to enable blocking.
IE-TPL	Blocks tracking, enables DNT headers	Easy to install TPLs from provider websites.	Configuration interface confusing. Participants unaware that default settings don't block trackers. Participants did not realize they had to choose a TPL in order to be protected. Even when prompted, participants were unable to choose a TPL using the interface. Difficult to unblock specific trackers.
AdBlock Plus	Blocks tracking	Facilitates awareness of trackers when users click icon. Users are forced to pick a filtering list so have protection right away. Blocks ads.	Configuration interface confusing, includes jargon. Difficult for participants to find specific trackers to unblock. Difficult for participants to understand differences between filtering lists.
Opt-out			
DAA	Sets opt-out cookies for 79 advertising companies	Provides links to more information about each tracker. Easy to select specific trackers.	Initial configuration took a long time. Difficult to navigate to actual opt-out page. Not obvious that opting out of all trackers requires switching out of default tab on opt-out page. Participants incorrectly believed that they were opting out of tracking. Participants did not realize that deleting cookies nullifies opt-outs. Opt-outs sometimes fail. Participants unable to confirm opting out was effective.
Evidon	Sets opt-out cookies for 184 advertising companies and provide links to opt out of 118 additional companies	Provides links to more information about each tracker. Easy to select specific trackers. Provides links to non-standard opt-outs. Provides the most comprehensive list of tracker and advertising opt-outs.	Initial configuration took a long time. Participants incorrectly believed that they were opting out of tracking. Difficult to navigate to actual opt-out page. Participants did not realize that deleting cookies nullifies opt-outs. Difficult for users to complete non-standard opt-outs. Opt-outs sometimes fail. Participants confused by "opt-out request sent" messages with no additional information. Participants unable to confirm opting out was effective.
PrivacyMark	Sets opt-out cookies for 160 advertising companies	One-click opt-out.	Participants did not realize that deleting cookies nullifies opt-outs. Participants unable to confirm opting out was effective. Requires dragging icon to bookmarks toolbar, which participants could not find. Tutorial video states incorrectly that tool will stop tracking. Participants thought clicking icon would delete cookies.
Built-in			
IE-Settings	Blocks specified cookie types	Default settings provide some protection.	Configuration interface confusing, includes jargon. Participants couldn't figure out how to block all third-party cookies.

Continued on next page . . .

Tool	Capabilities	Strengths	Weaknesses
Firefox	Blocks specified cookie types, sets DNT headers	Participants could easily block all third-party cookies. Participants could easily turn on DNT.	Participants didn't know what protection DNT provided.

Table 6.1: Summary of strengths and weaknesses of each tool identified by observing participants during usability testing. While most tools have additional strengths and weaknesses, we report here only those that were revealed when study participants interacted with the tools.

6.5.1 Participants

Our participants were fairly well-educated, with concerns about online privacy. They included 15 males and 30 females between the ages of 19 and 57 (mean age 29); each condition had both males and females. Eight were undergraduate students, 15 were graduate students, two were unemployed, and 20 were employed in a variety of occupations. None had a background in computer science or web development. The level of initial knowledge about behavioral advertising was fairly uniform across conditions.

In our initial interview, a number of participants expressed awareness that the ads they see are sometimes tailored to their interests, though they conflated contextual and behavioral advertising. When asked how they think online advertising companies decide which ads may be relevant to users, half of the participants mentioned web browsing history and/or web searches, while many others mentioned social networking activities and the contents of emails. A few participants mentioned that cookies might be involved, though they did not know how. None of the participants demonstrated an understanding of the mechanisms used for tracking. After they viewed the behavioral advertising video, most participants were able to explain roughly behavioral advertising and third-party cookies. When asked about ways to stop receiving targeted ads, most participants mentioned deleting cookies, while some mentioned antivirus software. Only a few mentioned built-in browser settings.

6.5.2 Opt-out tools

Configuration

Participants had difficulty using the DAA's opt-out website both when attempting to navigate from the site's homepage to the opt-out page and also when choosing the companies from which they wished to opt out. Two of the five participants assigned to test the DAA's website (DAA-1 and DAA-4) were unable to find the opt-out page, which is linked from the homepage, until the moderator provided written instructions. Both of these participants accidentally navigated to the page on which advertising companies register to join the DAA, mistakenly believing that this was the opt-out page. DAA-1 remarked, "The application to opt out it is a bit expensive, $5,000 a year." Other participants also experienced difficulty finding the link to the opt-out page.

Once they arrived on the DAA's opt-out page, participants had trouble choosing companies due to the page layout. The DAA's opt-out is organized with the tabs "All Participating Companies," "Companies Customizing Ads For Your Browser," and "Existing Opt-Outs." The default view is "Companies Customizing Ads For Your Browser," which means that many users only opt out of companies that have already begun tracking them. In our test, in which each user began with a new Firefox profile, Yahoo! always appeared alone on this list. Both DAA-3 and DAA-5 only opted out of Yahoo! even though both expressed a desire to opt out of all behavioral advertising. They didn't realize that they needed to go to the "All Participating Companies" tab to choose all companies. The other three DAA participants all opted out of all participating companies. Figures D.2 and D.3 in the appendix show the DAA home page and DAA opt-out default page. Since participants had difficulty navigating the DAA site, the opt-out process took a relatively long time. Participants also expressed displeasure when the DAA website displayed an error message stating that certain opt-outs had failed.

All five participants who tested Evidon successfully located the opt-out mechanism, although EV-2 complained that "the opt out option is hidden." EV-1 initially had problems finding it, saying, "I am not sure where to go to opt out," and EV-3 requested assistance finding the opt-out tab once he landed on the "Manage your online profile" page. EV-1 and EV-3 both chose to "Select All" companies whose opt-out could be completed on Evidon's page, while EV-4 chose to opt out of all companies except Google, 24/7 Real Media, AOL Advertising, and YouTube, which he identified as those he uses and trusts.

Although Evidon provides the most comprehensive list of trackers, including links to manually opt out of sites, we observed that users who wish to opt-out of all companies linked from Evidon's page can expend a large amount of time doing so. Both EV-2 and EV-5 wanted to opt out of all companies available, including those that required manual opt-out. EV-2 explained, "I need to opt-out of everything, otherwise it will be useless." EV-5 spent 47 minutes completing the opt-out process, including landing on opt-out pages in five different languages. "How am I gonna opt-out of this one?" he remarked when he arrived on a Japanese language opt-out page. He completed these non-English opt-outs by using Google Translate, as seen in Figure D.6 in the appendix.

The installation process for PrivacyMark, which entails dragging an icon to a browser's bookmarks toolbar, was confusing for users because of its unfamiliarity. PM-1 was initially confused about where the bookmarks toolbar was located. PM-4 remarked, "Usually software goes through a different installation process." The instructions provided, shown in Figure D.4 in the appendix, incorrectly assume that the user has previously enabled the bookmarks toolbar. This toolbar is not enabled by default in recent versions of Firefox.

Understanding

No participants who tested the DAA website understood what opting out means in this context. Four of five participants incorrectly stated that opting out will stop tracking. Only DAA-5 did not mention tracking, but she thought that opting out "makes it easy to block advertisers from sending you ads." She expected to see 50% fewer ads while browsing, stating that if opt-out doesn't result in fewer ads, "I would think that opt-out is pointless."

All participants who used Evidon's opt-out tool similarly misunderstood opt-out to mean that they could not be tracked or would receive fewer ads. However, Evidon's opt-out website explicitly states, "If you opt out, you will still see ads online, and in some cases data may be collected about your browsing activity."[11] After opting out initially, EV-1's expectation was that she would see "probably only 10% of the ads that I used to see." After completing the browsing tasks, she concluded that she "saw slightly less ads." Most participants mistakenly believed they could no longer be tracked. EV-3 thought that Evidon's opt-out configures "who gets your information and whether they can/cannot use it," while EV-4 believed he was "telling ad companies that I do not wish to participate in tracking behaviors." EV-5 thought he could now browse without "worrying about my information being collected."

The mechanism for opting out confused users. None of the five participants who tested the DAA's website, and only two of the five participants who tested Evidon's website, understood that opting out sets an opt-out cookie on their computer. All other participants who mentioned cookies mistakenly thought that cookies were being blocked. DAA-1 thought he was temporarily stopping cookies, DAA-2 expected that opting out "prevents third-party cookies from being installed on my computer," and DAA-3 said, "it blocks cookie creation and transfer." Evidon participants also thought opt-out blocks access to cookies. For instance, EV-2 said, "Somehow, it will prevent those companies from looking at the cookies that accumulate in my computer." Although they misunderstood the opt-out process, some participants liked that both the DAA and Evidon sites include links to learn about the companies that participate in the opt-out program.

None of the PrivacyMark participants initially understood that the purpose of the tool was to set opt-out cookies. Three of the participants watched the video on PrivacyChoice's website, which states incorrectly that this tool stops online tracking. Common misconceptions were that PrivacyMark either prevented cookies from being sent or deleted cookies. When asked what PrivacyMark does, Participant PM-1 stated, "[PrivacyMark] deletes information, whatever you search for, and that will not be connected to the advertisers." In the eyes of PM-2, PrivacyMark "clears cookies, prevents cookies from being sent, or encodes cookies so that advertisers cannot see them." Participants retained their misconceptions of PrivacyMark's purpose even after performing a number of browsing tasks with the tool installed.

Three of the ten participants who tested either the DAA or Evidon websites drew parallels between opting out and Do Not Call lists. DAA-4 expressed a negative attitude, saying that the DAA opt-out is "almost like Do Not Call lists, not like that works." DAA-5 said, "Everyone gets ads. You have to intentionally remove yourself, like Do Not Call."

The Evidon website's possibility of displaying either "opted-out" or "opt-out request sent" also dissatisfied users. Four of the five participants who tested Evidon's opt-out mechanism disliked receiving the "opt-out request sent" message. EV-1 was typical of these users, saying, "I do not have a way to verify that I successfully opted out. The request was sent, but I am not sure if I actually opted out." Another participant received an "opt-out failed" message, leading him to further question the opt-out process' effectiveness.

Users were also unhappy that Evidon's 'Select All' option only selected the subset of advertising companies whose opt-out could be completed on Evidon's page. EV-1 felt that the idea that "if you select all, you will not opt-out of *all* is misleading." EV-2 echoed, "I liked that you could select all. Unfortunately, you cannot do it." Figure D.5 in the appendix shows the web page that users were shown after selecting "all" and opting out.

[11] http://www.evidon.com/consumers/profile_manager#tab3

Overall, users were unsure of how successful their opt-outs were, with EV-2 stating, "You just have to hope that it is working." EV-4 similarly wondered, "I do not know if I actually did anything." He was also confused about the meaning of the trade group affiliations listed on Evidon's opt-out page, saying, "It would be nice to know what these [DAA, NAI] affiliations are." EV-5, who was redirected to the NAI website a handful of times during his 47-minute Evidon opt-out process, said that he believed that the NAI is an "ad agency" used by a number of companies.

Although PrivacyMark empowers users to opt out with one click, its lack of communication with users was its major usability issue; users wanted an indication that PrivacyMark was working. For instance, PM-2 described the feature she wanted to see in PrivacyMark as "a little notification telling you that it is working, blocking something." PM-5 suggested that she "would like to be able to check from which companies I have opted out. I want to choose specific companies I want to block." PM-4 felt that the lack of communication meant that it was not doing anything, explaining, "In theory, it sounds like a good idea. In practice, it didn't seem to be effective."

Finally, most participants who used cookie-based opt-out tools mistakenly believed that deleting their cookies would further protect their privacy. However, unless they use a tool designed to prevent opt-out cookie deletion (e.g. TACO, Beef TACO, "Keep my opt-outs" by Google, "Keep more opt-outs" by PrivacyChoice), users who delete their cookies inadvertently delete their opt-out cookies, undoing their opt-out.

6.5.3 Built-in tools

Informed users try to block third-party cookies

Although Internet Explorer does block some (but not all) third-party cookies by default, privacy-sensitive participants had difficulty choosing configurations that matched their expectations. Most participants were able to find the privacy settings page, although they were confused by the page's interface and jargon, and also unsure how the P3P-based settings related to third-party cookies. IE-1 spent more than 10 minutes trying to find the Internet Options Window. Although she eventually found the window, she never clicked on the 'Privacy' tab. The other four participants were able to find the settings page, but the settings they chose differed from their expectations in all cases. For instance, IE-4 incorrectly expected that the default settings "will block third-party cookies." IE-5, who chose the 'High' privacy setting, was unsure what that setting actually meant. She said, "I hope what I chose, 'high,' will block cookies from dangerous websites, but from safe ones everything will get through." IE provides explanations next to the privacy levels, but uses terminology related to P3P compact policies, unlikely to be familiar to an average user.

In contrast, participants testing Firefox were able to both configure and accurately describe their privacy settings. For example, FF-1 blocked both first- and third-party cookies, but added exceptions to allow websites she uses, including Amazon.com and Pandora.com. She explained that Firefox "seems to be effective at limiting cookies... I like more stringent privacy setttings, but I have some exceptions, mainly entertainment." FF-4 accepted first-party and blocked third-party cookies, saying that her configuration "clears away all the cookies that you do not want...I wanted less cookies, less tracking, less invasion." The three other Firefox participants kept the default cookie settings, which allow both first- and third-party cookies. However, these participants demonstrated awareness of their settings. For instance, FF-3 explained that she “didn’t want it to not track completely since I’m sometimes interested in ads.”

Users like ‘Do Not Track’ option but are skeptical about its effectiveness

When asked to configure Firefox’s privacy settings as they would on their own computer, four of the five Firefox participants enabled DNT. This suggests that participants like the idea that they can stop tracking with a single click. Nevertheless, users were skeptical about DNT’s effectiveness. For example, FF-5 said, “[DNT] would probably just put a wrench in their program, but they could probably figure something else out.” Both FF-1 and FF-3 correctly realized that DNT relies on advertisers’ good faith. FF-1 mentioned that she learned this from the Firefox privacy webpage we had directed her to at the beginning of the study, explaining, “Firefox says that DNT is voluntary. I would like to think websites will actually respect my preferences, but I am not sure.”

Participants did not understand the details of the DNT mechanism, though they expressed their desire for it to stop tracking. For example, FF-3 felt that DNT meant, “Don’t allow behavioral advertising to happen. Don’t share...my browser history or my information,” whereas FF-4 thought it meant that “websites will not be allowed to collect cookies on me. They will not be able to remember what I have done.”

Browsers differ in the ease of changing settings

We observed a stark difference in the performance of participants testing Internet Explorer and Firefox. When asked to do so, none of the five Internet Explorer participants were able to allow first-party and block third-party cookies. The option to block third-party cookies is contained in the ‘Advanced’ menu, which only IE-2 opened. Rather than blocking third-party cookies as they had been instructed, IE-2, IE-3, and IE-5 chose the ‘Low’ setting on Internet Explorer’s privacy slider, falsely believing they had accomplished their goal. In contrast, all five Firefox users were able to configure the specified settings, including blocking third-party cookies, in 1 to 4 minutes. Figures D.7 and D.8 in the appendix show the privacy settings in Firefox 5 and IE 9, respectively.

Fine tuning settings to fix broken elements

Both Internet Explorer and Firefox users were able to remove Facebook from a blacklist in order to log in. All five Internet Explorer users and all five Firefox users correctly recognized that they were unable to login to Facebook because Facebook had been blacklisted. Although all participants removed Facebook from the blacklist, IE-1 never refreshed Facebook’s page after changing her settings and thus she was not able to login after 10 minutes of trying. It took the other four users between 1 and 5 minutes from when they noticed there was a problem to successfully logging in.

Removing Facebook from the list of blacklisted domains was sufficient for Internet Explorer users to complete the task, but Firefox users needed to perform an extra step that proved difficult for most. Only two of the five Firefox participants were able to invite their friends to

Farmville by enabling third-party cookies. Although FF-4 solved the problem, she was confused by why her solution worked, stating, “I think I am getting confused between third-party cookies and others.” FF-1 displayed similar confusion during her unsuccessful attempt to load Farmville’s ‘Invite Friends’ feature, commenting, “I do not know why cookies are required to invite friends.”

6.5.4 Blocking tools

While participants were able to install all four of the blocking tools, they had trouble configuring them to match their preferences. In many cases, participants erroneously believed they had chosen configurations that would block most or all third-party tracking. When the tools blocked content participants needed to complete browsing tasks, they were often unable to take appropriate corrective action, instead either failing to complete the task or disabling the tool entirely.

Installing blocking tools is easy

Overall, participants experienced few difficulties installing blocking tools. All participants who tested Ghostery, TACO, and IE-TPL were able to install the tool without any assistance, although TACO took participants longer to install. Four of the five participants testing AdBlock Plus installed the tool without assistance, while one participant required assistance finding the options menu. Particiants found the installation process for Ghostery, in particular, to be especially simple.

Participants tried and failed to configure strong protections

Although participants were able to install the blocking tools with relative ease, they experienced difficulty configuring these tools appropriately. Participants were confused by jargon in the interface, and in some cases thought erroneously that they had chosen the most protective configuration when the tool was actually doing little.

Ghostery permits users to block tracking cookies and web bugs, but these options are off by default. Users must navigate multiple steps filled with jargon to turn on blocking, which participants found cumbersome. Only one of five participants blocked all available trackers, the highest level of protection. Three participants did not block any trackers, but two of these participants nonetheless believed they had configured the tool to block all trackers. The remaining participant selected a handful of trackers and cookies to block. Figures D.9 and D.10 in the appendix show Ghostery’s main configuration interface.

All five participants who tested TACO selected the default blocking and opt-out features, which set (and prevent the deletion of) opt-out cookies, yet do not block any trackers. This configuration does not exploit the tool’s significant privacy-enhancing features. Two TACO participants attempted to take advantage of the tool’s diverse identity protection features, even though neither configured any options to opt out of or block web tracking. TACO-2 spent 15 minutes installing the tool and setting preferences, attempting yet failing to configure TACO’s “safe e-mail” and “safe phone number” features. Although she stated that she hoped to block cookies, she was unable to; although she remembered seeing an option to block cookies, she forgot where this option was amid TACO’s many features. TACO-4 stated that she was very concerned with privacy and was determined to use all of TACO's features. After spending 24 minutes trying to configure the tool and watching its video tutorials, she questioned TACO's trustworthiness. She remarked, "Who says Abine is a company to trust? They will collect information about me... I think this is a false sense of security. Give us your information and we will anonymize it. Yeah sure!" Figure D.14 in the appendix shows TACO's main configuration interface.

Four of the five AdBlock Plus participants chose the default filtering subscription list without any further changes, while ABP-4 chose the default list but unblocked Google AdSense. However, none of our participants understood what they were blocking, and most were unsure how to differentiate between the filtering lists offered. Figure D.16 in the appendix shows AdBlock Plus' main configuration screen.

All five participants testing Internet Explorer Tracking Protection also kept the default settings. However, this default setting does not subscribe the user to any TPLs, leaving users with minimal protection. Although all this configuration does is to send a DNT header, participants believed they were configuring the tool protectively. For instance, TPL-2 explained the rationale for his configuration as, "I just tried to get like the maximum privacy." Similarly, TPL-4 stated, "I did not configure anything, but I think it will block all tracking." Figure D.13 in the appendix shows the TPL configuration interface. Participants encountered several usability problems, some previously discussed by Cranor [156], leading them to select less than optimal privacy settings.

Changing configurations is difficult

When asked to configure blocking tools according to a specified configuration, participants' initial problem was often finding the tool again in order to change its settings. Although the add-ons toolbar was enabled, participants ABP-2, ABP-3, GH-2, and TACO-4 all required assistance finding their respective tools. Many of these participants misunderstood the idea of browser add-ons, mistakenly looking for these tools in the "All Programs" area of the Windows Start Menu. Others clicked on "Add-Ons" to open the add-ons manager, but never realized that they needed to click on "Extensions" to see which add-ons were already installed.

Only two TACO participants were able to configure TACO according to the specification we provided, spending 6 minutes and 16 minutes to do so. The three other TACO participants were unable to block web trackers. TACO-2, who spent 8 minutes before giving up, never realized that she could click on the "Not Blocked" text listed under web trackers to block them. TACO-4, who worked for 12 minutes before giving up, expressed, "It is very confusing...How can I block all?" She didn't realize that clicking on a particular category of trackers produced a drop-down menu of the companies whose trackers were blocked. All participants who realized they could click on this drop-down menu complained that companies were presented in a seemingly random, rather than alphabetical, order. Participants noted that an alphabetical list would have been much faster for them. Participants also experienced problems with jargon, confusing the "Targeted Ad Networks" and "Web Tracker" categories in TACO's interface.

Similarly, only two AdBlock Plus participants were able to configure the tool as we specified. Two other participants didn't select the specified filter subscription. Participants had trouble navigating AdBlock Plus' interface and understanding the jargon that accompanied filtering rules. The remaining participant gave up. However, four of the five Ghostery participants correctly configured the tool. The remaining participant required assistance finding the tool's options page and also neglected to enable one specified feature.

When asked to add a specific IE TPL, all five participants were able to do so. However, three participants were unsure how to use the IE interface to add Tracking Protection Lists, instead going to search engines to look for the Fanboy TPL (the TPL we specified) and then downloading it from the Fanboy website. Participants were also unsure whether they actually downloaded any TPLs. TPL-5 wondered aloud,"Did I add it?" after he received no confirmation. IE TPL participants were also asked to configure the personalized TPL to allow and block content from two specific domains, respectively. None of the the participants were able to configure custom preferences that unblock specific trackers.

Fine tuning settings to fix broken elements

Participants testing AdBlock Plus, Ghostery, and TACO all encountered websites that did not work because of the tool. IE TPL participants did not encounter any problems, probably because the TPL that was installed did not block critical content at the visited sites.

In the nytimes.com task, it was easy for participants to notice that there was a problem since they could not watch the required video. All five AdBlock Plus participants and four out of five Ghostery participants realized that the tools were preventing the video from showing up. Every participant who noticed the problem eventually solved it. One AdBlock Plus participant unblocked a single tracking domain, while the other four participants disabled AdBlock Plus on nytimes.com. For instance, ABP-3 realized in less than a minute that something had been blocked, and he spent eight minutes trying unsuccessfully to unblock particular trackers. In the end, he disabled AdBlock Plus on nytimes.com. Figure D.17 in the appendix shows the complexity of trying to unblock a specific tracker using AdBlockPlus. Some participants hovered their mouse cursor over the ABP icon to learn which items were blocked, yet these notifications did not help them to unblock particular trackers. All four Ghostery participants who solved the problem unblocked a single tracking domain, while GH-2 gave up after 4 minutes of attempting to unblock trackers.

In the Dell scenario, it was more difficult for participants to notice problems. The mouse pointer started blinking and the site never responded after participants clicked the checkout button, leading many participants to believe that the Internet was temporarily slow. Five Ghostery and three TACO participants experienced problems; the two other TACO participants did not experience problems due to changes in the Dell website during the course of the experiment.

Three of the Ghostery participants realized that there was a problem on their own, albeit after waiting for over two minutes. However, the two other participants waited for over four minutes until they were primed by the moderator to consider whether Ghostery might be causing the problem. At this point, GH-4 speculated that it was "maybe because I am about to enter personal information," whereas GH-5 attributed the delay to Dell's website. Four of the five Ghostery participants solved the problem by unblocking specific trackers, while the other participant uninstalled Ghostery.

In contrast, none of the three affected TACO participants realized by themselves that something was wrong. After the moderator waited four minutes and then asked the participant whether TACO might be causing the problem, TACO-1 concluded that TACO was the cause. However, TACO-2 still attributed the delay to the webpage, thinking that because she had successfully navigated past the first page of Dell's website, TACO was not causing problems. She said, "I'm like into the page now, so I'm thinking if anything it's just the webpage itself is slow or something... I don't know why it would have anything to do with TACO." TACO-3 also attributed the delay to network issues, explaining, "It just seems to be taking a few minutes. I hit the 'review and checkout' button. It's just not loading." When prompted whether TACO might be causing the problem, she decided that TACO might be protecting her from entering personal information. The only TACO participant who solved the problem, TACO-1, unblocked one web tracker and solved the problem in about two minutes.

The Facebook/Farmville task was easier for many participants than the Dell task, both because they had learned about unblocking trackers in previous tasks and because the failure was more evident, as in the nytimes.com task. In the Facebook/Farmville task, all Ghostery participants experienced problems inviting friends yet were able to solve the problem in about one minute. Four of these participants unblocked specific trackers, while the other participant simply uninstalled Ghostery. Four of the five TACO participants experienced problems inviting friends. TACO-1 did not experience problems since she noticed TACO's message that other participants have recommended different settings for this site, and she chose to accept those changes. None of the other TACO participants noticed this message even though all received it. TACO-3 again thought that TACO might be blocking her actions because she was about to enter personal information, although she was not certain that TACO was causing the problem. The two other TACO participants never considered TACO as the culprit. TACO-3 gave up after seven minutes without ever noticing the alert about recommended changes. After it was pointed out by the moderator, TACO-4 noticed the TACO alert at the top of the page, but she decided to reject the changes and gave up. TACO-5, however, found an alternate route through the page that circumvented the blocked objects, never realizing that TACO had caused any problems.

Understanding and willingness to use

Participants found the feedback provided by Ghostery and TACO useful, helping them gain a better understanding of what the tools were doing. For example, participants liked that Ghostery listed the trackers blocked on each web page visited. GH-4 explained, "[Ghostery] shows me who is collecting my data." However, GH-2 mistakenly believed that Ghostery "helps companies [recommended by Ghostery] to track my browsing history."

Most Ghostery participants indicated that they were willing to use the tool. GH-3 said, "It tells you exactly what trackers are on the web page and gives you control to block them." Participants did indicate a desire for a better explanation about what web trackers are and how to use the tool, as well as an ability for the tool to adjust its settings automatically to fix broken elements on websites. For example, GH-3 said, "It would be nice if it could realize what the context is. For example, if you are on Facebook, apps should work."

Similarly, participants liked TACO because they can click the TACO icon to see who is attempting to track them. TACO-1 said "It tells you what companies are tracking you, and you can click [them] on and off." Figures D.11 and D.15 show the alerts provided by Ghostery and TACO, respectively. These alerts improved participants' awareness of tracking and understanding of the purpose of these tools.

Four of the five TACO participants said they would use TACO in their daily browsing because it reduces the amount of tracking. Nevertheless, TACO-4 was not confident about using the tool, finding it cumbersome.

Participants were commonly confused about IE TPLs. All five participants misunderstood what TPLs do and were unable to differentiate between them. Participants did not seem to trust the third-parties that produce TPLs. For example, TPL-4 erroneously believed that Fanboy, a popular TPL curator, "is probably a top advertising company."

In contrast, all five AdBlock Plus participants said they would use the tool in their daily browsing. Participants liked the tool's easy installation and that it blocked ads, although they found configuration difficult. ABP-4 explained, "Filter subscription: I do not really know what that is... Most of these are kind of jargon to me... To be honest, I do not really know what these things are apart from the Google one."

6.6 Discussion

None of the nine tools we tested empowered study participants to effectively control tracking and behavioral advertising according to their personal preferences. We identify the usability problems that appear endemic to this space, and we split these usability errors into thematic strands.

6.6.1 Users can't distinguish between trackers

The opt-out websites, as well as the Ghostery and TACO browser add-ons, provide users with lists of companies that they can block or from which they can opt out. However, users don't recognize the majority of these companies. We observed that users generally chose the same settings for all companies on the list. A few users made exceptions for a handful of companies with names they recognized, but mostly users attempted to block trackers from all companies. Users were unable to set opt-out or blocking preferences meaningfully on a per-company basis. In order to better match user expectations, blocking and opt-out tools should allow users to easily opt-out of all tracking. They should provide more fine-grained choices as an advanced setting and allow users to configure exceptions if they so desire, but not assume that most users are going to exercise such fine-grained control. Filter subscriptions and TPLs allow users to delegate these decisions to trusted experts; however, tools need better interfaces for selecting and installing these lists. In addition, tool providers should develop and test other ways of grouping trackers into meaningful categories that allow users to block or set opt-outs on a per-category basis rather than a per-company basis.

6.6.2 Inappropriate defaults

None of the tools that are not bundled with browsers have default settings that are appropriate for their target audience. If a user proactively downloads a browser add-on like Ghostery or TACO, or proactively visits an opt-out website, their action indicates that they likely intend to block tracking. However, Ghostery and TACO do not block any trackers by default, and enabling tracking involves multiple clicks. Similarly, no advertising companies are selected by default on the DAA and Evidon opt-out sites.

The general population of Firefox and IE users may have a different set of expectations. Thus, it might be appropriate for browsers' built-in privacy settings to have less protective defaults. However, once a user enables a browser privacy feature such as TPLs, a protective default for that feature seems reasonable. IE Tracking Protection requires users to subscribe to a TPL before the feature provides additional protections. While automatically subscribing users to a TPL would require Microsoft to select a default TPL, user interface changes could make users more aware that they need to select a TPL, guiding them to do so.

6.6.3 Communication problems

The tools we tested were ineffective at communicating their purpose and guiding users to properly configure them. The tools tended to present information at a level that is either too simplistic to inform a user's decision or too technical to be understood. For instance, Internet Explorer 9 provides a simplistic privacy slider whose six levels (e.g. "medium") do not describe their functionality. In contrast, participants were unable to understand the jargon-filled technical explanations next to the slider. Ghostery and TACO used the following terms whose distinction was meaningless to participants: Web Tracker, Web Bug, Flash Cookie, Silverlight Cookie, Tracking Cookie, Script, IFrame, and Targeted Ad Network. In addition, participants testing opt-out tools did not understand what the tools would opt them out of, mistakenly believing that they were protected against tracking. Furthermore, opt-out tool users thought deleting cookies would protect their privacy even more, not realizing that deleting their cookies would also delete their opt-out cookies and undo their opt-out.

6.6.4 Need for feedback

Many of the tools we tested provide insufficient feedback to users. Users were left unaware whether or not most tools were working, and oblivious to what was happening behind the scenes.

None of the opt-out tools tested notify users while they are browsing that their preferences are being respected. Furthermore, participants were unsure of what it meant to be opted-out and how they could tell whether opt-out was working. Participants who tested the browser cookie settings also had no mechanism for understanding what exactly was happening behind the scenes unless websites didn't work. DNT mechanisms also provided no feedback; however, there is currently no way for tools to confirm that DNT preferences are being honored.

While AdBlock Plus did not provide explicit feedback, users noticed the absence of all ads on pages they visited and inferred that the tool was effective.

In contrast, Ghostery and TACO users received notifications on every website visited about what companies were attempting to track them and whether trackers had been blocked. Users appreciated this feedback and gained an understanding of what the tool was doing. However, future work is needed to determine whether these notifications become less useful or annoying over time, and whether users stop noticing them.

6.6.5 Users want protections that don't break websites

Participants had difficulty determining when the tool they were using caused parts of websites to stop working. In cases where some content was not displayed or features stopped working, it appeared to participants that the problem was due to their Internet connection. They were especially confused when problems did not occur on the first page of a particular site, but only on subsequent pages.

Some participants suggested that the tools should be able to detect these problems automatically and change their settings accordingly. TACO is able to detect browsing problems and suggest changes based on feedback from other users. However, most participants didn't notice TACO's notification about these recommendations. An improved notification might be helpful. Another option would be to adjust the settings automatically without waiting for user confirmation. However, there is a risk that tracking companies might game the crowdsourcing system to have their trackers unblocked. TPLs have the potential to address this problem by allowing users to subscribe to a list that has been curated to block most trackers, except those necessary for sites to function. However, participants in our study were unaware of the need to select a TPL and unsure how to decide which TPL to select. In addition, users expressed a desire to easily delete all tracking cookies without losing essential site functions, improving privacy without compromising functionality. This suggests that built-in browser tools should provide an easy way not only to block third-party cookies but also to delete third-party cookies without deleting first-party cookies.

6.6.6 Confusing interfaces

The tools we tested suffered from major usability flaws. For instance, multiple participants opted out of only one company on the DAA's website despite intending to opt out of all. Others mistook the page on which companies register for the DAA as the opt-out page. Participants testing TACO never realized that they were not blocking any trackers. Furthermore, it seems that TACO bundles too much functionality; multiple participants never realized they could block tracking or third-party cookies since they were confused by features related to anonymous email. Participants did not understand AdBlock Plus' filtering rules. None of the participants who tested IE Tracking Protection realized that they needed to subscribe to TPLs until prompted in a later task. When we asked them to subscribe to a particular TPL, most participants did not use the IE TPL interface but instead performed a Google search for the name of the specified TPL and subscribed via its website.

6.6.7 Conclusion

In our 45-participant lab study, we evaluated the usability of tools that limit OBA. We found serious usability flaws in all nine tools evaluated, demonstrating that the status quo is insufficient for empowering users to protect their privacy. Although we recognize the efforts of the advertising industry, browser providers, and third-parties for contributing an assortment of tools to this ecosystem, we encourage a greater emphasis on usability moving forward.

Our results suggest that the current approach for advertising industry self-regulation through opt-out mechanisms is fundamentally flawed. Users' expectations and abilities are not supported by existing approaches that limit OBA by selecting particular companies or specifying tracking mechanisms to block. Users have great difficulty distinguishing between tracking companies. They also lack sufficient knowledge about tracking technology or privacy tools to use existing privacy tools effectively.

There are significant challenges in providing easy-to-use tools that give users meaningful control without interfering with their use of the web. The list of advertising companies and the technologies for tracking are changing constantly, making it difficult for tool providers, let alone users, to keep up. It is difficult and time consuming to determine the purpose and privacy practices associated with every tracker on a website. It is also difficult to determine which trackers can be blocked without breaking desired website features. Even with additional education and better user interfaces, it is not clear whether users are capable of making meaningful choices about trackers.

Chapter 7

What Do Online Behavioral Advertising Privacy Disclosures Communicate to Users?

Online Behavioral Advertising (OBA), the practice of tailoring ads based on an individual's online activities, has led to privacy concerns. In an attempt to mitigate these privacy concerns, the online advertising industry has proposed the use of OBA disclosures: icons, accompanying taglines, and landing pages intended to inform users about OBA and provide opt-out options. We conducted a 1,505-participant online study to investigate Internet users' perceptions of OBA disclosures. The disclosures failed to clearly notify participants about OBA and inform them about their choices. Half of the participants remembered the ads they saw but only 12% correctly remembered the disclosure taglines attached to ads. When shown the disclosures again, the majority mistakenly believed that ads would pop up if they clicked on disclosures, and more participants incorrectly thought that clicking the disclosures would let them purchase advertisements than correctly understood that they could then opt out of OBA. "AdChoices," the most commonly used tagline, was particularly ineffective at communicating notice and choice. A majority of participants mistakenly believed that opting out would stop all online tracking, not just tailored ads. We discuss challenges in crafting disclosures and provide suggestions for improvement.

7.1 Introduction

Online advertising companies increasingly use a sophisticated mechanism called Online Behavioral Advertising (OBA) to gather data about users' online activities, build models inferring users' interests, and display advertisements accordingly. OBA can benefit advertisers by increasing click-through rates [119, 120]. While some users may enjoy receiving more relevant ads, many users are concerned about being tracked. In response to privacy concerns about OBA [4],

This chapter is based on "What Do Online Behavioral Advertising Privacy Disclosures Communicate to Users?" [157].

the online advertising industry has established a self-regulatory program based on user education, transparency, and consumer control over OBA [96].

Advertisers primarily use OBA disclosures in the form of icons and accompanying taglines to provide consumers transparency and control. These icons and taglines are placed near behaviorally-tailored ads. Clicking on these disclosures directs users to *landing pages* that explain OBA in more detail and outline the choices users have for managing their opt-out preferences.

We present the results of a 1,505-participant online, between-subjects study investigating the messages that icons, taglines, and landing pages actually communicate to Internet users. We tested the "advertising option icon," three taglines that the online advertising industry currently uses, an alternative icon, three additional taglines, and five landing pages from major online advertisers.

We found that the OBA disclosures and landing pages fell short both in terms of effectively drawing participants' attention and communicating clearly about notice and choice. Only 28% of participants remembered seeing the icon, and fewer than 12% correctly recalled the tagline they had been shown. The "Why did I get this ad?"[1] tagline was most memorable. It was also most effective for communicating notice, followed by "Interest based ads" and "Learn about your ad choices." However, no tagline was effective at communicating choice. More than half of participants believed clicking the disclosure would trigger pop-up ads, and a similar fraction thought clicking would indicate interest in the advertised product. "AdChoices," currently in wide use, was ineffective at communicating notice and choice. While landing pages were effective at communicating notice, the majority of users mistakenly believed that opting out would stop online tracking.

We discuss background and related work in the next section. In Section 7.3, we explain our methodology. In Section 7.4, we present our results. Finally, we discuss the implications of our results and potential opportunities for improvement in Section 7.5.

7.2 Background and related work

We first provide a brief background on online behavioral advertising and industry self-regulation. We then review related work evaluating OBA disclosures and tools, followed by work on the effectiveness of disclosure icons and taglines in a variety of domains. Finally, we discuss research on the design of privacy notices.

7.2.1 Online Behavioral Advertising

The U.S. Federal Trade Commission (FTC) defines *online behavioral advertising* as "the practice of tracking an individual's online activities in order to deliver advertising tailored to the individual's interests" [4]. Online advertisers track users as they navigate the Internet, constructing profiles for the purpose of delivering targeted advertisements. Third-party HTTP cookies are the main mechanism used for online tracking [124]. Unlike first-party cookies, which are placed by the domain a user is visiting, third-party cookies are placed by another domain, such as an advertising network. Studies have found that users are more likely to click on targeted ads. Yan

[1] In Fall 2011 Google used this tagline. As of March 2012, Google appears to be using a variety of taglines, including "AdChoices" and "Why these ads?"

et al. found that behavioral targeting led to improvements of up to 670% in the clickthrough rates of ads [120].

Although OBA is popular with advertisers, Internet users have concerns about the practice. In a 2009 telephone survey of 1,000 Americans by Turow et al. [113], 68% of respondents "definitely would not" and 19% "probably would not" allow advertisers to track them online if given a choice. McDonald and Cranor found that only 20% of respondents to their online study of 314 Americans preferred targeted ads to random ads, while 64% of respondents found the idea of targeted ads invasive [136]. Hastak and Culnan found in 2010 that only 24% of respondents were comfortable with OBA in the absence of transparency and choice [137]. In a 2012 Pew telephone survey of 2,253 participants, 68% of respondents said they were "not okay with targeted advertising because [they] don't like having [their] online behavior tracked and analyzed" [114]. In Chapter 5 we interviewed 48 users, finding that they perceived benefits in OBA, but that both privacy concerns and misunderstandings made them reluctant to embrace the practice.

7.2.2 Industry Self-Regulation

In February 2009, the Federal Trade Commission released a set of principles designed to guide industry groups' efforts to self-regulate OBA practices [4]. The FTC's principles focus on transparency, disclosure, and consumer consent. The Network Advertising Initiative (NAI) and Digital Advertising Alliance (DAA) industry organizations responded with self-regulatory principles. Both organizations maintain websites[2] where users can set opt-out cookies signaling a desire not to receive OBA.

One form of notice adopted by the industry is the use of uniform icons, links, and accompanying text—which we term "taglines"—disclosing that advertisements are behaviorally targeted. For instance, the DAA advises that participating entities "use the Advertising Option Icon and one of the approved wordings to represent adherence to the Self-Regulatory Principles for Online Behavioral Advertising and as a means for providing enhanced notice of online behavioral advertising practices" [140]. These disclosures are typically placed just above an ad. The icon and tagline serve as clickable links to a landing page, which describes the advertising company's OBA practices and gives the user the option to opt out of OBA or change his or her OBA preferences with that company.

7.2.3 Evaluation of OBA Disclosures

In 2009, the Future of Privacy Forum (FPF), a think tank, contracted with the WPP advertising company to develop icons to label OBA. The FPF commissioned Hastak and Culnan to conduct a study to test possible icons and taglines. Based on results from two focus groups, two icons and seven taglines were selected for study. The two icons tested were the "Power I" and "Asterisk Man." Power I looks like a letter "i" with a circle around it, similar to a computer's power button. Asterisk Man looks like a cross between an asterisk symbol and a stick figure.

In an online study of 2,604 participants, Hastak and Culnan measured the effectiveness of the icons and taglines at communicating OBA. Although they found that two of the tested taglines were statistically better at communicating than the others, taglines and icons were in

[2] `http://www.networkadvertising.org/choices/` and `http://www.aboutads.info/choices/`

general not effective at providing notice and choice about OBA. They found that Asterisk Man performed slightly better than Power I on several comprehension measures [137]. However, the circular Power I icon was modified to be triangular and became the industry-standard Advertising Option Icon. Our work differs from Hastak and Culnan's since we evaluate disclosures in the context of a simulated browsing scenario, rather than on a single page. We also evaluate the landing pages to which users are taken when they click the icon.

In Chapter 5 we found that most interviewees had difficulty interpreting OBA disclosure icons and both the "AdChoices" and "Interest Based Ads" taglines. Multiple participants misread "Interest Based Ads" as "*Internet* Based Ads." These results informed our selection of statements for evaluating OBA privacy disclosures.

7.2.4 Communicating with Icons

A number of studies have examined icons as a means to communicate information. Huang and Bias compared how visual representations and textual information were interpreted by 78 students; participants understood the semantics of an object or concept more quickly and more accurately when communicated with text [158]. Wiedenbeck studied the performance of 60 undergraduates using a computer interface that communicated information using only icons, using only text, or employing both icons and text. Although participants initially rated the text-only interface poor on perceived ease of use, participants performed poorly when using the icons-only interface for the first time, suggesting that text is important for initial communication in unfamiliar situations [159]. Haramundanis surveyed the use of icons in software, arguing that text performs an essential role in accompanying icons; she posits that icons cannot stand alone [160]. Taken together, this work suggests that text taglines should accompany unfamiliar icons.

Studies have also focused on the visual design of icons. For instance, Kunnath et al. compared the learning and performance of 53 graduate students when information was communicated using one of three types of icons: abstract, pictorial (photos), and line drawings. They found pictorial icons resulted in better learning and performance than abstract icons or line drawings [161]. This work suggests that the abstract icons used for OBA disclosures might not communicate semantic meaning or concepts effectively.

Icons have been evaluated in domains ranging from pharmaceuticals to foods, often in the context of communicating risk information. In a study of 406 students, Wang used a yellow "warning symbol" to draw attention to health disclosures on pharmaceutical advertisements, finding that this method of visual priming leads study participants to express greater trust towards the advertisements [162]. Employing a sample of 520 adults, Andrews et al. studied the "smart choices" nutrition icon, which is designed to condense a product's nutritional information into a single front-of-package indicator. They found that participants more positively evaluate the nutritional content of products displaying this icon, even products with debatable nutritional content [163].

7.2.5 Evaluating Taglines

Taglines and other phrasal, textual communications have been studied in the context of advertising slogans, particularly as they relate to brand recognition. Lee found that including a tagline with a brand name can cue a person to recall the brand from memory [164]. In a study of 174 undergraduates, Boush found that slogans can either ease or undermine attempts to extend a brand to new products [165]. Dahlén and Rosengren found that slogans carry brand equity and are better liked when associated with stronger brands [166].

Taglines have been studied in both the healthcare and consumer marketing domains. Williams and Koepke evaluated 18 potential taglines for promoting Medicare information sources. They found "answers to your health care questions" and "helping you help yourself" were preferred by participants when the participants ranked a set of taglines. In contrast the less context-specific taglines, "so much more than you think," "it's all you need to know," and "get the most out of it," were rated lowest by participants [167].

7.2.6 Communicating Privacy

A growing body of work has examined how to make privacy disclosures more usable. Most privacy disclosures are presented as long plain-text documents. Studies have indicated that people do not read these policies, do not understand them, and do not like them [20,69,83]. McDonald and Cranor estimated that if Americans actually read privacy policies, it would take 244 hours per year per person, corresponding to a national opportunity cost of $781 billion dollars [85].

Researchers have evaluated alternatives to text privacy policies. Kelley et al. proposed and tested a tabular "privacy nutrition label," taking cues from the standardized presentation of the nutritional information of foods. They found that standardized privacy policy presentations allowed users to better understand privacy policies and do so more quickly [168]. Garrison et al. [74] found that a table format significantly improves comprehension of a privacy notice in comparison to other formats, including those currently popular.

Reducing privacy policies to icons has proved challenging in past work. Internet Explorer 6 introduced a status bar privacy icon that shows when cookies have been blocked [169]. The icon—a stylized eye with a red, do-not-enter road sign—can be difficult to notice and understand. Cranor et al. developed "Privacy Bird," a browser helper object that uses bird-shaped icons with word balloons to indicate whether web sites comply with a user's privacy preferences. However, in lab tests, users commonly misinterpreted these icons [149].

7.3 Methodology

Our goal was to evaluate the ability of OBA disclosures to empower users to make privacy choices. We conducted a between-subjects online study. We randomly assigned each participant to an experimental condition that consisted of an icon, tagline, and landing page. Half the participants were also assigned to a condition in which they were primed to believe that ads they were shown during the study were behaviorally targeted.

We recruited 1,548 users of Amazon's Mechanical Turk (MTurk) crowdsourcing service to participate in what we described as an "Internet Usage Survey" in December 2011. We required that participants be at least 18 years old and live in the United States. Participants were compensated $1 for the study, which took 24 minutes on average. 20 participants were excluded from the data set for using web blocking tools that prevented them from seeing part of the study. 23 other participants were excluded for providing answers unrelated to the study in response to the majority of open-ended questions. The remaining 1,505 respondents comprise our data set.

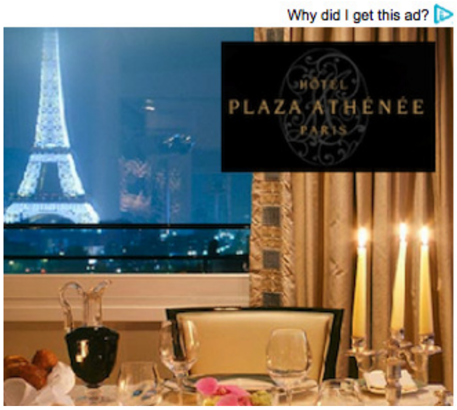


Figure 7.1: Advertisement shown to all participants for a Paris hotel. The icon and tagline on the top-right corner were assigned randomly from two and six options, respectively. The total area used to display both the ad and OBA disclosure was a square of 330px per side.

7.3.1 Study Protocol

The study was conducted entirely online in a participant's web browser. Each participant was first presented a consent form[3] and told they would be asked about their Internet usage and opinions of webpages and online advertising. This broad description was intended to prevent users from initially realizing what was being studied.

Each participant was randomly assigned an experimental treatment that specified the form that OBA privacy disclosures presented in the study would take, as described in Section 7.3.2. Participants began by providing demographic information and rating their agreement or disagreement with general statements about Internet advertising. We then asked participants to conduct a Google search on one of two possible topics: "traveling to Paris" or "buying a Nissan car." We asked them to visit two websites from the search results and report briefly on their impressions. Participants next answered several general questions about Internet usage. We then asked them to go to a simulated version of the *New York Times* front page and provide their impressions of the page, report the most interesting headline, and identify any privacy-protection mechanisms that they saw. The top of the news page contained two advertisements for Air France, each of which was 210px wide by 75px high and contained the privacy disclosure specified by the participant's treatment. Consistent with current industry practice, disclosures in the study were located above the ad and justified to the right side. The area used to display a disclosure was 165px wide by 20px high. The page also contained a 330px wide by 310px high ad for a Parisian hotel, as depicted in Figure 7.1. This advertisement, located on the right side of the page below a list of op-ed article headlines, contained the same privacy disclosures as the AirFrance ads.

After viewing the simulated news page, participants clicked a button to continue, automatically closing the *New York Times* window or tab so that they could not refer back to it. We then asked participants about the products advertised on the website, and whether they had seen a symbol or short phrase near the advertisements. These symbols and phrases were the OBA privacy disclosures. If they answered affirmatively, participants were asked to answer more detailed

[3] This study was reviewed and approved by the Institutional Review Board at Carnegie Mellon University.

questions about the symbol or phrase. Taken together, this portion of the study investigated the extent to which participants noticed OBA privacy disclosures in context.

Next, we showed participants an ad with the OBA disclosure specified by their treatment, absent any website context. We asked participants to interpret these disclosures through multiple-choice questions, open-ended prompts, and a series of true/false statements to which participants responded on a 5-point scale ("Definitely not," "Probably not," "Not sure," "Probably," or "Definitely").

In the final portion of the study, we asked participants to click on an OBA disclosure icon and visit a company's landing page. On a landing page consumers are presented information about OBA and given the opportunity to opt out of receiving behavioral advertisements. A participant's assigned treatment dictated which of five landing pages, described in Section 7.3.2, he or she would see. We again asked participants questions in a variety of formats about their interpretation of this landing page. We concluded the study with a final set of questions about participants' privacy concerns and uses of privacy-protection mechanisms.

7.3.2 Treatments

We assigned participants randomly to experimental treatments across three major dimensions: the priming they received before the simulated browsing scenario, the privacy disclosures they saw during the scenario, and the landing page they were shown.

Priming

The first dimension of the experimental treatment was designed to give half the participants reason to believe the advertisements they saw in the simulated browsing scenario were behaviorally targeted. Before the simulated browsing scenario, participants were asked to search for, visit, and describe two websites on one of two randomly assigned topics: "traveling to Paris" or "buying a Nissan car." Those who searched for Paris travel were considered to be primed for behavioral advertising. During the browsing scenario that followed, all participants were shown a simulated version of the *New York Times* website containing ads for travel to Paris, regardless of their priming. Participants primed toward a trip to Paris could have reason to believe that these ads had been tailored based on their search, while participants primed toward a car purchase would not believe that the ads were behaviorally targeted.

Icon

All study participants were randomly assigned to see one of two icons: the blue Asterisk Man icon previously tested by Hastak and Culnan [137], or the Advertising Option icon consisting of the letter "i" in a blue triangle. The Advertising Option icon is the current standard required by the Digital Advertising Alliance [140]. The icons we tested are shown in Figure 7.2.

Tagline

Independent of the icon shown, participants were randomly assigned one of seven conditions for the tagline, a phrase to the left of the icon. In one condition, no tagline was displayed. All other

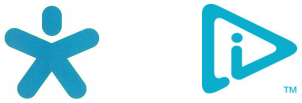


Figure 7.2: The two OBA icons studied. Each participant was randomly assigned to see either the "Asterisk Man" icon (left) or the "Advertising Option" icon (right).

conditions were shown one of the following taglines: "Why did I get this ad?" "Interest based ads," "AdChoices," "Sponsor ads," "Learn about your ad choices," or "Configure ad preferences."

We selected the first three taglines ("Why did I get this ad?" "Interest Based Ads," and "AdChoices") because they have been approved by the Digital Advertising Alliance [140]. Hastak et al. previously tested these three taglines and found that they were not effective at communicating notice [137]. "AdChoices" is the tagline that has been most widely used by advertising companies, and it is currently being used in multiple languages. "Sponsor ads" was used by Hastak et al. as a control and was not expected to communicate effectively about notice and choice [137]. We tested "Learn about your ad choices" as an alternative to "AdChoices" that includes an action. We tested "Configure ad preferences" to test the impact of "configure" and "preferences" as key words.

Landing page

The final dimension of our experimental treatment randomly assigned participants one of five landing webpages currently in use. These webpages are intended both to notify consumers about data collection and use as well as to provide consumers with the opportunity to opt out of receiving OBA. The five landing pages we used come from the advertising companies AOL, Yahoo!, Google, Microsoft, and Monster Career Network.

7.3.3 Statistical Analysis

Most of our data for this study was categorical. For instance, we provided participants with statements about online advertising to which they responded on a 5-point Likert scale ("Strongly Agree," "Agree," "Neutral," "Disagree," "Strongly Disagree"). We binned participants' responses into *agreement* ("Strongly Agree," "Agree") and *non-agreement* ("Neutral," "Disagree," or "Strongly Disagree"). We also showed participants statements about the OBA disclosures tested, some of which were true and some of which were false. Participants again responded on a 5-point scale ("Definitely Not," "Probably Not," "Not Sure," "Probably," "Definitely"). We again binned responses into *yes* ("Definitely" or "Probably") and *non-yes* ("Not Sure," "Probably Not," or "Definitely Not").

For omnibus comparisons among conditions, we used Pearson's chi-squared test (noted in our results as χ^2) on the binned responses. We also performed pairwise comparisons of all treatments. Since the frequency of responses in some categories could potentially be low, rendering χ^2 p-values unreliable, we used Fisher's Exact Test (noted in results as FET) to perform these pairwise comparisons. Post-hoc comparisons, including the pairwise comparisons, were corrected for multiple testing using the Holm-Bonferroni method, indicated in our results as "HC."

To test interaction effects between icons and taglines, we performed a logistic regression using the Asterisk man and "blank" (i.e. no tagline) as control categories for icon and tagline treatments, respectively. We did not find any significant interactions. Further, the results from the logistic regression analysis were consistent with the pairwise comparisons performed using chi-squared tests. Therefore, p-values reported in the remaining of this paper are those that resulted from our pairwise comparison analysis.

7.3.4 Limitations

Our study was conducted online, which enables a large number of participants to take part, yet introduces a number of limitations. As with any online study, we were not able to prevent participants from answering randomly or disregarding instructions. However, we manually verified key responses to open-ended questions to verify that the participant's answers related to the study and excluded participants for whom the majority of responses to open-ended questions were unrelated to the study (23 participants). Participants could view the study on any operating system and browser, with their preferred plugins installed. We could not perfectly control the context in which the participant took the study. Participants could have searched for information about OBA disclosures online or tried to keep the simulated browsing session open while answering questions about it. We used Javascript to close the simulated browsing session when the participant moved forward in the study, although it is possible that some participants found a workaround. In addition, regardless of their priming condition, participants might have perceived the ads shown as tailored had they been interested in vacation trips at the time the study took place.

As we conducted our study on MTurk, we are subject to its demographic biases. United States MTurk workers are not representative demographically of U.S. Internet users. MTurk workers in the U.S. trend younger, more female, and more educated than the general population [170]. They also tend to be lower income than overall U.S. Internet users [171]. Despite the known biases, previous studies have shown that Mturk participants behave similarly in studies to subjects recruited from other sources [172], and that MTurk can provide a sample that is at least as diverse as participants recruited from other online or laboratory channels [173].

Some of the icons and taglines we tested, and all of the landing pages, are deployed in the wild. It is possible that some participants had seen their experimental treatment previously, potentially influencing their responses. Furthermore, due to the time-limited nature of an online study, the context in which study participants viewed OBA disclosures is not a perfect proxy for viewing these icons over a long period of time. However, we think that if a user does not understand the purpose or message of a disclosure the first time it appears, repeated exposure to this disclosure will not necessarily clarify its meaning.

7.4 Results

We analyzed responses from 1,505 participants, finding that the OBA disclosures we tested perform poorly. Participants ranged in age from 18 to 82 (mean = 32, SD = 11.5). We did not observe any statistical differences in education, technical background, gender, age, or Internet usage across treatments.

We first discuss the effects of our priming conditions. We then present results on the extent to which participants recognized the disclosures as privacy mechanisms, whether they noticed them, and whether they could recall them later. Next we discuss the messages conveyed by the disclosures, including the extent to which they conveyed notice and choice, as well as the expected results of clicking on them. Finally, we present participants' perceptions and understanding of the landing pages to which OBA disclosures link.

7.4.1 Effect of Priming

In an attempt to simulate the experience of seeing tailored ads, half of the participants were assigned to perform a Google search about taking a vacation to Paris, while the other half were asked to perform a search about buying a Nissan car. When participants later saw ads from Air France and a Paris hotel, we expected (but could not validate) that participants who had searched for travel to Paris would perceive these ads as targeted, while participants who had searched for cars would not.

Responses from participants who were primed for traveling to Paris did not differ significantly from those of participants who were primed for purchasing a Nissan car for any question in our study. We cannot conclude whether our priming task was ineffective or whether participants were oblivious to tailored advertising. Although we have anecdotal evidence suggesting that users do not correlate their Internet browsing with the ads they see, why our priming was ineffective remains an open question.

Since the priming did not seem to have any significant effect, the results presented in the following sections consider all participants together, regardless of their priming condition.

7.4.2 Identifying Privacy Mechanisms

While showing the news website, we asked participants to "Explain as completely as possible what privacy protection mechanisms (if any) do you see on this news webpage." A handful of participants expressed uncertainty about how to identify privacy mechanisms, and a few dozen explicitly mentioned that there were not any privacy protection mechanisms on the news website.

Overall, participants did not perceive the icons and taglines as being associated with privacy protection mechanisms. However, a small number of participants (fewer than 10% of participants in any treatment) recognized that some of the taglines might be associated with privacy protection mechanisms. In particular, in the "Configure your ad preferences," "Why did I get this ad," "Learn about your ad choices," and "Interest based ads" treatments, 16, 14, 8, and 6 participants, respectively, mentioned the icon or tagline. One participant in the "AdChoices" treatment expressed "...seems you may be able to filter or choose what ads you don't mind viewing." One "Interest based ads" participant misread the tagline, saying "there is a little icon to opt out of *internet* based ads" [emphasis added].

Regardless of the tagline treatment, many participants referred to the privacy policy link and TRUSTe seal at the bottom of the page. A few others mentioned the "Terms of Service," "Your Ad Choices," and "Contact Us" links at the bottom of the page. In addition, some participants mentioned that the opportunity to create an account or log into the news website could be seen as mechanisms to protect their privacy. Some said they believed that registered users would

Tagline	Remembered	Not remembered
Why did I get this ad?	49 (22.3%)	171 (77.7%)
Interest based ads	27 (12.6%)	187 (87.4%)
Learn about your ad choices	24 (10.7%)	200 (89.3%)
Configure ad preferences	22 (10.8%)	181 (89.2%)
AdChoices	17 (7.9%)	199 (92.1%)
Sponsor ads	15 (7%)	200 (93%)
Overall	154 (11.9%)	1,292 (88.1%)

Table 7.1: Tagline recall across conditions. “Why did I get this ad?” was recalled at a significantly higher rate than all other taglines except “Interest based ads.”

receive better privacy protection. On the other hand, a small number of participants noted that by logging in they would be identifying themselves to the website, which could reduce their privacy. Several participants mentioned that the sole fact that the news website was not asking for personal information could be seen as a privacy protection mechanism. Finally, a few participants conflated privacy with security and referred to the lack of security on the page (i.e. no https) as something that could affect their privacy.

7.4.3 Recall of Ads and OBA Disclosures

After participants closed the news page, we evaluated whether they remembered the OBA disclosure icon and tagline by asking, “Was there a symbol placed near, but not inside, at least one of the advertisements?" Only about a quarter of participants (27.6%) remembered having seen the disclosure icons, with no significant differences between the Asterisk Man and the advertising option icon. Participants were significantly more likely to remember the ads than icons ($p < 0.0005, \chi^2$). Only 11.9% of participants both said that they remembered a tagline and correctly selected the particular tagline they had seen from a list. In comparison, approximately half (49.3%) of the participants remembered the ads shown on the news webpage, with no significant differences between participants in different icon or tagline treatments.

However, the memorability of taglines did differ significantly across conditions ($p < 0.0005, \chi^2$). When we performed pairwise comparisons, we found that participants who were shown the “Why did I get this ad?” tagline remembered it at a significantly high rate than participants in all other tagline conditions except “Interest based ads.” ($p < 0.05$, HC FET). Nevertheless, “Interest based ads” was not statistically significantly more memorable than any other tagline. Tagline recall rates are summarized in Table 7.1.

7.4.4 Messages Conveyed

We again showed participants the Paris hotel advertisement with a disclosure icon and tagline, as shown in Figure 7.1. We asked the free-response question: “What, if anything, does this symbol [and phrase] communicate to you?” Participants’ opinions varied considerably by treatment. Across most treatments, the icon and tagline did not communicate effectively the concepts of notice and choice about targeted advertising. The “Why did I get this ad?” tagline was most effective at communicating notice. While some of the taglines communicated that users had choices, they did not communicate that the choices were related to OBA.

"Why did I get this ad?" Many participants who received this tagline associated it with behavioral advertising. For example, one participant explained, "It communicates that there is a logical reasoning behind the ad, most likely tracking my cookies." Similarly, another participant wrote, "This conveys that my web usage may be monitored so that the ads are tailored to my particular interests." Another common response was that this tagline was intended to explain why ads were shown on the news page. For example, one participant wrote, "The New York Times understands that people may not like ads and may be wondering why they are there."

"Learn about your ad choices" communicated three main messages: users can set preferences about what ads (if any) to see, the ads were selected based on previous browsing activity, and users can purchase advertising space.

"Configure your ad preferences" This tagline primarily suggested that a user could change the layout of the ad or set preferences regarding the types of ads he or she is interested in seeing. For example, one participant wrote, "It means you can make the ad smaller if you want," while another mentioned the "ability to control the nature of ads (i.e. static vs. animated ads)."

"Interest based ads" Many participants correctly inferred that it communicates about tailored ads. Similarly, some participants also inferred that online tracking was involved. In addition, many participants wrote that the ads displayed were exclusively for the Internet, suggesting that participants might have misread the word "Interest" as "Internet," which has been noted in prior work [**?**]. For example, one participant commented, "This advertisement is based only on the Internet. Not on a television or newspaper."

"AdChoices" Opinions about this tagline were more varied. Although many participants wrote that they had no idea about the purpose of the disclosure, a few correctly mentioned that the tagline was providing notice about ads being tailored based on previous pages visited. Other common beliefs included: it indicates that it is possible to select the types of ads you want to view, it provides a link to the ad supplier's website, and it provides a way to differentiate between web page content and advertisements. Other participants inferred that "AdChoices" was the ad's sponsor.

"Sponsor ads" Participants most commonly believed this tagline offered ad space for sale. For example, one participant expressed, "you as an individual (the symbols looks like a little person) can put your ad on this site" and another explained, "I can click on the emblem for the possibility to advertise there myself." Another common thought was that the ads were from a third party.

Symbols alone did not communicate anything related to tailored ads. The Advertising option icon alone was mostly seen as a play button with a few participants suggesting it meant "click to play advertisement" or "click to see next picture." Similarly, many of those who saw the asterisk man symbol thought it was intended to point the user to read more detailed information at the bottom or inform them about terms and conditions that might apply.

7.4.5 Communicating Notice and Choice

We evaluated the effectiveness of icons and taglines at communicating notice and choice by presenting participants with true and false statements describing the purpose of these disclosures. Participants evaluated these statements on a five-point scale ("Definitely not," "Probably not,"

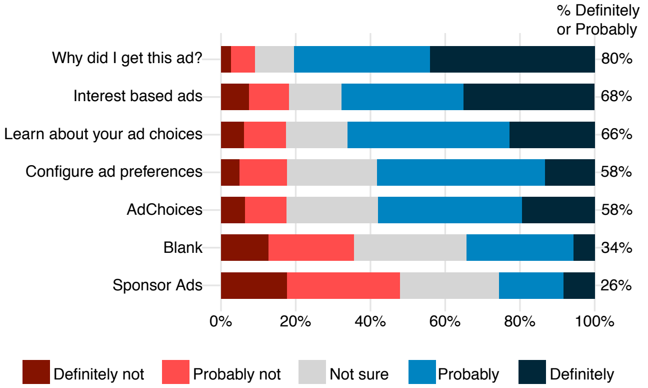


Figure 7.3: Agreement with the statement that the symbol and phrase suggest that "This ad has been tailored based on websites you have visited in the past." As shown, "Why did I get this ad?" was significantly better than most other taglines at communicating notice about OBA.

"Not sure," "Probably," or "Definitely"), which we then binned into *agreement* ("Definitely" or "Probably") and *non-agreement* (all other responses).

Communicating notice

We evaluated the degree to which different icons and taglines provided *notice* that OBA was occuring. We found that the "Why did I get this ad?" tagline performed significantly better than all other taglines, with no significant differences between icons.

Our evaluation focused on responses to the question,

"To what extent, if any, does this combination of the symbol and phrase [icon+tagline shown], placed on the top right corner of the above ad suggest the following?" Participants rated their agreement with the true statement: "This ad has been tailored based on websites you have visited in the past." Participants' agreement with all other statements are summarized in Table E.2 in the appendix. Agreement with this statement did not differ significantly between icon treatments ($p = 0.4, \chi^2$), whereas agreement differed significantly across tagline treatments ($p < 0.0005, \chi^2$); responses are summarized in Figure 7.3. "Sponsor ads" and blank treatments were least effective at communicating notice. The five other taglines performed statistically better than both the "Sponsor ads" tagline and not having a tagline ($p < 0.005$, HC FET). "Why did I get this ad?" performed the best. In particular, 80% of participants who received this tagline agreed with the statement evaluated, compared with 68% in "Interest based ads" ($p = 0.03$, HC FET), 66% in "Learn about your ad choices" ($p = 0.01$, HC FET), and 58% in both "Configure ad preferences" and "AdChoices" ($p < 0.0005$, HC FET).

Communicating choice

We also investigated the degree to which different icons and taglines communicated that participants could make a *choice* about receiving OBA. Although we found "Configure ad preferences" to be significantly better than all others at communicating choice, none of our icons or taglines was particularly successful. Our evaluation was based on the question, "What do you think would happen if you click on that symbol or that phrase?" We focused on the level of agree-

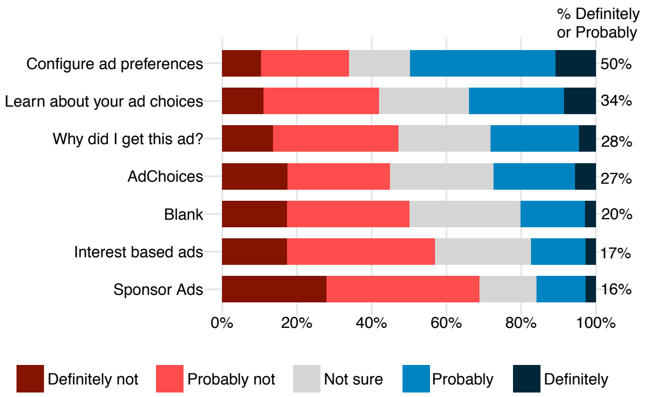


Figure 7.4: Agreement that clicking the OBA disclosures "will take you to a page where you can tell the advertising company that you do not want to receive tailored ads."

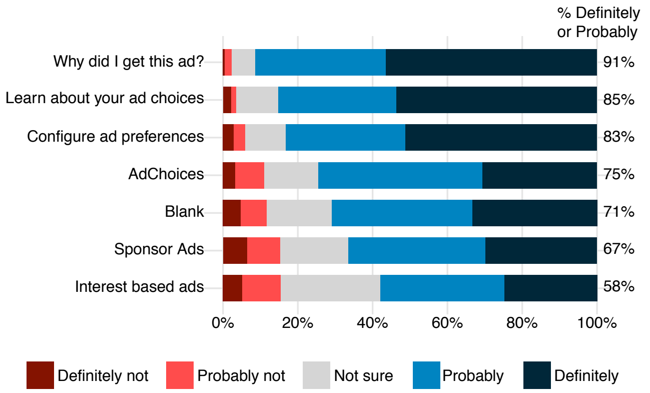


Figure 7.5: Agreement with "You can click on that symbol [and phrase]." "Why did I get this ad?" better conveyed clickability than all other taglines. Overall, those taglines containing actionable words communicated better clickability.

ment with the true statement: "It will take you to a page where you can tell the advertising company that you do not want to receive tailored ads." Participants' agreement with all other statements is summarized in Table E.3 in the appendix.

As with *notice*, the two icons did not differ significantly at communicating *choice* In contrast, tagline treatments did differ significantly ($p < 0.0005, \chi^2$). Figure 7.4 summarizes participants' responses. "Sponsor ads," "Interest based ads," and blank were least effective at communicating choice, while "Configure ad preferences" was significantly better than all other taglines ($p < 0.01$, HC FET).

7.4.6 Communicating "Clickability"

A primary mechanism for visiting network advertisers' landing pages is clicking on the icon or tagline located near ads. We evaluated the extent to which participants believe they can click on the icon and tagline, which we term "clickability." We found that clickability was fairly high in most treatment conditions, but there were significant differences between tagline and icon treatments.

We asked participants, "To what extent, if any, does this combination of the symbol and phrase [icon+tagline shown], placed on the top right corner of the above ad suggest the following?" Our analysis focuses on participants' agreement with the true statement: "You can click on that symbol [and phrase]."

Overall, participants believed the disclosures to be clickable, with 76% of participants agreeing. A larger fraction of participants given the advertising option icon (82%) agreed with the statement evaluated, compared with 69% of those given the asterisk man icon ($p < 0.0005, \chi^2$).

Taglines also differed in the clickability they conveyed. Figure 7.5 summarizes participants' levels of agreement, for which we found significant differences across tagline conditions ($p < 0.0005, \chi^2$). "Why did I get this ad?" performed the best, significantly better than the "AdChoices," "Interest based ads," "Sponsor ads," and blank treatments ($p < 0.0005$, HC FET). Differences between "Why did I get this ad?" "Learn about your ad choices," and "Configure ad preferences" were not significant.

7.4.7 Attitudes About Clicking

What participants believe will happen when they click on a disclosure is important because it may influence their willingness to click. We found that most participants had misconceptions; more than half believed that clicking on the disclosure would lead to pop-up ads or signal interest in the advertised product.

We evaluated participants' agreement with the following statements, which were provided in response to the question, "What do you think would happen if you click on that symbol or that phrase?"

> More ads will pop up. [false]
>
> You will let the advertising company know that you are interested in those products. [false]
>
> It will take you to a page where you can buy advertisements on this website. [false]

Overall, 53% of participants responded that clicking on the icon or tagline disclosure would probably or definitely trigger more ads to pop up. Figure 7.6 summarizes participants' responses by condition. A lower percentage of participants shown the asterisk man icon thought incorrectly that additional ads would pop up if they clicked on the disclosure. 50% of participants shown the asterisk man icon believed more ads would pop up, compared with 57% of those who were shown the advertising option icon ($p = 0.003, \chi^2$). There were also differences across tagline conditions. The fraction of participants who saw "Sponsor ads" who responded "probably yes" or "definitely yes" (63%) was significantly greater than the fraction who saw "Configure ad preferences" (42%) or "Why did I get this ad?" (46%) ($p < 0.02$, HC FET).

The majority of participants also mistakenly believed that clicking on the disclosure would signal to the advertising company interest in the product advertised. 51% of participants believed clicking would "let the advertising company knows that you are interested in those products," with no statistical differences across treatments.

Participants differed across tagline treatments in their level of agreement with the false statement that the OBA disclosures are intended to sell advertising space ($p < 0.0005, \chi^2$). Figure 7.7

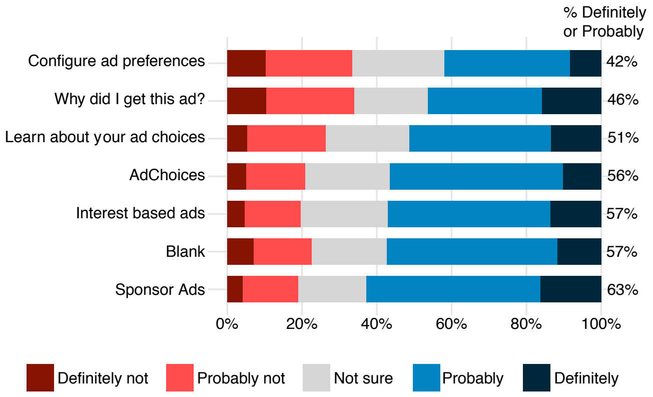


Figure 7.6: Agreement with the statement, "More ads will pop up," if they click the OBA disclosures. Overall, participants believed that clicking the disclosures would cause additional ads to pop up.

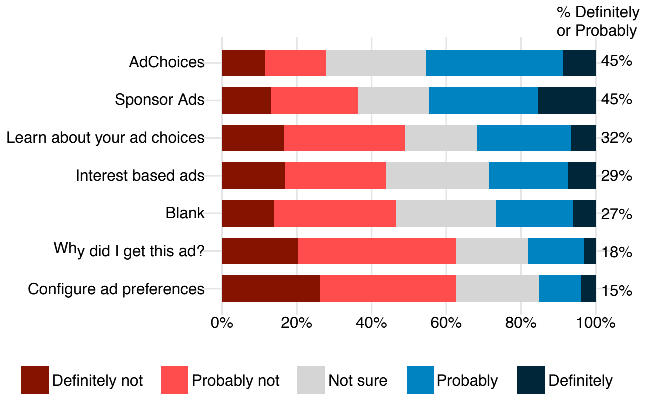


Figure 7.7: Agreement that clicking the OBA disclosures "will take you to a page where you can buy advertisements on this website." Participants in the "Configure ad preferences" and "Why did I get this ad?" treatments were less likely to believe that the disclosures aimed to sell advertising space.

summarizes participants' responses. Participants in "Configure ad preferences" and "Why did I get this ad?" were significantly less likely than those in other treatments to believe that the disclosure was intended to sell advertising space (all $p < 0.0005$, HC FET). Overall, these clickability results suggest that users have significant misconceptions about the purpose of OBA disclosures. Although 27% of participants correctly believed that clicking on the disclosure would take them to a webpage on which they could stop receiving tailored ads, larger percentages of participants believed they would receive pop-up ads (53%), signal interest in a product (51%), or learn about placing advertisements themselves (30%). Of the taglines, "Configure ad preferences" and "Why did I get this ad?" did the best job of conveying what happens when someone clicks on the disclosure.

7.4.8 Landing Pages

Landing pages, the pages that appear when a user clicks the icon or tagline disclosure, were the final element we tested. First, we report on what choices participants inferred from these pages and on user sentiment towards these pages. We then report on participants' understanding of the opt-out process after visiting the landing page.

Opinions About Landing Pages

To evaluate participants' sentiment toward the landing page they saw, we asked participants to rank the information it presented on three different dimensions: informativeness, understandability, and level of interest. Responses to these dimensions were significantly and positively correlated. The majority of participants felt the information on the landing pages was "very easy" or "easy" to understand (70%) and "very informative" or "informative" (75%), but only 41% felt it was "very interesting" or "interesting."

The Monster opt-out page performed poorly. It was seen as less understandable than each of the others ($p < 0.0005$, HC FET). Only 54% of participants believed the page was very easy or easy to understand, compared with significantly higher percentages for AOL (74%), Microsoft (74%), Google (74%) and Yahoo! (72%). Similarly, the Monster opt-out page was perceived as less informative (all $p < 0.0005$, HC FET), with 52% of participants believing the page was very informative or informative, compared with Google (83%), Yahoo! (82%), Microsoft (80%), and AOL (77%).

Notice Provided by Landing Pages

To test the extent to which a landing page conveyed notice about OBA, participants rated completions to the phrase, "To what extent, if at all, does the information on the 'landing page' suggest to you that..." We focus on agreement with the true statement,

"The ads you see in the news website are based on your visits to this news website and other websites."

Overall, 77% of participants agreed or strongly agreed, "The ads you see in the news website are based on your visits to this news website and other websites." This result suggests that opt-out pages are effective at communicating notice that OBA is occurring. In particular, 82% (Yahoo!), 79% (Google and Microsoft), 77% (AOL), and 67% (Monster) of participants agreed or strongly agreed with this statement. However, a significantly lower percentage of participants who saw the Monster landing page agreed with the statement than those who saw landing pages from Yahoo!, Google, or Microsoft ($p < 0.03$, HC FET).

The Meaning of "Opting Out"

All landing pages tested gave participants the opportunity to opt out of OBA. After visiting the landing pages, about half of participants misunderstood the meaning of opting out, either believing that it would stop online tracking or remove all advertisements. For example, one participant who visited the AOL landing page wrote, "It gives users the ability of opt out of having our data taken." Another participant who visited the Yahoo! landing page wrote, "It gives you the option to tell websites to not monitor your browsing history." Similarly, a participant who visited the Google landing page explained that the page offered "the ability to stop companies from monitoring your web activity." One participant who visited the Microsoft landing page thought the page provides "ways to advertise or ways to opt out of seeing advertisements (for a

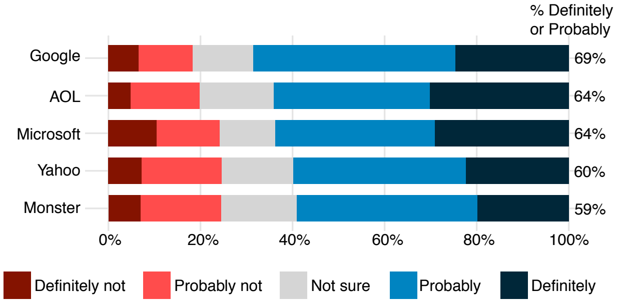


Figure 7.8: Agreement with the statement "Stop advertising companies from collecting information about your browsing activities." Most participants (63%) believed that by opting out they could stop online tracking.

fee)." Another participant thought the Microsoft landing page allowed him to decide "what ads you see or if you see any at all."

In addition, some participants expressed mistrust about the opt-out process. For example, a participant who visited the Microsoft landing page complained, "This is really hardly a choice at all since nothing stops them from continuing to gather the information." A participant who visited the AOL landing page felt it contained "information to cover the company's butt for taking my info."

To further validate these anecdotal results, we asked participants to "indicate your agreement with the following statements defining what 'opt out' means in the context of internet advertising."

- Stop advertising companies from collecting information about your browsing activities. [false]
- Stop seeing ads based on your browsing activities. [true]

Overall, 63% of participants agreed that opting out would stop advertising companies from collecting information about browsing activities, and 80% believed they would stop seeing advertisements based on their browsing activities. Figure 7.8 summarizes participants' agreement with the first statement, showing that, independently of the landing pages seen, participants understood that by opting out they could stop online tracking.

Only 13.4% of participants chose the correct answers for both questions. In contrast, the majority of participants (57.9%) incorrectly believed that opting out would stop both tailored ads and online tracking.

7.5 Discussion

Our investigation of OBA disclosures informed our understanding of what the different icons, taglines, and landing pages communicate to Internet users. While some disclosures stood out as being more effective at communicating notice and choice, we found that none of these disclosures are currently communicating clearly to consumers. In this section we discuss our main findings and suggest ways to make OBA disclosures more effective.

Notices are not noticed. One challenge of informing users about OBA through icons and taglines placed on ads is that most users do not notice them. After viewing the news webpage with ads that included our icon and tagline treatments, half of the participants correctly remembered the ads shown, but only a quarter of participants remembered the icons and fewer than 12% of participants recognized the correct taglines. While design improvements might lead to more people noticing OBA disclosures, it seems unlikely that small icons and taglines would be widely noticed on a page full of content and ads, especially when users are focusing on the content of the page. Salient links to user-friendly privacy polices with explicit information about OBA practices on the site being visited should serve as an alternative means of providing notice and choice about OBA.

"AdChoices" is ineffective. "AdChoices" is one of the official DAA taglines, as well as the one that has been observed in use by the most advertising companies [?]. However, we found that other taglines provide more effective notice, including "Why did I get this ad?" and "Learn about your ad choices." Although it contains the word "choices," it was not particularly effective at communicating that users could make choices about receiving OBA. "AdChoices" performed similarly to our control tagline, "Sponsor ads," with 45% of participants believing that the purpose of these two taglines was to communicate the availability of advertising space for sale. We suggest avoiding the use of meaningless phrases or contractions, which might be perceived by users more as a brand than as something informing them about OBA. "Configure ad preferences" and "Learn about your ad choices," which contain action words, were most effective at communicating that users have a choice to make. Further, it may be worth investigating additional taglines that pose questions or contain words like "privacy."

Users are afraid to click. The most effective taglines, "Why did I get this ad?" and "Learn about your ad choices," performed reasonably well at providing notice and were perceived as clickable, yet were ineffective at communicating that participants could use them to exercise choices about OBA. In particular, more than half of participants believed clicking on the icon or phrase would trigger pop-up ads, and a similar fraction believed that clicking them would signal interest in the advertised product. These misconceptions may be due to beliefs that the icon was part of the ad. Furthermore, a third of participants believed the disclosure was intended for selling advertising space. If users do not understand the purpose of clicking on the icon, it is unlikely that many users will click on it. Consumer education campaigns might be helpful to educate users about the purpose of these disclosures. In addition, the use of tooltips and callouts might help convey information to users who otherwise would be afraid to click on the icon.

Users are confused about the meaning of opt out. After reading the landing page, participants were unable to understand the meaning of opting out. Two-thirds of participants believed that opting out would stop online tracking. Effective and transparent disclosures should clearly communicate users' options for managing OBA. The distinction between opting out of tailored ads and opting out of online tracking should be clearly stated to avoid misleading users, or opt-outs should be made to match user expectations.

User education is needed. Arguably, the main challenge to the effectiveness of OBA disclosures is that users do not understand OBA and are unaware that disclosures link to choice mechanisms. Although user education is part of the self-regulatory principles for OBA, little user education has been done to date. The online advertising industry is currently providing consumer education about OBA through an industry website,[4] but this website is mainly accessed through the OBA disclosures that are currently not being noticed. In January 2012, the DAA launched the "Your AdChoices" campaign.[5] However, we have seen little evidence of this campaign beyond the campaign website and industry press releases.

[4] `http://www.aboutads.info/`

[5] `http://www.youradchoices.com/`

Chapter 8

Factors That Affect Users' Willingness to Share Information with Online Advertisers

Much of the debate surrounding online behavioral advertising (OBA) has centered on how to provide users with notice and choice. To better inform the design of OBA notice and choice methods, we conducted two large-scale online studies investigating what factors are relevant for users to make privacy decisions regarding OBA. We measured how different facets of ad companies' current privacy practices such as data retention, scope of collection and use, and access to collected data affected participants' willingness to share different types of information with online advertisers. We asked participants to visit a health website and a news website in studies one and two, respectively. After visiting the website, we explained OBA to them, and outlined policies outlining scenarios with varying data practices. These policies varied by condition. In study one we followed a quantitative analysis and identified classes of information that most participants would not share, as well as classes that nearly half of participants would share. More restrictive data-retention and scope-of-use policies increased participants' willingness to allow data collection. In contrast, whether the data was collected on a well-known site and whether users could review and modify their data had minimal impact. In study two we confirmed that participants' willingness to share information is not only based on the sensitivity of the information, but also on the scope of collection and use. However, qualitative analysis allowed us to gather insights regarding participants' decision making process. Participants considered the perceived necessity of collection, and perceived benefits or harms of disclosing specific data types. Participants were particularly adverse to sharing information that they perceived as irrelevant for advertising or personal. However, our results also reveal that, under the right circumstances, participants may be willing to share their data with advertisers to enhance the utility of shown ads. We discuss public policy implications and improvements for user-interfaces to align with users' privacy preferences.

This chapter is based on "What Matters to Users?: Factors That Affect Users' Willingness to Share Information with Online Advertisers" [174].

8.1 Introduction

Online behavioral advertising (OBA), the practice of targeting online advertising based on users' past online activities, has been the subject of a major privacy debate in recent years. Reports released in 2012 by the U.S. Federal Trade Commission [5] and the White House [1] discuss the privacy tradeoffs inherent in this practice. At the same time, browser vendors have recently taken steps to reduce tracking: Microsoft sends a Do Not Track signal by default in IE 10 [175], and Mozilla has announced that Firefox will eventually block third-party cookies by default [176].

As battles rage about default behaviors and options, average users are asked to make choices about their privacy preferences regarding online behavioral advertising. In some cases, these choices have limited granularity. For instance, with the Do Not Track signal under debate [177], users have the choice of actively turning Do Not Track on or off, or leaving it unset. In many other cases, however, users have a more complex decision to make. As part of the advertising industry's self-regulation program, users can opt out of behavioral advertising from individual companies [95]. Similarly, third-party privacy tools like Abine's DoNotTrackMe[1] and Evidon's Ghostery[2] enable users to see which companies are tracking their activities on a particular site, and to block particular companies. In Chapter 5 we found that familiarity with a third-party tracking company influences users' attitudes about data collection.

However, little is known about other factors that may influence users' preferences. For instance, does the length of time behavioral data is retained actually matter to most users? Does it make a difference whether data is used to target advertisements only on a single first-party website, or on Facebook, or on any website on the Internet? Does it matter if the purpose of collection may be other than advertising? This understanding is crucial for the design of future OBA privacy tools. For instance, when a privacy tool asks the user to decide whether to permit or block the collection of data by a particular entity, the tool could highlight that entity's privacy practices that most strongly affect users' decisions. Better understanding the drivers of user behavior might also influence public policy. For instance, laws and regulations designed to support consumer privacy could focus on practices that most affect users' comfort with data collection and sharing, rather than focusing on distinctions that have little bearing on users' preferences.

In study one, we examined how four dimensions of privacy practices impact users' willingness to permit the collection of data for OBA. These dimensions are the length of time data will be retained, whether or not a user will have access to review and modify this data, the range of websites on which advertising will be targeted based on this data, and whether the data was collected on a well-known website.

To this end, we conducted a 2,912-participant online survey. We asked participants to visit a health website. After they explored this page, we explained the value proposition of online behavioral advertising: that advertising and the collection of data for targeted ads enable websites to be free. We then showed the participant this website's data-collection practices, with details varied based on the participant's condition. In different conditions, participants were told that data would be retained for one day or indefinitely; they were told or not told that they would

[1] `https://www.abine.com`

[2] `http://www.ghostery.com`

be provided access to review and modify collected data; participants were told that data would be used for targeted advertising only on the health site, on both the health site and Facebook, or on any website; and the health site itself was either well known or a site we invented. We then asked participants to rate their willingness to allow the collection of 30 different types of information, and to answer additional questions related to their OBA preferences.

Nearly half of our participants were unwilling to allow the collection of any data, while the site's privacy practices impacted the remaining participants' attitudes. Of the four dimensions we examined, the scope of use and the period of data retention had the greatest impact on participants' willingness to allow their information to be collected. Having access to view and modify data collected, as well as participants' familiarity with the website on which data was being collected, did not appear to affect their willingness to allow data collection, at least in the narrow scenario we investigated.

In study two, in addition to investigating the impact of scope of collection and use, we investigated more granular retention periods (one week, three months, and one year) in the context of a news website. We further explored why users were willing or unwilling to share their data and participants' opinions of advertising profiles.

1,882 participants took part in study two. We found that participants' (un)willingness to share information with online advertisers was affected by the context and purpose of collection, as well as participants' general attitudes towards targeted ads and privacy. Participants were particularly adverse to sharing information that they perceived as irrelevant to advertising, such as income range. Participants believed that their online activities did not reflect their purchasing interests. Many participants were indeed willing to share information with advertisers, such as their actual interests, or even correct collected data in order to improve the relevance of the ads shown to them.

Our results suggest a need for better transparency regarding ad companies' practices and more granular OBA control mechanisms.

We provide background on the debate surrounding OBA and highlight related work in Section 8.2. Then we discuss the methodology of study one and its results in Sections 8.3 and 8.4, respectively. In Sections 8.5 and 8.6, we discuss the methodology and results of study two, respectively. We discuss our overall results in Section 8.7.

8.2 Background and Related Work

We first discuss current OBA practices and then users' privacy expectations.

8.2.1 OBA practices

In OBA, third-party advertisers track users as they browse websites. The purpose of this tracking is to build profiles of users in order to target ads. Tracking can be performed using third-party cookies or more complex techniques [127].

Third-party tracking for OBA is widespread. In 2011, third-party trackers were present on 79% of pages examined among the Alexa top 500 websites [178]. Among a set of selected U.S. and Canadian health websites, 85% contained at least one tracker [179].

Online social networks also track user data and leak data in potentially privacy-invasive ways. In a study of twelve online social networks, Krishamurthy and Wills found that the sites tended to leak unique identifiers to third parties, allowing users to be linked to one or more social networking profiles. They also found that some websites directly leaked personally identifiable information [180]. Roosendaal found that Facebook tracked both users and non-users of Facebook across the Internet using cookies attached to "Like" buttons embedded in other pages [181].

This large data footprint leads to privacy concerns. Retailers can combine credit or debit card histories with data from online tracking to create detailed customer profiles revealing potentially sensitive "lifestyle or medical issue[s]" [100]. Even when data is collected in an aggregated, ostensibly anonymized manner, bulk collection leaves the potential for re-identification [182, 183].

Advertisers may use the collected data for the purpose of behavioral advertising, and also for other purposes, such as website analytics, or marketing research. The data may be retained for varying periods of time. For example, Google may retain information for an undisclosed period of time,[3] and Lotame, a data aggregation company, may retain information for up to nine months.[4] In order to increase transparency, some advertisers allow users to access profiles created about them, e.g., with BlueKai Registry[5] or Google Ad Settings.[6]

OBA is a significant contributor to advertising revenue, due to higher prices and click-through rates compared to non-targeted ads [119].

However, the effectiveness of OBA is nevertheless questionable. Farahat and Bailey found that when pre-existing consumer interest is considered, sophisticated targeted advertising may not benefit the advertiser [122]. Lambrecht and Tucker analyzed data from a travel website and found that general-audience ads perform better on average than targeted ads, and that personalized ads are only effective when users have already developed strong preferences on the subject [184]. Similarly, Tucker examined social advertising on Facebook and found that users responded more positively to social ads targeted using the standard Facebook algorithm than to ads referencing their social connections [185].

8.2.2 Users' privacy concerns

Many studies have found that users are generally concerned about OBA. Many do not want third parties to track and profile them online [136, 186].

Awad and Krishnan [187] found that users who valued transparency of information were more concerned about online profiling than those who did not (n=401). Wills and Zelijkovic created a Javascript tool to display user information that could be collected by third-party sites. They found in their study (n=1,800) that half of their participants were concerned about third-party tracking, data collection, and trackers' ability to guess demographic information [188]. User understanding of OBA is poor, despite its prevalence. In a 2011 survey, participants thought websites collected more information than was possible [189]. Perceptions of shown ads further affect users' attitudes towards online advertising. In a survey (n=266), Cho and Cheon found that participants avoided online ads because of the overall number of ads, previous negative experiences with online ads, and believed that the ads were contrary to their browsing

[3] `http://www.google.com/policies/privacy/`

[4] `http://www.lotame.com/legal`

[5] `http://bluekai.com/registry/`

[6] `http://www.google.com/settings/ads`

goals [190]. Agarwal et al. found (n=53) that users were particularly sensitive to being shown embarrassing ads as a result of OBA [115].

Although factors influencing OBA decision-making have not been well studied, a handful of researchers have examined factors that impact information sharing more broadly. In a 1998 study of 401 participants, Awad and Krishnan found that users who valued information transparency were more concerned about being profiled online than those who did not [187]. Across two studies, Acquisti et al. found that the context of an information request affected users' willingness to share information. If more sensitive information was requested prior to less sensitive information, participants were more likely to reveal more information overall [191]. Taylor et al. found that general online trust made users less concerned about privacy [192]. Joinson et al. asked distance-learning students to sign up for a panel that requested a variety of sensitive personal information. They found that participants were least willing to share financial information [193].

Two studies have examined information sharing with online music sites. When asked to provide information to a mock online music retailer, Metzger found that participants were more likely to disclose information if they saw a strong privacy policy than if they saw a weak privacy policy. She also found that participants were most likely to be willing to provide information necessary for a retail transaction, specifically name and address, as well as basic demographic information. Participants were least likely to be willing to provide financial information [194]. In a study of how participants felt about revealing information to a music recommender system, van de Garde-Perik et al. found that some participants wanted to reveal information anonymously because of privacy concerns, while other participants were willing to reveal information tied to their identities to help improve the system. In both cases, the researchers found that participants wanted to know how the data would be used and who would have access to it [195].

In this work, we extend and complement the body of knowledge about user sentiment toward OBA by assessing and contrasting the effect of different data collection and use practices on participants' willingness and comfort to share certain information. We complement our quantitative findings with extensive qualitative analysis of participants' reasons behind their preferences. We further studied participants' opinions about the content and perceived benefits of behavioral advertising profiles.

8.3 Methodology of Study One

We conducted a between-subjects online study to investigate how online advertising companies' privacy practices impact users' willingness to allow the collection of information for OBA. Participants completed an online survey in which they were asked to visit a health website, were given notice about privacy practices that governed data collection for OBA on the site they visited, and answered a series of questions about their willingness to allow different types of personal information to be collected. Each participant was assigned to a condition that specified the exact privacy practices that would be presented to him or her. Participants answered additional questions investigating their attitudes toward OBA and online privacy.

In this section, we discuss participant recruitment, the conditions to which participants were assigned, and the design of the survey. We then provide an overview of our analysis methods.

8.3.1 Recruitment

We recruited our participants using Amazon's Mechanical Turk crowdsourcing service.[7] Recruitment materials indicated that the study would be about how individuals experience the Internet. They provided no indication that either OBA or privacy would be major components of the study. We required that participants live in the United States and be age 18 or over. All participants who completed the study were paid $1.00, which is typical for a task on Mechanical Turk that takes approximately twenty minutes to complete. The Carnegie Mellon University IRB approved our protocol.

8.3.2 Conditions

We assigned participants round-robin to a condition. This condition specified the privacy practices participants were told governed OBA on the website they visited. Our study's design was full-factorial across three dimensions of privacy practices. For our first dimension, we investigated three types of scope of use and sharing policies. Our second dimension varied the period for which the data collected would be retained. Finally, the third dimension investigated the impact of providing users the ability to review and modify data collected about their behavior. As our investigation was primarily exploratory, we considered only extremes; for example, data would be retained for a day or indefinitely. If diametrically opposed policies do not impact participants' attitudes, it is unlikely that gradations of these policies would.

Each participant's condition specified one of the following levels for each of these three dimensions:

- **Scope of use** (3 levels). Participants assigned the first treatment level were told that the XYZ Advertising Company would collect behavioral data only on the health website they were visiting, and that collected data would be used only to target advertisements on that website. Participants assigned the second level were told that the XYZ Advertising Company would collect behavioral data on any website on the Internet, and this data would be used for targeting ads on any website on the Internet. Those assigned the third level were told that Facebook, acting as the ad network, would collect and use data for targeting advertisements on both the health website and Facebook.

- **Data retention period** (2 levels). Participants were told either that all data collected for online behavioral advertising purposes would be retained for *one day*, or that the data would be retained *indefinitely*.

- **Level of access** (2 levels). Participants were either told the advertising company would provide "access to a webpage where you can review, edit, and delete the information that is being collected about you," or told nothing regarding data access.

We also investigated whether participants' familiarity with the health website they visited as part of the study, and on which behavioral data would be collected, would impact their willingness to allow data collection. As this investigation of familiarity with the first-party site was a secondary goal of the study, we did not include it in our full-factorial design. We assigned participants one of two levels for familiarity: Participants either visited WebMD, which is a popular

[7] `https://www.mturk.com`

health website; or they visited WebDR, a clone of WebMD that we invented and with which participants would presumably be unfamiliar.

8.3.3 Survey Flow

After reviewing and agreeing to a consent form, participants answered general questions about their impression of advertising on the Internet, exploring whether it was useful, relevant, or distracting. In order to gain a better understanding of our participants, we then asked them to answer demographic questions, as well as general questions about their use of the Internet and social networks.

In order to simulate the experience of visiting a website more closely, we instructed participants to follow a link in the survey to visit either the WebMD or WebDR website, depending on their condition. To eliminate variability caused by pages changing over time, we hosted an exact copy of the WebMD homepage as of February 5th, 2013. We disabled all hyperlinks and forms on the page so that participants would concentrate on the homepage, yet retained all other functionality on the page, such as interactive drop-down menus and scrolling news stories. The WebDR homepage was identical to WebMD's, except that all branding and logos had been changed to read "WebDR." In order to verify that participants examined the site, we asked them to identify three health conditions discussed on the site's homepage. To gauge whether WebMD was actually a familiar brand to participants, while WebDR was not, we asked questions about participants' history of visiting either WebMD or WebDR, as well as their impressions of the site's reputation and trustworthiness.

We next presented participants with a description of OBA, along with its value proposition. We explained that websites "are able to offer free services to their visitors by contracting with online advertising companies. The advertising companies pay websites for every ad they show, allowing the websites to provide free services for users like you."

Participants were told to imagine they were "experiencing a flaky scalp condition" and therefore visiting a health website. We explained that an advertising company contracting with the health site "collects information about your interactions with the {WebMD | WebDR} website in order to predict your preferences and to show you ads that are most likely to be of interest to you. These ads are known as targeted ads." Following this description, we presented participants with the details of the privacy practices governing online behavioral advertising according to their condition. They were immediately asked questions testing their comprehension of the privacy practices specified (e.g., "Based on the information that you just read, for how long may *company* use the information collected about you?"). Data from participants who answered any of these comprehension questions incorrectly were removed from our analysis.

We next asked participants to "answer the questions below indicating what information you would allow {XYZ Advertising Company | Facebook} to collect for the purpose of showing you targeted ads on {the WebMD website | the WebMD website and other websites that you visit | your Facebook page and the WebMD website}." We then asked about the 30 items of *personal information* shown in Table F.1 in the appendix. To facilitate participant comprehension, we organized these 30 data items into five categories: computer-related information, demographic and preference information, interactions with the website, location information, and personally identifiable information (PII), which we referred to only as "information" in the survey. For each item, participants rated their agreement with the statement "I would be willing to allow *company* to use and store the following information related to my interactions with the *name* website" on a five-point Likert scale ("Strongly Disagree" to "Strongly Agree"). The data collected during this part of the study are used for the bulk of our analyses.

We then asked a number of additional questions related to privacy and online behavioral advertising. For instance, we asked participants about their willingness to share for different data-retention periods, whether participants might be willing to pay money for stopping data collection or advertising, and how they felt about online behavioral advertising on a number of different types of websites. We also presented participants with six features a hypothetical browser plugin might have that could help users understand or control data collection. We also asked whether the presence of each feature would increase their willingness to allow advertisers to collect their personal information. The final page of the survey asked participants about their general privacy attitudes and whether they had taken privacy steps, such as opting out of OBA or enabling Do Not Track in their web browser.

8.3.4 Analysis

We were interested in understanding how the practices participants were told governed data collection impacted their willingness to share the 30 types of information we asked about. As shown in Table F.1 in the appendix, we examined both sensitive and non-sensitive information. The Network Advertising Initiative (NAI) requires opt in or "robust notice" for some, but not all, of the sensitive items we studied [95]. To reduce these 30 types of information to a smaller number of output variables, we performed exploratory factor analysis. Factor analysis reveals underlying associations by evaluating which variables are closely related, combining variables that are highly correlated into a single latent factor. If such underlying factors are observed, further analysis considers these factors in place of the individual variables.

We performed exploratory factor analyses and found that 22 data types were grouped into five factors, while 8 data types did not conform to any particular factor. These five groups closely mirrored the categories from our original survey. We used the standard procedure of considering a factor part of a group if it had a factor loading of at least 0.6 for the particular group, as well as factor loadings under 0.4 for all other groups. In Section 8.4.1, we discuss the results of this process, including which types of information were grouped or excluded. We further verified our groupings by calculating Cronbach's alpha for each group, using the standard value of 0.8 or higher to indicate good reliability.

Our further analyses focus on these five resultant factors. We created an index variable for each of the five factors by averaging participants' responses to the question items included in each factor. Using the participant's treatment for each dimension of privacy policy as independent variables and the five factors' indices as dependent variables, we performed a multivariate multiple regression, evaluating the effect of multiple independent variables on multiple dependent variables. Our model considered covariates including age, gender, and privacy attitudes, as well as interactions between independent variables. We confirmed our results by running repeated measures ANCOVA and MANOVA, which yielded similar results. For all statistical tests, $\alpha = 0.05$.

8.4 Results of Study One

We analyzed responses from 2,912 participants between the ages of 18 and 74 (mean $= 31$, $\sigma = 11.1$). Almost half (47%) of the participants were female. Most of them (97%) indicated using the internet everyday for at least one hour. Participants exhibited a diverse range of occupations, and were well educated (90% at least some college, 45% Bachelor's or Graduate degrees). We did not observe any statistical differences in the education, technical background, gender, age, or Internet-usage patterns of participants assigned to different conditions.

Around half of our participants were unwilling to disclose any personal information in exchange for targeted ads. The remaining participants were willing to disclose their gender, low-granularity location, operating system, and web pages they had visited at a higher rate than other types of personal information. We found the type of information collected, the scope of use of the information, and the retention period impacted participants' willingness to disclose information.

We first describe the results of our exploratory factor analysis that used participants' responses to group different types of information (Section 8.4.1). We then identify which factors affected participants' willingness to disclose different types of information (Section 8.4.2). In Section 8.4.3, we discuss participants' attitudes toward targeted ads in different first-party browsing contexts. We then discuss our qualitative results investigating participants' willingness to pay to remove ads and stop data collection (Section 8.4.4). Finally, in Section 8.4.5, we discuss the impact of mechanisms for controlling data collection on participants' disclosure preferences.

8.4.1 Factor Analysis

Our exploratory factor analysis created five groups that included 22 of the 30 types of information. Table 8.1 lists these groups by the names we gave them, as well as the types of information in each group. We provide greater detail about the factor loadings for each type of information, as well as how we created these groups, in Appendix F.1.

Throughout the remainder of this paper, we refer to these five groups by name: *browsing* information, *computer* information, *demographic* information, *location* information, and *personally identifiable* information. The remaining 8 types of information were not associated with any of the five groups and are excluded from our regression.

To verify that the types of information in each group were highly correlated, we calculated Cronbach's Alpha for each group. Our results for this correlation analysis supported the groups from factor analysis. Alpha values of 0.8 or higher are considered to be correlated, and our five groups had overall alpha values ranging from 0.81 to 0.94.

8.4.2 Willingness to Disclose Information

Nearly half of our participants were not willing to disclose information for the purpose of receiving targeted ads. The remaining participants distinguished between the types of information they would disclose, as shown in Figure 8.1. For instance, 45% of participants were willing to disclose the operating system they used, while under 1% were willing to disclose their Social Security number or credit card number.

Browsing information ($\alpha = 0.92$)
Medications taken (inferred from browsing)
Pages visited
Search terms entered
Survey responses
Time spent on each page
Computer information ($\alpha = 0.93$)
Operating system
Web browser version
Demographic information ($\alpha = 0.94$)
Age
Gender
Highest level of education
Hobbies
Income bracket
Marital status
Political views
Religion
Sexual orientation
Location information ($\alpha = 0.91$)
Country
State
Town/City
ZIP code
Personally identifiable information ($\alpha = 0.81$)
Email address
Name

Table 8.1: The five factor groups that resulted from factor analysis, comprising 22 of the 30 types of information from the survey.

The data-retention period and scope of use significantly impacted participants' willingness to disclose the types of information for which participants had varied responses. Providing the opportunity to access and edit information that had been collected, as well as familiarity with the website on which data was collected, had minimal impact. A participant's level of privacy concern, frequency of Facebook usage, age, and positive opinions about targeted ads also impacted their willingness to disclose information.

Impact of Type of Information

Participants were willing to disclose different types of information at vastly different rates. Unsurprisingly, most participants strongly objected to the collection of personally identifiable information (PII), and these attitudes did not vary significantly by condition. For example, across all treatments, under 3% of participants would disclose their phone number. On the other extreme, participants were most willing to disclose arguably innocuous information, such as their country (53%) and gender (46%). Between these two extremes were types of information for which users' willingness to disclose was affected by the scope of use of the information, and for how long it would be retained.

Figure 8.1 summarizes participants' responses across all conditions to the 30 different types of information in our survey. While participants' willingness to disclose many types of information differed significantly by condition, participants had relatively homogeneous answers for the most and least sensitive types of information. Very few participants were willing to disclose

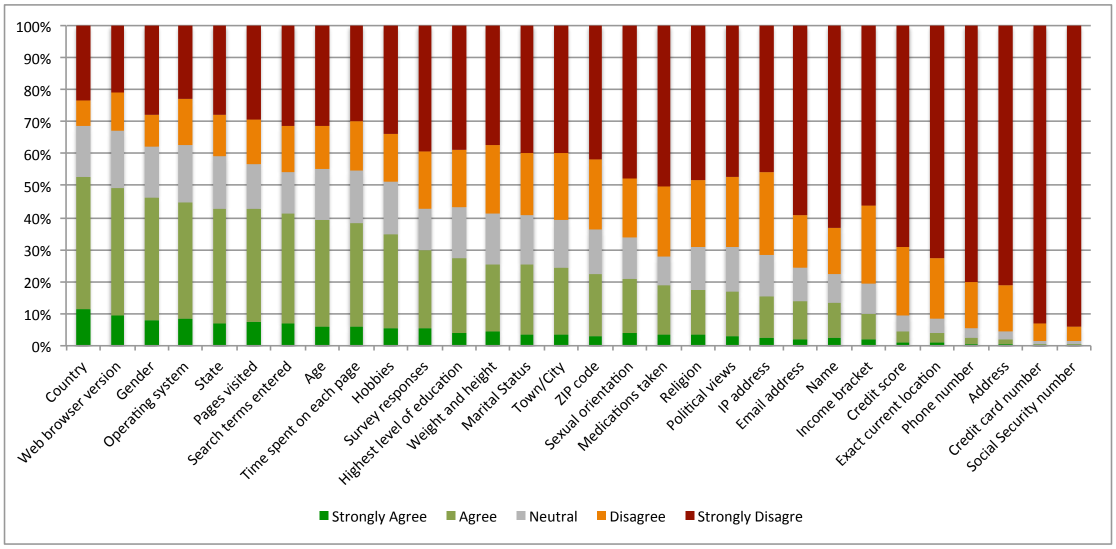


Figure 8.1: Participants' responses to the statement, "I would be willing to allow *advertising company* to use and store" 30 different types of information. The two shades of green represent willingness to share, while the two shades of red indicate unwillingness to do so.

sensitive information. For instance, only a handful of participants were willing to disclose their SSN (<1%), credit card number (<1%), address (2%), phone number (3%), exact current location (4%), and credit score (5%). We did not observe significant differences across conditions for these types of information. Participants' unwillingness to disclose these types of information is particularly notable in light of Krishnamurthy et al.'s finding that a majority of popular websites actually leak some types of sensitive information to advertising companies [196].

In contrast, nearly half of our participants were willing to disclose less sensitive information. Many participants were willing to disclose their web browser version (43%), operating system (45%), and gender (46%). Participants were similarly willing to disclose coarse-grained information about their location, such as the state (43%) and country (53%) from which they were visiting the health website. These results also did not vary significantly by condition.

Impact of Retention Period

The data-retention period significantly impacted participants' willingness to disclose various types of information for three of the five groups of information, as shown in Figure 8.3a. In particular, participants who were told that data would be retained only for one day were significantly more willing to disclose browsing information ($p < .001$), demographic information ($p = .025$), and location information ($p = .001$) than those told data would be retained indefinitely. We did not observe significant differences for computer information or personally identifiable information.

We next asked participants, "How would your willingness to allow {XYZ Advertising Company / Facebook} to collect your information change if it retained your information" for periods ranging from the duration of a browsing session to indefinitely. These results, shown in Figure 8.2, further suggest that the data-retention period impacts preferences. In particular, 39%

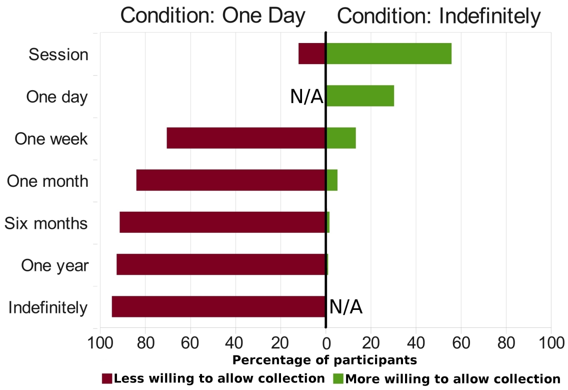
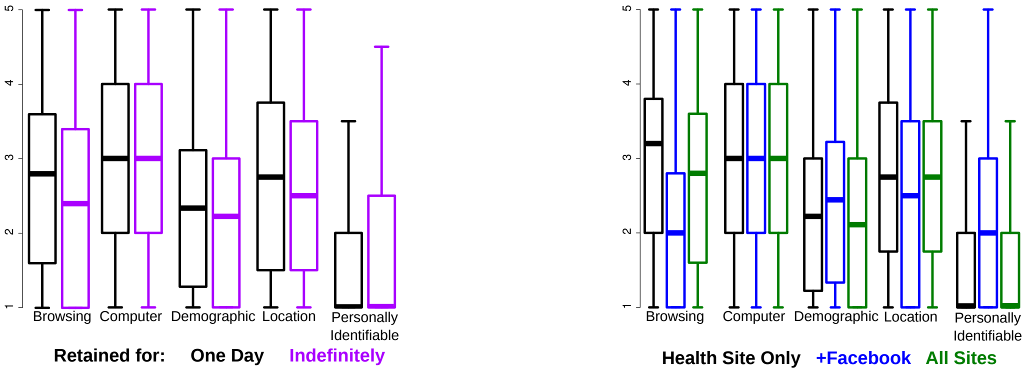


Figure 8.2: The percentage of participants originally told that data would be retained for one day who would be less willing to allow data collection for different retention periods, as well as the percentage originally told data would be retained indefinitely who would be more willing for different periods.

of participants in the one-day treatment and 56% of participants in the indefinite treatment indicated that they would be more willing to disclose if their information were retained for the duration of their browsing session. On the other hand, participants were considerably less likely to disclose information for periods greater than one week. Further research is needed to determine whether a data-retention period longer than the duration of a browsing session would align with Internet users' preferences.

Impact of Scope of Use



a) Effect of retention period

b) Effect of scope-of-use

Figure 8.3: A comparison of participants' willingness to share the five groups of information given different retention and scope-of-use policies. The y axis represents participants' willingness to share (5=strongly agree, 1=strongly disagree), averaged over all types of information in that group. Differences between the one-day and indefinite treatments are significant for browsing, demographic, and location information. Differences between the health-site and Facebook treatments are significant for browsing, location, and personally identifiable information. Differences between the health-site and all-site treatments are significant for browsing information

How collected information would be used outside the first-party site also impacted participants' willingness to disclose information, as shown in Figure 8.3b. Based on their condition, participants were told that XYZ Advertising would collect and use their data only on the health website, that XYZ Advertising would collect and use their data on any website on the Internet, or that Facebook would collect and use their data on both the health website and Facebook.

Participants in the Facebook treatment were significantly less willing to disclose browsing information ($p < .001$) and location information ($p < .001$), than participants told that XYZ Advertising would collect and use their information only on the health website. In contrast, participants in the Facebook scenario were significantly more willing to disclose personally identifiable information ($p < .001$). As users share personal information on Facebook, this result might be explained by the contextual nature of privacy [197].

In addition, participants told that XYZ Advertising would collect and use their information on any website on the Internet were significantly less willing to disclose browsing information ($p < .001$) than participants told that information would only be collected and used on the health website. We did not observe significant differences across conditions for computer or demographic information.

Impact of Access to Collected Data

Access to the collected information had a more moderate impact than data-retention and scope-of-use policies. In particular, we did not observe significant differences between participants told they could review and edit the information collected and those not told about this opportunity for any of the five groups of information.

A number of factors might explain this lack of an effect. The concept of "access" to data collected by third parties (e.g., advertising networks) might have sounded strange or vague to participants. Additionally, reviewing and editing collected information represents a cost that may outweigh the expected benefit. The concept of "access" was also abstract in that we did not specify what participants might find on the hypothetical page where they could review the data that had been collected.

Later in the survey, we also asked questions about access to participants whose condition dictated they not be told originally about having access to the information collected. In particular, we asked these participants whether they would be more or less willing to share information if they were able to view and edit it after it was collected. Of these participants, 48% responded they would be more willing to disclose information if given access, 41% responded they would be equally willing, and 11% responded they would be less willing. Companies including Google, Microsoft, and Yahoo! currently provide users access to information through privacy dashboards. Nevertheless, very little is known about how people use these dashboards. More research is needed to understand at a deeper level how access impacts users' privacy decision making.

Impact of Site Familiarity

Our manipulation for site familiarity, having participants visit either the real WebMD or fictional WebDR, appeared to work as intended. While 81% of participants who visited WebMD felt the website was trustworthy, only 62% of those who visited WebDR felt the same ($p < 0.001$, χ^2). Similarly, 73% of participants who visited WebMD said they were familiar with the website, compared with 20% of those who visited WebDR ($p < 0.001$, χ^2), and 82% of participants believed that WebMD had a good reputation, compared with 39% of WebDR visitors ($p < 0.001$, χ^2).

However, whether the participant visited WebMD or Web-DR did not significantly affect the participant's willingness to disclose information. This result suggests that participants' opinions were mostly based on the third party collecting the data, rather than the first-party site. Although we must be careful about extrapolating this result, participants' willingness to disclose information in other first-party contexts may not be drastically different.

Other Factors Impacting Disclosure

We found that aspects other than the website's privacy practices and type of information collected also impacted participants' willingness to disclose information. Participants with higher privacy concerns, identified through survey questions about privacy attitudes (Question 46 in Appendix F.3), were less willing to disclose all five types of information (all $p < .001$). In contrast, participants who expressed positive opinions toward targeted ads were significantly more willing to do so for all five groups of information (all $p < .001$). Participants who used Facebook more often were also more willing to share all five groups of information (all $p < .001$).

For certain types of information, we observed other significant covariates. Older participants were less willing to share demographic information ($p = .026$) and more willing to share location information ($p < .001$). A participant's stated background in technology, such as holding a degree or job in IT, was a significant covariate for demographic information ($p = .021$). We also observed a significant interaction effect between the Facebook scope-of-use scenario and indefinite data retention for browsing information ($p < .001$), demographic information ($p = .045$), and personally identifiable information ($p = .043$).

8.4.3 Site Context

As privacy attitudes depend on context [197], we also investigated how participants would feel about the type of website (e.g., banking website, travel website) on which data was collected. In particular, for nine categories of sites, we asked participants to rate on a five-point Likert scale their agreement or disagreement with the statement, "I am interested in receiving targeted ads on the websites that I visit based on my online activities on *category* sites."

Participants' willingness to have data collected and used for OBA purposes differed across the category of site. Figure 8.4 presents detailed results. More than half of our participants would not be willing to permit data collection on any of the nine categories of sites we presented. Participants were most willing to allow data collection on arts and entertainment websites (40% of participants), travel websites (34%), and news websites (32%). Only around 8% of participants would be willing to have their actions on dating or online banking sites used for targeting ads, and only 15% of participants felt the same for photo-sharing websites.

Although health information has been classified as sensitive by both the advertising industry [95] and government regulators [?], 25% of participants were willing to have data from health sites used for OBA purposes. On the one hand, this result might reflect bias in that participants had just answered questions about the collection of personal information on a health website. On the other hand, since a health website formed the basis for the scenario in our study, this result might suggest a baseline for how participants' willingness might have been different had the scenario taken place on another type of website.

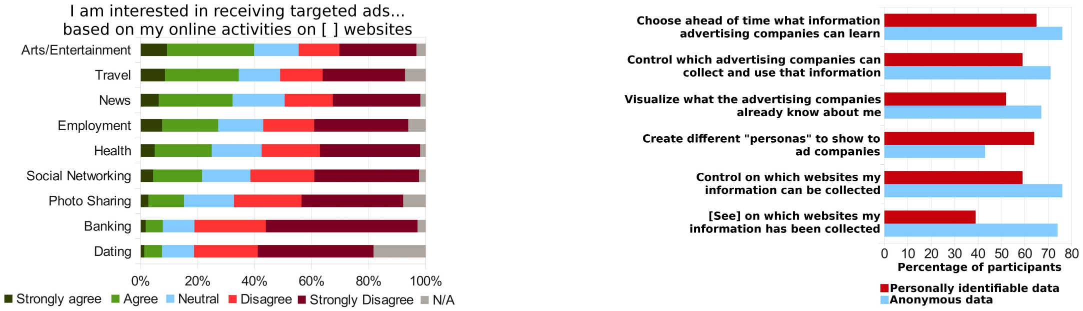
8.4.4 Willingness to Pay

A user's willingness to pay for a feature can be used as a proxy for how much the user values that feature. We asked three questions about participants' willingness to pay a monthly fee for different advertising and data-collection scenarios. These scenarios were "not showing you any ads," "not showing you targeted ads, but only generic ads," and "stopping collection of any information about you or your online activities." Items 29, 31, and 33 in Appendix F.3 show the specific questions that we asked.

We found that the majority of participants were not willing to pay anything for these changes. Across all conditions, 62% of participants would not pay to stop data collection, 69% would not pay to remove ads, and 80% would not pay to see generic ads in place of targeted ads. Participants cited several reasons for not being willing to pay. They commonly felt they could obtain the information they wanted on other websites without paying, or use free software to block ads. They also felt that websites should be free, and that privacy is a right they should not have to pay for.

Participants who were willing to pay said they would pay a median amount of $3.00 to stop data collection, $2.25 to remove all ads, and $2.00 to show generic ads in place of targeted ads. That a larger proportion of participants were willing to pay money to stop data collection than to remove ads, and that those who were willing to do so would pay more, indicate that many participants value stopping data collection more than removing ads. However, participants' low willingness overall to pay for any of these scenarios suggests that their perceptions are rooted in the belief that websites and ad-blocking tools should be free.

8.4.5 Levels of Control



a) Willingness to share for different website's contexts

b) Effect of different controls given

Figure 8.4: a) Participants' interest in having their behaviors on different types of websites used to target ads. b) Percentage of users who would be "more willing to allow collection of {anonymous | personal} information for the purpose of receiving targeted ads," if their web browser provided them six different options for control.

Current web browsers do not provide usable, fine-grained control over data collection for OBA purposes, nor over the display of targeted ads. To explore whether the introduction of new, fine-grained controls would help users feel more comfortable sharing data with advertisers, we asked six questions about a hypothetical browser plugin that would give the user control over, as well as better visibility into, the information collected by online advertising companies for the purpose of showing targeted advertisements.

For each feature, we asked users whether they would be more willing to disclose information if they were able to take advantage of a plugin with this feature. The six plugin features were "choose ahead of time what information to disclose," "control which ad companies can collect the information," "visualize and edit information already disclosed," "create different 'personas'," "control which websites [can collect information]," and "visualize which websites already collected information." Independent of their condition, half of our participants were asked about plugins for controlling anonymous information, and half about plugins for controlling personally identifiable information.

Respondents reported that a plugin with some of the six proposed features would make them more willing to share anonymous (84% of participants) and personally identifiable (74% of participants) information with advertisers. This difference—where participants asked about sharing anonymous information were more likely to share more in the presence of a plugin—was consistent across five of our six proposed plugin features. The exception was the "create different 'personas' " feature, where the trend was reversed. This result was consistent with our expectations, since the concept of a persona is less relevant to anonymous than to personally identifiable information.

Which plugin features participants thought would most increase their willingness to share differed between the set of participants who were asked about sharing anonymous information and those who were asked about sharing personally identifiable information. Perhaps unsurprisingly, participants who were asked about sharing personally identifiable information were most frequently interested in plugin features that allowed them to prevent information from being sent to advertisers in the first place (59–64% reported that they would share more). A smaller proportion would share more if they could see after the fact what information had been gathered (52%), or on what sites (39%). Most participants who were asked about sharing anonymous information reported that these features would increase their sharing.

Overall, we believe these results suggest that although participants were not willing to disclose much information online, offering more adequate control over disclosure could mitigate some of their privacy concerns.

8.5 Methodology of Study Two

To understand how advertisers' data practices influence an individual's willingness to share information for advertising purposes, we conducted a between-subjects study as an an online survey on MTurk. We recruited 1,882 participants residing in the United States and at least 18 years of age. Compensation was $1.5 and average completion time was 22 minutes. Our study was approved by CMU's IRB. Below, we discuss our study design, survey implementation, and our analysis approach.

	Scenario	Description
S1	AllNews	Best Ads may collect information and show targeted ads only on the AllNews website.
S2	OtherPurposes	S1 + Best Ads may use the collected information for other purposes.
S3	Websites	Best Ads may collect information and show targeted ads on the AllNews and other websites.
S4	Offline	S3 + Best Ads may collect information from a local department store and give targeted coupons for the store.
S5	Websites&OtherPurposes-NoShare	S3 + Best Ads may use collected information for other purposes, but not share it with other parties.
S6	Websites&OtherPurposes	S3 + Best Ads may use collected information for other purposes, no restrictions given.
S7	Facebook	Facebook may collect information on the AllNews website and users' Facebook page to show targeted ads on Facebook.

Table 8.2: Scenarios of the different conditions, each was tested with retention periods of one week, three months, and one year (21 conditions in total).

8.5.1 Study Design

Our goal was to understand how advertisers' data practices would impact users' willingness to share if users were aware of such practices. (1) Does the type of data collected matter? For example, does willingness vary by personal, financial or predictive information? (2) Do variations in scope of collection and sharing affect users' willingness to share information? For example, are users more comfortable when their data is collected and used on a single website versus on multiple websites? Similarly, does a shorter retention period make them more comfortable? (3) Is a limited purpose specification more conducive to sharing than a vague purpose? Lastly, (4) what do users think of the profiles that advertisers create about them, and do they see benefit in accessing their own profiles?

To answer our questions, we assigned each participant to one condition that described a particular data practice scenario. We described our scenarios in the context of a hypothetical news website (AllNews) where an advertising company (Best Ads or Facebook) would collect and use data from visitors. We considered seven scopes of collection and use, and three retention periods (one week, three months, one year), resulting in 21 conditions in total. Table 8.2 provides an overview of our scenarios.

While each participant saw only one of the scenarios, all participants were shown a realistic example of a behavioral profile that advertisers may create about users. We were interested in what surprised or concerned participants about the profile, and perceived benefits of being able to access or edit the information in their profiles.

8.5.2 Survey Implementation

The survey consisted of four sections. The first section asked for participant demographics, Internet use, and opinions about online advertising. To signal that the survey required more than minimal effort, we started with an open-ended question asking for their opinion about online advertising.

In the second section, we confronted participants with one scenario from Table 8.2. We first asked participants to visit the AllNews website, a static website we modeled after the CNN.com homepage with changed branding logos and text. Hyperlinks and forms were disabled. To verify that participants were following instructions, we required them to identify the title of a news article that appeared on the AllNews homepage, presented among four decoy titles.

Next, we asked participants to imagine that they were users of the AllNews website and provided a short explanation of how targeted ads work. Then we told them that the AllNews website had contracted with a company that was interested in showing users targeted ads and informed them about the scenario-specific data practices. Appendix F.3 shows a sample scenario. We asked participants to read the given scenario thoroughly and then assessed their understanding with a follow-up question about the stated data practices. Participants who answered incorrectly were shown the correct answers, asked to read the scenario again, and then tested again. We then collected participants' willingness to share the 10 different types of personal information shown in Figure 8.5. These data types were chosen based on what advertising companies typically collect or infer. For each item, participants rated their comfort with sharing that item on a 5-point Likert scale. We followed up with open-ended questions asking them to explain why they would or would not be comfortable with the advertiser collecting those data types. Next, we presented follow-up questions for at most four randomly selected data types to avoid fatigue.

In the third part, we showed participants a realistic example of a profile. Rather than asking participants to access their own profiles kept by major online advertisers, which may fluctuate in content and may require registration, we showed them a sample profile (see Appendix F.5) created by combining data collected from real user profiles [104]. We contextualized the sample profile to each participant by dynamically adapting the first two categories in the profile (location, individual demographics) to the participant's IP address and their information provided at the beginning of the survey. We asked them to select from a list two items that appeared in the sample profile to ensure that they read it. We then elicited participants' surprise and concerns about the profile's content and asked for perceived benefits (if any) of having access to their profile. We ended the survey with eight questions from the Internet Users' Information Privacy Concerns (IUIPC) instrument to gauge participants' general privacy concerns [198].

8.5.3 Analysis Approach

We cleaned the data by removing participants from outside the US (39); that completed the survey in <5 minutes (3); were inconsistent in whether they visited news websites regularly (27); failed the AllNews website (120) or sample profile (13) content questions. We analyzed valid responses from 1,882 participants aged 18–79 (mean=34, σ=12.2). Half the participants were female. Participants predominantly indicated low (40%), medium (33%), or high (20%) Internet literacy, and few participants indicated no Internet literacy (7%). Participants exhibited a diverse range of occupations, and were well educated (31% some college, 10% Associate's degree, 35% Bachelor's degree, 15% Graduate degree). We did not observe any statistical differences between conditions concerning education, tech savviness, gender, age, or Internet literacy.

Quantitative Analysis

We performed Kruskal-Wallis rank sum tests for each of 10 assessed data types to determine for which data types the scenario and retention period had a statistically significant impact. We found that scenario had a significant impact on willingness to share some types of information, but retention period did not.

We then performed binary logistic regressions on all data types using the scenarios and retention periods (to verify its null effect) as independent variables. The willingness to share questions served as dependent variables with "strongly agree" and "agree" responses binned as "agreement," and "neutral," "disagree," and "strongly disagree" responses binned as "non-agreement." In addition to scope and retention, our regression models controlled for participants' age and gender, and included indicator variables for privacy concerns, positive opinion of targeted ads, usage of ad blocking tools, positive opinion of the AllNews website, Facebook account, tech savviness, and whether participants answered correctly at least one of the scenario understanding questions.

Qualitative Analysis

We qualitatively analyzed multiple open-ended questions: participants' reasons for comfort/discomfort with sharing certain information types, surprise and concerns about the sample profile content, perceived benefits of accessing their own profiles. All questions were shown dependent on a participant's answers to preceding Likert-scale questions. For instance, we asked about their reasons for being concerned if they indicated concern about the sample profile. In addition, participants assigned to conditions mentioning "other purposes," were asked what those might be, in order to understand whether they had positive or negative associations.

For each open-ended question, we randomly selected a 10% sample of the respective responses, drawn evenly from all 21 conditions, for qualitative data analysis. Considering the large total of participants, this provided us with a sufficiently large sample per question for qualitative data analysis (138–199 responses per question). Due to random sampling within each condition we were confident that the selected responses are representative of the whole dataset, which we confirmed with cursory inspection of the remaining responses. For the "other purposes" responses, all 792 responses were coded as *positive*, *negative*, or *ambiguous* in order to enable integration in the regression models.

For each open-ended question, two researchers independently evaluated the same subset of responses to derive relevant codes from which a question-specific coding taxonomy was jointly developed. Next, they coded the full sample of responses. Initial inter-coder reliability was evaluated with Cohen's Kappa coefficient. Coding disagreements were subsequently resolved on a per-statement basis in an iterative process between the two coders, resulting in fully reconciled response annotations for each open-ended question, which were used in subsequent analysis. In total, 2,245 statements were examined as part of qualitative data analysis, resulting in 2,919 assigned codes.

8.5.4 Limitations

Our analysis is based on self-reported data regarding participants' sharing comfort. Users actual behaviors may differ. We created a browsing simulation scenario, asking participants to visit a news website to emulate a real Internet experience, while not a perfect substitute for behavioral data, the fact that participants reacted differently to different scenarios and showed to be invested in the study by providing rich qualitative data, suggests that users provided with effective notice

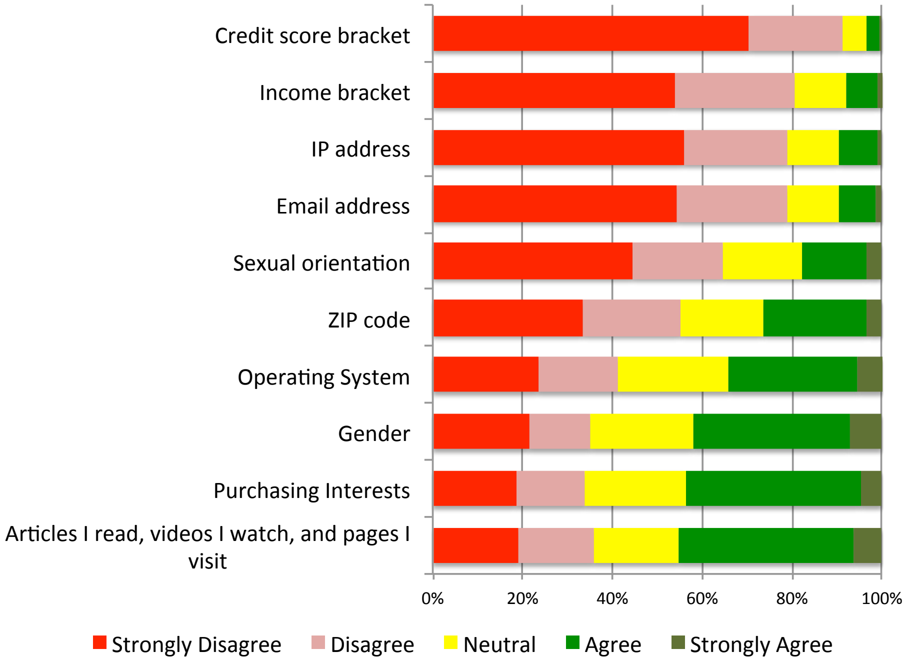


Figure 8.5: Participants' responses to the statement, "I would be comfortable if *[Best Ads / Facebook]* collected or otherwise inferred the following information about me." The green shades represent sharing comfort, while the red shades indicate discomfort.

may behave accordingly to their stated preferences. MTurk users have known demographic differences compared with the general Internet population; however, it has been demonstrated that MTurk participants behave similarly in studies as subjects recruited from other sources [172].

8.6 Results of Study Two

First we discuss how different factors affected participants' comfort with sharing different types of data. We provide both statistical and qualitative evidence of participants' sharing preferences. We then discuss participants' reaction to a sample contextualized behavioral profile.

8.6.1 Sharing Preferences

Overall, almost half the participants (45%) were comfortable sharing information with advertising companies. They were most comfortable sharing the pages visited, articles read, and videos watched on the news website (45%), the products they might be interested in purchasing (44%), gender (42%), computer's operating system (35%), ZIP code from where they access the Internet (26%), as well as sexual orientation (17%). Only a small fraction of participants were comfortable sharing their email address (9.7%), IP Address (9.6%), income bracket (7.9%), or credit score (3.3%). Figure 8.5 shows overall willingness to share.

Participants' sharing comfort did not only depend on the sensitivity of information, but also the scope of collection and use, necessity of collection, and perceived benefits and harms of disclosure. Furthermore, personal attitudes such as trust in the visited website, the opinion of targeted ads, and privacy concerns had a strong effect on willingness to share. We discuss how the elements in the tested scenarios impacted participants' comfort with sharing different data types. First we present quantitative results, followed by qualitative results that provide further insights into participants' decision making process. Overall, our results help to explain why participants might not be willing to share even apparently innocuous information under some circumstances, but they might be willing to share arguably more personal information under other circumstances. Finally, using results from our regression models, we discuss how personal attitudes towards targeted ads, trust perceptions, and privacy concerns affected participants' sharing comfort.

Factors Affecting Sharing Comfort

Statistical analysis provided evidence that participants took into account the scope of collection and purpose of use to make information sharing decisions. Non-parametric analyses of variance showed general differences between scenarios for six information types: interactions with the AllNews website, purchasing interests, gender, ZIP code, sexual orientation, and email address. However, retention period was not a significant factor in predicting willingness to share for any information type. We validated these results with our regression models, which also allowed us to investigate the particular direction of the effects. The regression model results are provided in Table F.3 in the Appendix. Participants' willingness to share their online interactions and purchasing interests decreased as the scope of collection and use increased, while their willingness to share their gender, email address, ZIP code, and sexual orientation was more nuanced.

Online interactions. Participants' comfort sharing their online interactions was similar (49%-53%) in scenarios where this data type was exclusively used for targeted ads either on the first-party website (S1) or other visited websites (S3), as well as when it was used for targeted ads and other purposes only on the first-party website (S2). But, they were statistically less comfortable sharing (βs from -0.62 to -0.55, p-value < 0.001) this data type (34%-41%) in scenarios where the information would be linked with offline data to receive coupons (S4), used on other websites for unspecified purposes (S5 & S6) and shared with Facebook (S7).

Purchasing interests. Participants' comfort sharing their purchasing interests was similar (45%-51%) in scenarios where this data type was exclusively used for targeted ads either on the first-party website or other visited websites (S1 & S3, 46%), used for targeted ads and other purposes only on the first-party website (S2, 51%), and even when linked with offline data to receive coupons (S4, 45%). However, participants were statistically less comfortable sharing this data type (38%) in scenarios where the information was going to be used on other websites also for other unspecified purposes (β=-0.44, p-value=0.03) and shared with Facebook (β=-0.35, p-value=0.07).

Gender. None of the scenarios (35%-47%) were statistically different ($alpha = 5\%$) in our regression model from the baseline scenario (S1, 40%). However, a larger fraction of participants was comfortable sharing their gender with Facebook (S7, 47%) and when used for targeted ads and other purposes only on the first-party website (S2, 47%) compared with scenario S6 where the information would be used on other websites and for "unspecified purposes" (35%).

ZIP code. As in the case of gender, fewer participants (21%) in scenario S6 were willing to share this data type than participants in all other scenarios (25%-34%). This difference was significant (β=-0.64, p-value=0.005) when compared with the baseline scenario (S1, 29%). For both the previous and this data type, the "unspecified" purposes seemed to negatively impact willingness to share, likely because participants assumed more "negative" than "positive" purposes for these data types.

Sexual orientation. Participants' comfort sharing their sexual orientation was low and similar in scenarios where this data type was used for targeted ads either on the first-party website (S1, 13%) or other visited websites (S3, 16%), going to be linked with offline data to receive coupons (S4, 18%), and used on other websites for other unspecified purposes (S6, 13%). However, they were statistically more comfortable sharing this data type in scenarios where the information was going to be shared with Facebook (S7, 23%, β=0.8, p-value<0.001) or used on other websites for other unspecified purposes but without sharing it with third-parties (S5, 21%, β=0.56, p-value=0.04).

Email address. While only few participants (10%) were comfortable sharing their email address, a larger fraction of them were comfortable sharing it to receive coupons (S4, 13%, β=0.56, p-value=0.07) and with Facebook (S7, 16%, β=0.88, p-value=0.002) when compared to scenario S1 (8%).

Computer's and sensitive information. None of the scenarios affected participants' willingness to share either computer's information such as IP Address and OS or more sensitive information such as income bracket of credit score.

Effect of unspecified purposes. For the scenarios that included "other purposes" uses, we coded participants' interpretations of those purposes as *positive* (e.g., suggesting content, measuring success of ad targeting, observing consumer trends), *negative* (e.g., selling information to other companies, creating mailing lists, sharing with the government), or *ambiguous* (e.g., participant uncertain or response unclear). The majority of participants perceived other purposes as negative (52%), some as positive (35%), and a smaller number of responses were ambiguous (13%).

We included this variable in our regression models to evaluate whether opinions about "other purposes" had an impact on sharing comfort. A positive perception of "other purposes" had a positive impact on the level of comfort for sharing online activities (β=0.32, p=.07), gender (β=0.32, p=.07), and ZIP code (β=0.46, p=.002). It is important to mention that to measure the effect of "other purposes" we had to split our sample into three subsets (positive, negative, ambiguous), decreasing statistical power.

Taken together, these quantitative results suggests that participants paid attention to the tradeoffs presented in their given scenario. We now turn to a qualitative discussion that provides further insights into the participants' reasoning.

Why people would share information

We asked participants open-ended questions to understand why they were comfortable sharing some data types with advertisers. We analyzed a 10% sample of 206 participant responses, resulting in 255 coded statements leading to the reasons shown in Table 8.3. The two main reasons why participants were willing to share data with advertisers were receiving relevant advertisements (25.8%), and feeling that the data was public rather than personal, or private (18.8%). The top data types that participants considered public were operating system information, gender and online activities.

Reason	Count	Percent
Receive relevant ads	66	25.8%
Not personal/secret/private	48	18.8%
Does not matter	32	12.5%
Not personally identifiable	23	9.0%
Required to provide relevant service	18	7.0%
No harm in sharing	12	4.7%
Easy to infer	8	3.1%
Technical aid	7	2.7%
Not embarrassing	7	2.7%
Receive better deals	7	2.7%
Location targeting	7	2.7%
Not privacy invasive	6	2.3%
Other	14	5.5%

Table 8.3: Reasons why participants were willing to share data with advertisers. Participants could provide multiple reasons (255 codes=100%, n=206).

Participants (12.5%) felt that sharing data types such as gender or operating system did not matter to them. For example, they stated, "...it is not big deal" or "I really don't care," but did not provide further explanations. Some participants (9%) mentioned that they were comfortable sharing general information, like ZIP code and operating system, as it could not be used to personally identify them. Other participants (4.7%) saw no harm in sharing their data, because at most advertisers would send them ads.

Participants also expressed data type-specific reasons for sharing. For example, participants wanted to share ZIP code to receive location-specific benefits, like local deals and news. Participants felt comfortable sharing information about the articles, videos and pages they visit to receive better service, for example, recommendations for interesting news articles. Participants wanted to share information about the products they were interested in to receive discounts. A few participants felt that operating system information was required for the website to display properly on their computer. Some participants indicated that they were proud of their sexual orientation and were not embarrassed to share it.

We analyzed why participants were more comfortable sharing certain data types for specific scenarios. Recall that participants were significantly more willing to share email in the Facebook (S7) and offline (S4) scenarios compared to other scenarios. Participants in the offline scenario were willing to share email to get better deals and services. In the Facebook scope they felt that they had already voluntarily provided their email to Facebook. Participants were more willing to share gender in the Facebook scenario as some did not care if Facebook knew their gender, and others felt that their gender was not a secret. Participants were also more comfortable sharing their sexual orientation in the Facebook scenario (S7). They were also more comfortable when sexual orientation could be used for other purposes, but would not be shared with other parties (S5). Participants' felt that it was no big deal, no harm would come out of it, no shame, or that it did not point to their identity. They also felt that services could be tailored to their interests. One participant said, "They can't do anything knowing that," and another mentioned, "The ads and services could be tailored to feature products that are in line with my lifestyle."

Reason	Count	Percent
Personal information	187	23.8%
None of their business	114	14.5%
Unnecessary for advertising	96	12.2%
Invasion of privacy	81	10.3%
Location tracking	50	6.4%
Ad spam	43	5.5%
Lack of consent	42	5.3%
Inference of information	24	3.1%
Personally identifiable	22	2.8%
General tracking	20	2.5%
Computer harm	20	2.5%
Unreliable information	18	2.3%
Other	68	8.6%

Table 8.4: Reasons why participants would not share data with advertisers. (786 codes=100%, n=575).

Why people would not share information

We further investigated reasons for feeling uncomfortable with sharing data with advertisers. The analyzed 10% sample consisted of 575 participant responses, resulting in 786 coded reasons. The larger number of coded statements reflects the higher percentage of participants not willing to share information with advertisers. We extracted the reasons shown in Table 8.4. Overall, participants were not comfortable sharing data they considered personal information (23.8%), there was no need to know (14.5%), or was unnecessary for advertising (12.2%). Note that the reasons for unwillingness to share are almost opposites of those for willingness to share previously discussed. For example, top reasons for willingness to share included considering information to be non-personal and also relevant for advertising.

As with willingness to share, participants' unwillingness to share varied by data types. Participants were mainly concerned about their location being tracked if they shared ZIP code or IP Address. And the main reason why participants would not like sharing their email address was because they would not want to receive unsolicited emails. Participants were unwilling to share IP address with advertising companies, because they considered it personally identifiable information. Furthermore, certain data types such as gender, income bracket or online activities were not perceived as reliable indicators of their interests for ads. One participant stated, "I don't see how what I read about accurately reflects any products or services that I would or could actually purchase or even be interested in."

The majority of participants were not comfortable sharing credit score, income bracket, and sexual orientation because they considered these types of information personal, unnecessary for advertising, or thought it was nobody's business. Some participants viewed the collection of these types of information as an invasion of their privacy. Some participants expressed concerns about discrimination based on price, gender or sexual orientation. Some did not trust Facebook and other advertisers. They felt that advertising companies might sell or share their data with third-parties. A few believed that their data may be stored insecurely, or that they may become victims of identity theft. Lastly, some participants thought that sharing data such as gender may increase the risk of assault or physical harm.

Participants considered what they did on Facebook personal and believed that they could not access Facebook freely if Facebook tracked their habits and activity history. For the offline scenario, participants were uncomfortable combining offline and online information. Participants in the "other purpose scenarios" considered collecting online activities as too much collection of information and compared it to spying on them. One of the participants said, "That is personal information. I feel like they are spying on me if they know that. That makes me feel uncomfortable."

Participants were uncomfortable sharing purchasing interests for Facebook (S7) and other purpose scenarios (S2, S5 & S6). Participants did not want Facebook to know too much of their browsing and shopping habits. They were concerned about Facebook hounding them with ads, and that Facebook could announce or share purchases that were deemed personal. Participants in the "other purpose" scenario were uncomfortable sharing purchase interests for several reasons including invasion of privacy, not relevant to their interests, unwilling to purchase from ads, and because they did not specifically give permission for collecting such information. Participants expressed that location could be inferred from ZIP code, and their location, or where they live, was personal information. They felt that sharing such personal information was invasion of privacy. They also did not want to be targeted by local businesses.

Other factors affecting disclosure

Based on our regression models, we further explored the impact of participants' personal characteristics such as gender, age, and attitudes towards privacy, targeted ads and the AllNews website. Attitudinal variables showed a strong effect on sharing comfort for all explored data types. In particular, positive opinion about targeted ads and about the AllNews website significantly increased sharing comfort (βs ranged from 0.58 to 1.87, p-values <.001). On the other hand, the more privacy concerned participants were, the less comfortable they were sharing any data type (βs ranged from -0.79 to -0.42, p-values <.002). Demographics had a milder effect. Participants' gender was only significant for the ZIP code and gender data types, where male participants were more comfortable sharing those types of information than female participants (p-values <.002).

8.6.2 Reaction to User Profile

We showed participants a realistic example of a behavioral profile and elicited their reactions. Participants rated their comfort, surprise, and concerns about the profile's content on a 5-point Likert scale, see Figure 8.6a. The majority was surprised (73% agree or strongly agree), concerned (84%), and not comfortable (82%) with the contained information. We first discuss what concerned and surprised participants, followed by perceived benefits of accessing their profiles.

Concerns and surprise

Participants who indicated surprise or concern could express their reasons in an open-ended question. For concern, the analyzed sample consisted of 157 responses, from which we identified 20 unique concerns. For surprise, the analyzed sample consisted of 138 responses, from which we identified 17 categories. Due to large overlap between concerns and surprises, we only show the concern categories in Table 8.5. Major reasons for surprise and concern were the amount of information collected by online advertisers (17.9% concern, 35.5% surprise) and the level of detail (7.5% concern, 15.2% surprise). A major concern was that the collected information is

Reason	Count	Percent
amount of information	55	17.9%
personal information	45	14.7%
general concern	31	10.1%
level of detail	23	7.5%
lack of consent	21	6.8%
general harm	20	6.5%
inaccurate information	13	4.2%
household (info)	11	3.6%
credit score (info)	11	3.6%
accurate information	11	3.6%
income (info)	9	2.9%
unnecessary for ads	9	2.9%
inferred information	8	2.6%
location (info)	7	2.3%
other	33	10.7%

Table 8.5: Reasons for concern about the sample profile (307 codes=100%, n=157).

considered personal (14.7%), many participants also voiced general concern about the profile (10.1%).

Participants (6.8%) mentioned lack of consent to the collection of data as a concern and were concerned about selling and sharing of information with third parties. One participant stated, "Collection of that much information on a person without their knowledge seems dishonest. Also the possibility of that information being used against a person or ending up in the wrong hands is disturbing." Lack of transparency regarding data collection was also a concern. Further concerns included both the inaccuracy (4.2%) and accuracy (3.6%) of information. Particularly inaccurate information was associated with potential harm: "...Incorrect information could lead to problems to the person being profiled through no fault of their own."

There were concerns about a number of specific data types, such as household (3.6%) or credit score (3.6%), specially as the amount and detail of collected data was seen as unnecessary or excessive for advertising (2.9%): "The amount of information collected, most of which is very personal and unnecessary for their purposes." For some participants, awareness of the scale of data collection increased their distrust in advertisers, e.g., "...NO corporation has any business collecting this much information... This is really incredibly upsetting... my mistrust of internet advertisers has increased tenfold by seeing that. I will make it my goal to block their access to ALL of my information..."

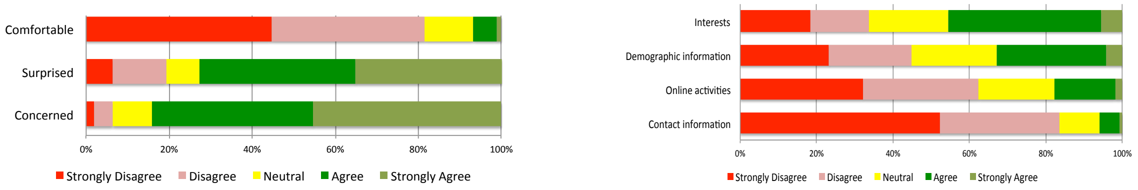
Surprise about the information in the profile largely aligned with stated concerns. One participant stated, "I was very surprised by just how much info there was; it was like a resume or background check." Participants were most surprised that the profile contained detailed information about an individual's household (5.6%) and financial information, such as credit score (5.6%) and income bracket (3.9%). One participant stated, "How much they made, where they lived, the credit bracket... stuff that I have been told all of my life to keep to yourself, but they have it on a spread sheet."

The level of accuracy was another reason for surprise (5.6%). Many participants felt uncomfortable with accurate information being collected about them, for example, "the profile almost matches me EXACTLY. It was a little terrifying." Participants were also surprised about the potential of creating inferences (2.6%) and data aggregation (2.2%). One participant summarized it as, "It has a lot of information that seems meaningless, but put together creates an alarmingly specific picture." Others were more surprised with the level of inaccuracy in parts of the profile (6.1%), e.g., "Many items on the profile did not fit my situation or personality. I am surprised at the assumptions made by the ad software based on very few points a reference."

Benefits of profile access

We asked participants to "[t]hink about the ability to view and edit the information that advertising companies know about you. How much do you agree or disagree with the following," showing them six statements. 90% of participants believed (agreed, strongly agreed) that they should be given the opportunity to view and edit their profiles. A large percentage wanted to be able to decide what advertising companies can collect about them (85%) and saw benefits in being able to view (79%) and edit profiles (81%). The majority thought that the ability to edit their profiles would provide companies with more accurate data (70%) and allow them to better serve the participants (64%). These results indicate that the motivation for accessing their own behavioral data is not only to remove information but actually enhance the utility of targeted ads.

After viewing the sample profile, participants rated their level of comfort with sharing different categories of information (see Fig. 8.6b). The majority was not comfortable sharing contact information (83% disagreement), which matches the results obtained before they viewed the profile (79% unwilling to share email). After viewing the profile, participants were either comfortable sharing interests (46%) or neutral about it (21%), which is similar to the willingness to share purchasing interests reported before (43% agreement, 19% neutral). However, the comfort level for sharing online activities, such as the articles read, was much lower (18% agreement) than the willingness to share this information reported before viewing the profile (45%). Thus, participants are more willing to share interests than online activities with advertisers, which conflicts with the current practice of primarily collecting activity data.



a) Reaction to profile

b) Willingness to share after seeing profile

Figure 8.6: Participants' responses to the statements, a) "I am *comfortable / surprised / concerned* with the information that such profiles may contain" and b) "In general, I am comfortable sharing with advertising companies," after they saw the sample profile.

Participants describe the benefits (if any) of having access to their profiles. We analyzed 199 responses, resulting in 259 coded statements categorized in Table 8.6. While 13.5% saw no benefit in gaining access to their profile or only benefits for the advertiser (3.9%), the main reported benefit was being able to know what information advertising companies were collecting (28.6%). Participants perceived it as "fair" to have access to the information collected about them, e.g., "I think transparency is best. At least i know what information they have gathered about me." A

Benefit	Count	Percent
know what is collected	74	28.6%
make ads more relevant	49	18.9%
no benefit for consumer	35	13.5%
consequences of online behavior	24	9.3%
make profile more accurate	20	7.7%
want to edit or delete info	15	5.8%
benefit for advertiser	10	3.9%
make better privacy decisions	9	3.5%
understand reasons for ads	8	3.1%
make profile less accurate	6	2.3%
other	9	3.5%

Table 8.6: Perceived benefits of access to own profiles (259 codes=100%, n=199).

Reason	Count	Percent
control what is collected	64	22.5%
make profile more accurate	47	16.5%
protect personal information	47	16.5%
know what is collected	43	15.1%
want to edit or delete info	40	14.0%
make ads more relevant	31	10.9%
make profile less accurate	9	3.2%
other reasons	4	1.4%

Table 8.7: Reasons for wanting edit profiles created by advertisers (285 codes=100%, n=173).

small fraction (1.2%) explicitly stated that access to the profiles would increase their trust in the advertising company.

Other stated benefits fall into two categories: a way for improving the utility of targeted ads or for enhancing privacy protection. Many participants would like to increase the relevance of ads shown to them (18.9%) and believed that by accessing their profiles they could correct inaccuracies (7.7%), and remove or edit some information (5.8%). A related reason was the interest in understanding why they see specific ads (3.1%). Participants also thought that accessing the profile would help them better understand the consequences of their online behavior (9.3%). Some participants saw that as a motivation for making better privacy decisions or changing their behavior in the future (3.5%), e.g.,"It might persuade more of them to be more careful about what information they give out and avoid websites that abuse their privacy." Only a small fraction of participants (2.3%) wanted to make profiles less accurate by removing personal information.

Participants who agreed that "they should be given the opportunity to view and edit their profile," explained their reasons. The analyzed sample of 173 responses led to 285 annotations across the categories shown in Table 8.7. The main reasons for wanting to edit the profile were the ability to control what is being collected (22.5%), to protect personal information (16.5%), to make the profile more accurate (16.5%), and to know what is being collected (15.1%). This reflects the perceived benefits of access as well as the major concerns and surprises about the amount of collected information. Other participants were generally interested in editing or deleting information from the profiles (14.0%), some with the goal of making ads more relevant (10.9%).

A number of participants thought that their online behavior did not reflect their interests and would prefer to provide or adjust their interests and ad preferences instead. One participant expressed,“it would help both myself and the advertising companies if they knew my preferences from me and not my searches. My searches don’t necessarily indicate my interests.” Another said, “I think it’s important because the articles I read might contribute to a different profile about me than what I actually like. For example, I might read a lot about Republican politics but be a registered Democrat.” This may explain why participants indicated to be more comfortable sharing their interests than their online activities, see Figure 8.6b. Some participants further stated that being able to control what information advertisers collect and use about them would increase their willingness to share certain information with them, “there are certain details of my life that I do not want known. I would be more willing to share some information, to allow properly targeted advertising, if I were also allowed to guard other information that I feel is personal and private. This would also help the advertisers because they would have accurate information and I would be more likely to respond to their ads.” While a few participants stated that they would purposefully make the information in a profile inaccurate (3.2%), a much larger fraction (16.5%) was willing to correct inaccuracies and they felt that correcting inaccuracies could be beneficial to them as well as advertising companies.

In summary, our participants were concerned and surprised by the amount and level of detail of information that advertisers collect. Yet, the reported reasons for and benefits of having access to their profiles show that this could potentially be addressed by being transparent about data collection and offering control about what data types can be collected and used for advertising.

8.7 Discussion

Public policy discussions related to OBA have focused on the need for privacy notices and opt-out opportunities [5]. The advertising industry has developed icons that indicate the use of targeted advertising on websites, and has provided tools and websites to enable consumers to opt-out of targeted advertising from individual or a list of advertising companies. However, these are “all or nothing” solutions that require users to make decisions without a full understanding of advertising companies’ practices. We investigated how practices routinely used by advertisers influence users’ willingness to disclose different types of information. Results from our studies could assist advertising companies to refine their practices and help shape regulation in this area.

Context Matters. More than half (55%) the participants in both studies were not comfortable sharing any type of information. However, those who were comfortable sharing some data showed nuanced preferences. In study two, participants were comfortable sharing information that they deemed necessary for advertising or delivery of a better service. For example, 49–53% of participants were comfortable sharing their online interactions when data was used for targeted ads on the visited website as well as other websites, but not if the data was used for other purposes outside the scope of the visited website, or combined with PII (e.g., Facebook and offline scenarios).

They were also more comfortable sharing their email addresses to receive coupons (13%) and with Facebook (16%) than with other websites or for other purposes (7–9%). Participant’s were comfortable sharing their gender, ZIP code and other data types under some circumstances, but not always. Disclosure preferences were generally influenced by perceived benefits and potential harms. These results suggest that binary approaches like Do-Not-Track or opt-outs, which do not consider the context of collection, as well as ad companies' current disclosure practices are less than optimal.

Meeting Users' Expectations. Participants were more comfortable sharing some data types when they assumed that the "other purposes" for which companies would use that data were positive. However, many participants tended to assume the worst. Consistent with our findings in Chapter 5, participants indicated that they were uncomfortable sharing data with companies that had not obtained proper consent to collect and use their data. Further, even users who recognized benefits in sharing remained uncomfortable sharing sensitive data types such as income bracket, credit score bracket, sexual orientation, email address and IP address. They considered them not necessary for advertising, and potentially harmful. Participants were surprised and concerned about the types and amount of information shown in the sample profile. Many participants did not see the need to collect so much data for advertising purposes. Participants also voiced concerns about sharing information with third parties. While participants in study two did not appear to distinguish between data retention periods of one week, three months, and one year; participants in study one were more comfortable with a defined retention period than an indefinite one. However, today advertising companies are often silent about retention periods. Overall, the misalignment between users' expectations and current OBA disclosure practices suggests that the advertising industry should modify their practices in this area.

Access is Needed. A large fraction (90%) of participants in study two wanted to view and edit their profiles for various reasons, including awareness and control over the collection of their data, to make their profile more accurate, to remove unwanted information, and to receive more relevant ads. Furthermore, participants believed that their online activities did not reflect their interests, yet were comfortable sharing their purchasing interests (46%). These results suggest that providing access to users could increase both the effectiveness of targeted advertising and users' comfort and trust. This contrast with today's practice of making inferences about users' preferences with the risk of making mistakes and infringing on their privacy. Ad companies could limit inferences to non-sensitive information such as demographic or non-sensitive interests and allow users to opt in to receive ads from more sensitive categories (e.g., according to income range, sexual orientation, or health interests). A handful of advertising companies are already giving users access to anonymous (i.e., cookie based) profiles, such as BlueKai Registry, but it is unclear whether users are aware of those access options, and the vast majority of advertising companies don't offer any access.

Improving Privacy Notices. Understanding the factors that are relevant to users is important for designing notices that communicate OBA practices in an effective and actionable manner. Our results show that users are interested in knowing not just what data is being collected, but also how it will be used and shared, and how those uses may benefit or harm them. However, it is not reasonable to expect users to read the privacy policy of every advertising and tracking company to determine whether their practices are acceptable. Therefore we need tools that can automatically identify the trackers that match users' preferences. Current efforts are attempting to extract important privacy policy elements through crowd sourcing and natural language processing [98, 99]. Success of these efforts, however, relies on companies being transparent about their data practices, and adhering to their stated collection, use and sharing practices.

Chapter 9

A Field Trial of Privacy Nudges for Facebook

Anecdotal evidence and scholarly research have shown that Internet users may regret some of their online disclosures. To help individuals avoid such regrets, we designed two modifications to the Facebook web interface that nudge users to consider the content and audience of their online disclosures more carefully. We implemented and evaluated these two nudges in a 6-week field trial with 28 Facebook users. We analyzed participants' interactions with the nudges, the content of their posts, and opinions collected through surveys. We found that reminders about the audience of posts can prevent unintended disclosures without major burden; however, introducing a time delay before publishing users' posts can be perceived as both beneficial and annoying. On balance, some participants found the nudges helpful while others found them unnecessary or overly intrusive. We discuss implications and challenges for designing and evaluating systems to assist users with online disclosures.

9.1 Introduction

Online social networks such as Facebook are designed to encourage sharing, facilitating the seamless and immediate broadcasting of all kinds of information. While sharing information through social networks generally benefits users, seemingly innocuous disclosures can lead to substantial negative consequences. Lack of awareness of the potential audience, posting while in highly emotional states, and hasty disclosures have been shown to lead social media users to experience regret [200]. Research in the fields of psychology, behavioral economics, and behavioral decision making has uncovered cognitive and behavioral biases that affect decision making. These biases are systematic deviations from what traditional economists call rational decisions. Furthermore, when limited resources (e.g., time or information) are available to make a decision, human beings often rely on heuristics or shortcuts. These biases and heuristics have been shown to impact privacy decisions [25, 201, 202] and privacy blunders in social media are vivid examples of the hurdles users face.

This chapter is based on “A Field Trial of Privacy Nudges for Facebook” [199].

Behavioral economists have proposed the use of soft paternalistic interventions to help people overcome behavioral biases that affect decision making. These interventions are designed to "nudge" (instead of force) people towards behaviors that have been shown to be publicly desired, but difficult to follow, without limiting people's autonomy [203]. Acquisti has proposed to use soft paternalistic interventions to improve security and privacy decisions [204]. We refer to soft-paternalistic mechanisms that nudge people towards more thoughtful and informed privacy-related decisions as *privacy nudges*.

Inspired by the literature on behavioral decision research and nudging, as well as by our prior work on regrettable Facebook behavior [200], we investigated the impact of Facebook privacy nudges. In this paper, we describe the design and evaluation of mechanisms that nudge Facebook users to consider more carefully the content and context of their online disclosures through visual cues and time delays. We developed a platform that enables us to deploy nudges and evaluate them with users in longitudinal field trials.

We conducted an exploratory 6-week field trial with 28 Facebook users. Our goal was to gain an understanding of how users perceive and interact with the features of our nudges. We analyzed both quantitative and qualitative data about participants' interactions with the nudges, the content of their posts, as well as their opinions collected through a final survey. We found that reminders about the audience of posts can prevent unintended disclosures without major burden; however, introducing a delay before publishing users' posts can be perceived as both beneficial and annoying. While many participants found the nudges helpful, others found them unnecessary or overly intrusive, suggesting that nudges may not be appropriate for everyone.

Our work makes two contributions. First, we developed an experimental platform that modifies Facebook's interface and collects users' behavioral data to operationalize and evaluate the concept of Facebook privacy nudges. Second, we identified key aspects worth considering when designing and evaluating a privacy nudging system.

9.2 Background and Related Work

Our work was inspired by scholarly research on problematic disclosures on social media, cognitive and behavioral biases, privacy decision making, and nudging.

9.2.1 Potential for Regret

Offline, people are naturally good at tailoring comments, gestures, and actions to specific audiences [205]. However, online (and in particular on social media such as Facebook), users tend to communicate with many groups (e.g., friends, co-workers) simultaneously, and as a result encounter difficulties in adapting messages for different audiences. Thus, shared content is often visible across groups, leading to a phenomenon called "context collapse" [206]. These issues are exacerbated by the fact that even experienced users have difficulties with Facebook privacy settings [207, 208]. Hence, a variety of dynamics lead to regrettable Facebook posts. Content is sometimes viewable by unintended audiences and users create posts "in the heat of the moment," which can lead to unintended disclosure and regret [200]. Unintended disclosures can lead to a range of consequences, including stalking, identity theft, blackmail [209], and reputation damage [210].

9.2.2 Biases, Heuristics, and Privacy Decision Making

Behavioral biases and heuristics can lead to systematic errors in decision making [211]. While, biases and heuristics have been studied in many contexts, here we focus on those that are closely related to our nudges. Bounded rationality forces individuals to rely on heuristics to simplify the choices available; however, sometimes choices with the best outcomes may be inadvertently discarded [212]. The economic effects of asymmetric information, which limits rational decision making, has been studied in the market of used cars [213]. Nearly half a century later, problems of bounded rationality and incomplete information are alive in social media.

Bounded rationality and asymmetric information prevent individuals from anticipating the audience for their posts. While it is easier for users to think in terms of broad audiences (e.g., friends, friends of friends, public), more granular groupings (e.g., parents, neighbors, church) can help mitigate unintended disclosures. Similarly, as Facebook comments inherit the audience of the original status update, it may be impossible for the person commenting to determine the audience of his or her comment—an example of asymmetric information. In fact, recent research by Bernstein et al. found that Facebook users "consistently underestimate their audience size for their posts, guessing that their audience is just 27% of its true size" [214]. As a result, Facebook users often post content that can be viewed by an unintended audience, which may lead to regret [200]. One of the nudges we present here attempts to mitigate problematic online disclosures associated with bounded rationality and asymmetric information.

Another relevant bias is known as hyperbolic time discounting, the fact that individuals use variable and inconsistent discount rates over time and often assign higher utility to present choices than to future ones [215]. For instance, people tend to procrastinate because they over-estimate the enjoyment of not doing work now and under-estimate the future consequences of delaying work [216]. In the privacy domain, Acquisti has shown that people often trade their personal information for immediate gratification [201]. The work on dual process theory is also relevant. For instance, Kahneman posits the existence of two processing systems in our brains: intuition (System I) and reasoning (System II). Intuition tends to be fast, automatic, and rely on heuristics, while reasoning is slower and involves more conscious judgement [217]. Our prior work on regrets found evidence of impulsive behavior, often driven by highly emotional states [200]. The second nudge we present was designed to mitigate problematic disclosures potentially due to hyperbolic time discounting and impulsive behavior.

9.2.3 Soft-paternalism and Privacy Nudges

Soft paternalistic interventions attempt to help individuals by mitigating behavioral biases (or, in some cases, exploiting them) to achieve the outcomes that better align with users' preferences. Thaler and Sunstein popularized the idea of nudging. They defined a nudge as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" [218]. Acquisti described the role of hyperbolic discounting and immediate gratification in the so-called privacy calculus [201]. He further proposed that nudges could be used to influence privacy decision-making in a manner that decreases users' regret [204].

Although not explicitly referred to as nudging, researchers in the fields of HCI and Persuasive Technology have explored mechanisms to assist users with privacy and security decision making. Forget et al. built a system to nudge users to create stronger passwords [219]. Ur et al. showed that certain password meter designs may encourage users to create stronger passwords [220]. Wilson et al. studied the effect of predetermined privacy profiles to assist users with location sharing disclosures [221]. Choe et al. investigated the impact of the "framing" heuristic (i.e., people would prefer alternatives that are framed as gains over those framed as loses, even when the two alternatives are equivalent) in the context of mobile apps selection; and found that the framing effect had minimal impact on participants' trust perceptions [222]. A recent longitudinal Facebook study highlighted how profoundly Facebook interface changes can impact users' information sharing — indicating a potential for Facebook privacy nudges [223].

Several mechanisms have also been proposed to improve privacy decision making for social media. For example, Fang et al. described a wizard that creates sharing categories automatically [224]. Lipford et al. examined interfaces for online social network privacy controls, comparing expandable grids to visual policies [225]. Besmer et al. built a tool that allows Facebook users to negotiate about photo tagging [226]. Unlike these mechanisms, our approach aims to proactively nudge users away from posting potentially regrettable content.

9.3 Privacy Nudge Design

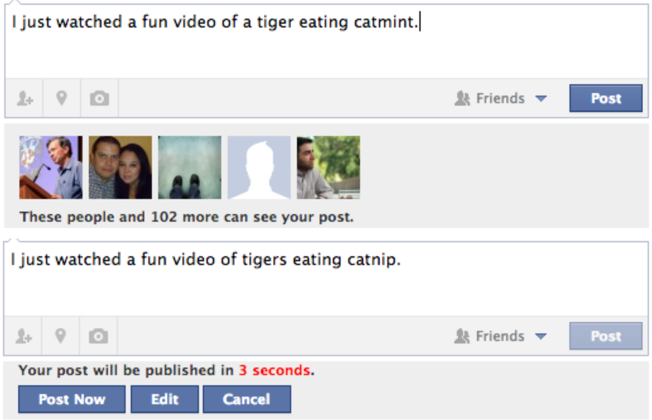
Our study focused on two types of nudges: one that reminds users about the audience for their post, and one that encourages users to pause and think before posting. In line with the concept of soft paternalism, neither of these nudges limits users' ability to disclose information, or affects the trade-offs associated with their disclosures. Instead, both nudges provide contextual clues intended to assist users in making better informed information-disclosure decisions. The nudge designs were improved based on two pilot studies conducted in April and July of 2012, respectively. We conducted a third pilot study in February 2013 to confirm that our experimental platform was stable enough to use in a field trial.

9.3.1 Audience Nudge

Inspired by the literature on bounded rationality and asymmetric information, the audience nudge aims to help Facebook users anticipate the audience for their posts.

This nudge initially included only the message: "these [number] people can see this post" based on the privacy setting of the post. However, after testing this feature in the first pilot study, we realized that participants did not find the intervention particularly useful, because the intervention message hardly varied over time — users often use the same privacy setting for their posts, and thus the message stayed the same.

In order to make the nudge more dynamic and salient, we considered using visual, rather than merely textual, cues. We designed a profile picture feature that displays five profile pictures, randomly selected from the pool of people who can view the post, based on that post's current privacy setting. This feature was inspired by Jenni and Loewenstein's work on the "identifiable victim effect," which finds individuals are willing to expend more resources on identifiable than unidentified victims [227]. In our context, the profile pictures provide some form of identifiability to both familiar contacts such as family members, close friends, and co-workers, as well as unfamiliar acquaintances or strangers, prompting Facebook users to think about who should see their posts.



I just watched a fun video of a tiger eating catmint.

These people and 102 more can see your post.

I just watched a fun video of tigers eating catnip.

Your post will be published in 3 seconds.

Figure 9.1: Audience+timer nudge. As the user types a post, five profile photos are displayed, selected randomly from people who will be able to see this post (top). After the user clicks "Post," a countdown timer appears and delays the post for ten seconds (bottom). During the count-down period users could edit or cancel the post, click a button to go ahead and post immediately, or do nothing and when the count-down expires the post will be made automatically.

We then combined both features (textual and visual information) into one "audience nudge," shown at the top of Figure 9.1. The nudge addresses two complementary aspects of audience perception: specific members of the audience, and the size of that audience. The pool of profile pictures and the audience size were determined by the privacy setting of the post. Our nudge implementation was able to detect and work with complex privacy settings of status updates such as "friends except certain people or groups" and "friends of friends." However, given the restrictions imposed by Facebook, the nudge cannot always precisely measure the audience size. In such cases, the message would provide qualitative rather than quantitative information, e.g., "These people, your friends, AND FRIENDS OF YOUR FRIENDS can see your post."

9.3.2 Timer Nudge

The timer nudge was inspired by literature on hyperbolic time discounting and dual process theory. It was designed to encourage users to pause and think (i.e., switch from System I to System II) before posting. It introduced a visual delay of 20 seconds after a user clicked the "post" button before publishing the submitted post. During the countdown, the user could cancel this post; otherwise the post was made automatically at the end of the countdown.

In the first pilot, participants found this nudge interesting, but also suggested reducing the time delay as well as allowing users to bypass the delay or edit the post. Accordingly, we reduced the delay to 10 seconds and added three links: "post now," "edit," and "cancel."

9.3.3 "Audience+Timer" Nudge

In the second pilot study, we tested the audience and timer nudges. Participants of this pilot study found both the picture and timer nudges somewhat helpful. However, some participants did not realize that the *post now, edit* and *cancel* links could be clicked [228]. To address this issue, we changed the design of these links to buttons that mimicked the Facebook look-and-feel. Furthermore, to create an improved Facebook privacy nudge, we combined the best aspects of two nudges we previously tested during the pilot into a single, new "audience+timer" nudge (Figure 9.1).

We built an experimental platform to both implement the nudge on Facebook and monitor how users interact with it. The platform consisted of a Facebook application to access users' Facebook data and a Chrome browser plugin to insert the nudge interface seamlessly into the Facebook interface.

In the second pilot study, we encountered a few technical issues. First, our nudges were not always shown. Second, our system did not reliably log users' Facebook behavior and interactions with the nudges. We discovered that both issues were partly due to significant changes introduced by Facebook after deploying our system. The lack of reliable behavioral logs prevented us from doing any quantitative analysis of the nudges [228]. We fixed these issues and implemented reliable logging of system events (e.g., display the nudge UI) and user behavior (e.g., click cancel). We then tested the updated system in the third pilot study. We identified some minor technical issues (e.g., Facebook pages loaded more slowly than usual, and some participants were unable to comment on certain public pages), but otherwise the system appeared stable and was able to log events reliably. We fixed these minor issues before the 6-week field trial. However, Facebook continued to make changes during our field trial and the comment problem reoccured.

9.4 Methodology

To evaluate the "audience+timer" nudge, we conducted a 6-week field trial with 28 Facebook users during April and May, 2013. We posted study ads on the Craigslist pages of the 12 most populated US cities as well as Syracuse and Pittsburgh. The ads directed prospective participants to a screening survey. The survey invited respondents to participate in the study if they met the following criteria: active adult US Facebook users who posted or commented at least once per day on average; native English speakers who posted in English and used Chrome, primarily, to access Facebook (because our platform was implemented for Chrome).

Study participants were required to install a Chrome plugin and an associated Facebook application. The 6-week field trial was divided into two phases. During the first three weeks (the *control period*) data collection took place without nudging interventions. At the end of the control period we asked participants to complete a mid-term survey to allow us to better understand participants' Facebook behavior during this period. We asked about unusual events, whether posts had caught the attention of unexpected audiences, and any regrets since the start of the study. During the remaining three weeks (the *treatment period*), in addition to data collection, all 28 participants were also shown our nudges. At the end of the treatment period, we asked participants to complete a final survey about their Facebook experiences during this period and their opinions of our nudges.

Participants were compensated with a $10 Amazon gift card for each week of participation and a $10 bonus for study completion. The study received IRB approvals at both Syracuse University and Carnegie Mellon University.

Since our nudges were highly dependent on the Facebook platform, we had to keep up with Facebook's frequent and unpredictable interface changes. To maintain our system during the 6-week study, we spent about an hour every day testing our system with all Facebook features it might interact with. We also had a programmer on standby to update the system if we found any issues. Despite these efforts, the system still encountered two technical problems during the six-week study. First, some participants were intermittently unable to post comments. This problem occurred most frequently for comments on public Facebook pages (e.g. for a company or celebrity). Second, participants experienced slow performance caused by our Chrome browser plugin.

9.5 Results

In this section we present the study results. We first describe participants' demographics and their posting frequency during the study. We then report participants' interactions with our nudges and changes to privacy settings made while our nudges were active. Finally, we present a detailed participant-level analysis looking into whether and how our nudges impacted each of our study participants.

9.5.1 Demographics and Facebook Posting Activities

Our participants included 19 females and 9 males between the ages of 19 and 51 (M=32, SD=10) from 16 U.S. states. All of them reported being active Facebook users, posting status updates or comments daily. Twelve (43%) self-reported having posted something on Facebook that they later regretted. Our participants came from a variety of occupations including medical staff, engineers, students, managers, teachers, homemakers, retired, and unemployed. Two had completed high school and the rest had at least some college education.

Table 9.1 summarizes participants' demographics, and the number of status updates and comments made during the control and treatment periods. We use a combination of one letter and one number to refer to each participant. As shown in Table 9.1, there is a large variability in the frequency of posting across participants. Overall, there is no obvious difference in posting frequency between the control and treatment periods. While our nudges might have impacted posting frequency, we cannot attribute those changes exclusively to the nudge. For example, participants explained that they had posted with unusual frequency, during both the control and treatment periods, due to factors such as vacations, illness, or new jobs. Thus, we do not use posting frequency to evaluate the impacts of our nudges.

ID	Gender	Age	Days in	Days in	Status updates		Comments	
			Control	Treatment	Control	Treatment	Control	Treatment
P1	F	49	18	22	1	0	171	68
P2	F	31	18	22	2	9	14	17
P3	F	28	18	22	2	5	57	158
P4	M	27	20	20	35	16	69	16
P5	F	30	18	22	20	24	31	50
P6	M	31	18	22	223	244	396	176
P7	F	45	19	21	45	87	822	950
P8	F	44	19	21	17	17	45	55
P9	F	39	18	22	1	7	14	18
P10	F	20	18	22	32	47	363	386
P11	M	23	18	22	5	9	52	27
P12	M	36	18	22	3	0	235	53
P13	F	50	18	22	1	3	14	15
P14	F	20	18	22	22	41	60	21
P15	F	23	18	22	4	17	4	12
P16	F	32	18	22	7	16	65	102
P17	M	28	21	21	2	5	144	17
P18	F	45	21	21	13	6	22	2
P19	F	19	21	21	86	37	290	132
P20	M	51	21	21	3	6	45	34
P21	F	28	21	21	34	33	77	48
P22	F	23	21	21	1	2	14	4
P23	F	27	21	21	10	7	34	42
P24	F	26	19	22	10	6	26	8
P25	M	27	19	22	4	1	3	3
P26	M	21	19	22	4	10	7	31
P27	F	22	19	22	20	1	46	11
P28	M	49	19	22	30	1	126	1
	Min	19	18	20	1	0	3	1
	Max	51	21	22	223	244	822	950

Table 9.1: Summary of participants' demographics and number of status updates and comments made during the control and treatment periods.

9.5.2 Interactions with Nudges

There are many ways a participant could interact with our nudges. We focused on four types of interactions with our nudges: hovering over profile pictures displayed by the audience nudge, and clicking *post now*, *edit*, and *cancel* buttons displayed by the timer nudge.

Hovering Over Profile Pictures

When a user hovers over the five profile pictures displayed by the audience nudge, the corresponding Facebook user's profile name appears. Twenty-four out of 28 participants hovered over profile pictures at least 3 times, and half of them did that throughout the treatment period, suggesting that most participants saw the pictures and interacted with them. However, participants only rarely clicked *edit* or *cancel* or changed the privacy setting after hovering over a profile picture. This suggests that participants were interested in identifying the people shown in the profile pictures, but generally did not feel the need to exclude them from seeing their posts.

Clicking Post Now

In the timer nudge, users could either wait for the post to be submitted automatically after the 10-second delay, or click the *post now*, *edit*, or *cancel* buttons. Twenty-four and 26 participants clicked *post now* at least once for status updates and comments, respectively. Participants clicked *post now* more often for status updates (64%) than comments (25%).

Clicking Edit

Participants used the *edit* button less than *post now,* but when they clicked *edit*, they did it more for comments (4.7%) than for status updates (1.7%). Eighteen participants clicked *edit* for comments, while only five participants did that for status updates. Seven participants clicked *edit* only once, while another 11 participants clicked it at least three times.

We also found that participants used the *edit* button in different manners. Clicking *edit* did not necessarily result in a different post, since in many cases the final post was not modified, suggesting that participants were using the *edit* feature to stop the timer and review their posts. A handful of participants ended up canceling their posts after clicking *edit*. Some participants used the *edit* option to correct typos, slightly rephrase, or complement their posts with additional information, while others made major changes to their original posts.

Clicking Cancel

We logged only seven cancellations for status updates (1.0%) and 15 for comments (0.6%) from eight participants. In some cases, participants refrained from submitting their post altogether, while in other situations they started a new rephrased post. In a few cases, as detailed in the per-participant analysis that follows, participants seemed to cancel potentially sensitive posts.

9.5.3 Privacy Settings Changes

Inline settings allow Facebook users to specify the audience for their status updates. Facebook users can select from a set of predefined groups (e.g., only me, friends, friends of friends, and public), create groups (e.g., high school classmates, co-workers, neighbors), and customize the setting to include or exclude specific people or groups. The setting remains the default for future status updates until it is modified again. The privacy setting of a comment inherits the setting from the corresponding status update. We expected that our nudges would help participants to more carefully select privacy settings for their status updates.

Six and eight participants modified their inline privacy settings during the control (number of changes per user: M=.6, SD=1.8) and treatment (M=.5, SD=.9) periods, respectively. Some of them changed the privacy settings both during the control and treatment periods, suggesting that our nudges were not necessarily associated with those changes. However, four participants changed their inline privacy settings only during the treatment period. In one of these cases a participant made his privacy settings more restrictive after hovering over a profile picture.

9.5.4 Interactions Over Time

To investigate novelty and habituation effects, we analyzed the temporal distribution of interactions with our nudges. Figure 9.2 shows the days on which participants clicked the edit and

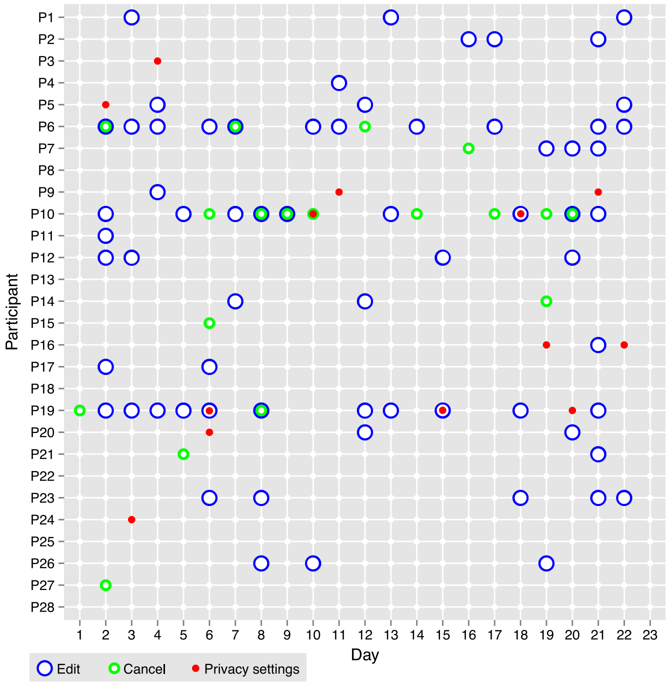


Figure 9.2: The days on which each participant clicked edit or cancel, or changed their privacy settings during the treatment period. The blue and green circles denote a participant who clicked edit or cancel at least once on that day, respectively. The red dot denotes a participant who has changed the inline privacy setting at least once on that day.

cancel buttons or made changes to their privacy settings during the treatment period. While there were six participants who did not exhibit any of these interactions and four who only interacted in the first few days, the majority of participants paid attention to and interacted with our nudges throughout the treatment period.

9.5.5 Participant-Level Analysis

To investigate the impact of our nudges on each participant, we analyzed each participant's interactions with our nudges as well as their survey responses. This allows us to understand why some participants liked the nudges and found them useful while others did not, and also helps us tease apart the impact of the audience and timer nudges. We categorized each participant into one of five descriptive groups defined by two dimensions: participants' attitudes and participants' level of interaction with our nudges.

Frequent Interactions and Positive Attitude

This group includes four participants (P4, P10, P20 and P23) who made extensive use of our nudges, and believed that at least one of the nudges could be helpful for themselves or others. These interactions include clicking the *cancel* or *edit* buttons and hovering over profile pictures.

P4 said, "I didn't post more or less, but I did post more cautiously. The constant reminder of who would be seeing my post was kind of an eye opener." He reported having used the time delay "to correct grammatical errors or statuses that looked 'off."' We found that while he often clicked the *post now* button, for a few posts he waited several seconds before clicking it, yet for others did not click it. He clicked *edit* for several comments, for example, he changed "long out" for "sign out" when writing about logging off of Facebook. He also hovered over pictures for many posts. For instance, after looking at the pictures, he checked the privacy setting without any change before posting, "I just started selling gold in the RMAH...I got a few auctions sold beforehand... :(" He told us that the nudges made him "a bit more aware. Especially the first day. That was almost the 'Oh wow' moment when I realized that more people could see my posts than I thought about." He also suggested that the nudges "should be default for all users."

P10 found the time delay helpful because it helped her avoid "getting into fights on Facebook because you have to stop and think." Despite experiencing some technical issues, her overall opinion was positive. She summarized, "I generally made better decisions." She normally clicked *post now* within 3 seconds, but later on she waited longer for a few of her posts. Besides, she clicked *cancel* for a few posts and then posted edited versions. For example, she canceled the status update, "not excited about still being sick wtf" and then posted, "not excited about still being sick after spending all afternoon in bed not doing my paper or having fun."

P20 canceled and edited a few posts. For example, he clicked *edit* and then ended up not posting: "Traded your Z in for that? :P." He also changed the privacy setting from public to only me after hovering over the profiles pictures shown when posting, "I've been nice up to this point, but the guy has to go! Eating all the bird seed. Where's my bebe gun?" In the final survey he said, "I think I was careful of what I said."

P23 found both nudges helpful. She explained, "I did like knowing when posts were going to be made public (like if a friend's wall is not protected to only their friends, etc.)" and elaborated, "I was going to respond to something snarky... I cancelled it because the application informed me that the entire internet could see my post." She also found the timer nudge helpful when she posted on other people's walls because it prevented her from "entering a discussion ... on someone's wall who posts religious or other annoying stuff all the time." However, she said she "was annoyed when I was using it to post to my own wall or to my close friends wall - those are not when I need the reminder." This suggests she might have preferred a nudge she could customize to fit her needs. Despite some technical glitches that prevented her from posting some comments, P23 had a positive opinion of the nudges overall because "It made me think twice about what I posted and who might see it." Our log data also show that she often hovered over the audience profile pictures and used the *edit* button to make minor changes to her posts.

Limited Interactions and Positive Attitude

Ten participants who had few interactions with our nudges stated that they thought at least one of the nudges could be useful for them or others (P8, P9, P11, P12, P15, P18, P21, P24, P25, and P26). For instance, P8 did not consider our nudges helpful to her, but said they could benefit "young people who are more likely to fly off the handle." She explained, "I didn't benefit, but trouble makers and kids would since it's an extra step and not just post and go-it may make someone think twice before posting hurtful comments." This participant used the friends privacy setting for all her posts and often clicked *post now* within 2 seconds. While she hovered over some profile pictures, she did not perceive any benefit from them since, "just wanna hit post and be done, not mess around with the delay or figure out who may or may not see it since

I have my privacy settings the way I want them." Nonetheless, she did edit some of her posts when she caught spelling or grammar errors.

While our logs show that P9 hovered over profile pictures, she reported not seeing any profile pictures. She based her opinion on the timer nudge and said she did not benefit much, but "it would be good for someone with a short temper." She said the nudge might be useful, "if I'd made a spelling error or tagged the wrong person." She clicked *post now* within 3 seconds on all her status updates and mostly used the friends privacy setting. In one case where she changed her privacy setting, she selected her Farmville friends group and explicitly excluded two friends before posting a Farmville-related request. She once clicked *edit* for a comment, but ended up reposting the same comment.

P11 reported that the profile pictures "helped me shape some of my posts." He further reported having canceled some posts because "I didn't want other people to think I'm stupid." However, he also expressed that, "the countdown timer annoys me a bit." He clicked *post now* within 2 seconds for all his status updates. Overall, he felt the nudge "made me think about what I was going to say."

P12 did not post any status updates during the treatment period and he used the public privacy setting for all his status updates during the control period. He clicked *post now* for most of his comments within 3 seconds and did not hover over pictures or click the *cancel* button. However, he clicked *edit* five times for his comments, two of which he ended up not posting, and three of which were reposted without changes. He concluded, the nudges are "not really for me, but I can see how it could be useful for others."

P24 found the audience nudge useful: "I think seeing all of the profile pictures made me rethink what I was going to post if it was slightly offensive or using curse words." However, the only interactions she had with ours nudges were hovering over pictures a few times. In the treatment period, she did change her privacy setting to exclude three friends when she posted, "So I bought a Rick Pitino Makers bottle for $50 and turned around and sold it for $180, lol..." Overall, she thought the nudge "made very little impact."

Limited Interactions and Negative Attitude

Three participants (P7, P14, and P22) neither interacted with our nudges, nor liked them. P7 was an active user who made over a thousand posts during the treatment period. She thought clicking *post now* was necessary to send her posts, which frustrated her. She explained, "I found it to be a pain because some of my posts I just like to post and go." She often clicked *post now* within a few seconds. Furthermore, while she hovered over profile pictures for about one third of her status updates, it did not seem to have any effect on what she posted as most of her status updates were public posts related to products, coupons, and promotions.

P14 also disliked the nudges. She did not click any buttons or hover over profile pictures when posting status updates. She clicked *post now* for three comments, and all were within three seconds. She also canceled one comment because "I was impatient!!" She explained, "I could see how the timer and the profile pictures would be beneficial but I just thought it was annoying."

P22 also did not like the nudges or interact with them. She only made six posts during the treatment period. She remembered seeing our nudges, but said that she did not pay attention to them. Our logs confirmed that she had no interaction with our nudges. When asked if there was any situation where the nudge negatively affected her, she replied, "for me, I don't care, so every time." She also seemed to post on Facebook just to get paid for participating in our study, posting the comment: "i'm only commenting on stuff b/c im being paid to by some app to spy on me and if i dont do enough social stuff they'll stop letting me do it."

Frequent Interactions and Negative Attitude

Seven participants did not like our nudges, but had extensive interactions with them (P2, P3, P6, P16, P17, P19, and P27). P2 experienced technical problems with our nudges and strongly disliked the timer. She complained, "the delayed posting thing which I HATE... makes posting harder." She believed that the timer is "not needed" and she did not "need anyone editing or censoring me." Although she indicated that she preferred the profile pictures over timer, she based her negative opinions on the timer. She explained, "I try to not put anything too embarrassing or horrible. I didn't mind that you were watching or anyone. I say what I want or feel," adding that, "there is no way to protect people from posting embarrassing or life impacting information online while mad or upset or whatever. It's human nature to be stupid sometimes." She often clicked the *post now* button when commenting, sometimes waiting several seconds before doing it and sometimes clicking it right away. For example, she waited 3 seconds before clicking *post now* for the comment: "David's burrito defeated him. It was HUGE," but waited 8 seconds for: "Off subject but the worst my arm pits have felt was during laser hair removal...Most painful area so far." She also clicked *edit* a few times without changing any comments, suggesting that she was using the *edit* option to take a second look at her comments. She also clicked *edit* to change one of her comments from "Or its like saying it transcends life. Negative Nelly aka Autumz..." to "Or its like saying it transcends life. I guess we all know your glass is half empty."

P6 hovered over profile pictures and clicked the *cancel* and *edit* buttons a few times. In addition, he often clicked *post now*, sometimes waiting a few seconds and sometimes clicking it right away. An active user, he made more than 400 posts during the treatment period. He thought that the plugin "made posting slightly more frustrating, but did not affect output." He explained that he was annoyed by "the fact I had to keep re-confirming that I wanted to post something I was sure about posting." This was another participant who thought that clicking *post now* was necessary to send his posts and did not realize that after the 10 seconds delay, his posts would be automatically posted.

P19 was another participant who frequently interacted with the nudges but did not like them. She said, "I didn't care about this feature at all, it did not affect my facebook usage at all, I just ignored it, like most people would." She also complained about encountering technical issues with the timer nudge. Interestingly, she acknowledged that she often regrets her posts, but said she solves the problem by just deleting them: "I almost always post things that I wish I wouldn't have, then I just delete them and the problem is solved." But, in some cases damage may have already been done before deleting a post. P19 did not seem to be concerned about this, and thus did not find utility in the nudges. While she disliked the nudge idea, she hovered over the audience nudge profile pictures 10 times, clicked *cancel* for a status update and a comment, clicked *edit* for 21 comments, and made two privacy setting changes during the treatment period. Most of her edits were minor rewording of the posts. For one lengthy status update about love and betrayal she changed the privacy setting to exclude one particular friend before she started typing her post, suggesting that she had a clear idea of her intended audience before she posted.

Overall, the nudges did not seem to help her avoid making potentially controversial posts, but did give her a chance to make minor edits of her posts.

Indifferent

Four participants expressed indifference about the nudges (P1, P5, P13, and P28). They either did not receive enough exposure to the nudges to form an informed opinion or simply expressed a neutral opinion even after having interacted with the nudges. Lack of exposure was due to participants posting less frequently or using different browsers or devices to post.

P1 reported having seen profile pictures and using the time delay to review her posts. She did not post any status update during the treatment period and often clicked *post now* within 3 seconds for her comments. She hovered over pictures before making several comments and clicked *edit* twice to correct typos. She was among those participants who had problems posting some of their comments which likely affected her overall opinion of the nudges. Although she used the nudges while posting and mentioned that the nudge could be useful "in case you are commenting on the wrong post," she was also not impressed and "could take it or leave it."

P5 neither expressed a positive nor a negative opinion and her behavioral data did not show any relevant interactions with the nudge as she only clicked *post now* right away. Similarly, P13 often clicked *post now* right away and did not have other interaction with the nudges. She explained, "In most cases I changed the time to posting to now so I didn't have to wait, I also didn't have to edit my posts because I wasn't saying anything I didn't want anyone to see." P28 hardly noticed the nudge as he only posted one status update and one comment using the plugin during the treatment period.

9.6 Discussion

The goal of our nudging interventions was to help users be more thoughtful when posting on Facebook, in order to reduce the potential for posting status updates or comments they might later regret. Consistent with the tenets of soft paternalism, we designed our nudges to encourage users to think about the audience and content of their posts without limiting their ability to post on Facebook. Here, we discuss participants' perceptions of our nudges and the challenges of conducting longitudinal field trials. We also offer recommendations for designing and testing nudges.

9.6.1 Varied Perceptions

Participants varied in how they perceived the general nudging idea and the two specific nudge features (audience and timer). Some participants were positive about the nudging intervention, e.g., P26 said "i like the plugin and think it is a great idea, i would love using this as the final product." In contrast, others had negative opinions. For example, P2 disliked the timer nudge because she did not "need anyone editing or censoring me."

We found that how participants use Facebook often played a role in their perceptions of our nudges. Generally, we found that those who use Facebook to post personal thoughts perceived the nudges as more beneficial than those who use it to broadcast news articles and other public information. On the other hand, those who use it for commercial or money making purposes (e.g. to share information about products and coupons) had negative opinions. We also found that those participants who had prior experience adjusting privacy settings and seemed to be careful about what they posted recognized the benefits of the nudges, but believed they did not need them.

9.6.2 Nudging toward Audience Awareness

The profile picture feature of the nudge was designed to remind Facebook users of the prospective audience for their posts. We found that most participants paid attention to and interacted with the profile pictures and several valued this feature, stating that it made them think about whether there post might offend someone.

Profile pictures were accompanied by a an indication of the number of people who could potentially see a post. Some participants said they found this information helpful, especially when posting comments on friends' posts.

In this study, we bundled the profile pictures and audience size information together. Further work is needed to determine their effectiveness in isolation.

None of the participants complained about the profile pictures, as they were less intrusive than the countdown timer. Users can ignore them, as some of our participants did, and go about their posting as usual. We found that most participants hovered over the profile pictures, and anecdotes from the final survey suggest that some of the participants benefited from having seen profile pictures. For example, one participant reported having decided not to post something after seeing the profile pictures. We also observed a participant change to a more restrictive privacy setting after hovering over a profile picture. This suggests that profile pictures can assist users in making better privacy decisions, but sometimes their effect can be subtle or difficult to measure.

9.6.3 Nudging with A Countdown Timer

The countdown timer was designed to encourage participants to stop and reflect on the content of their posts in order to avoid regrettable, "heat of the moment" posts. Participants were quite divided in their views. Some participants found the countdown timer valuable for giving them a chance to review their posts "a little more carefully" and "catch misspelled words or grammar errors." On the other hand, some participants voiced frustration about what they perceived as a requirement to take an extra step or wait 10 seconds to complete their posts.

We observed that the nudge was successful in helping some participants reconsider their posts. It had an additional benefit of helping users catch typos and other minor errors in their posts. A number of participants rephrased or even canceled their posts during the delay. However, this benefit came at the cost of delaying every post. The timer countdown was both the most liked and disliked nudge feature we implemented.

For the most part, participants were not as concerned about their posts being delayed for 10 seconds as they were about having to wait 10 seconds before they could move on to something else. In reality there was no need for them to watch the countdown, but some participants seemed unwilling to trust that their post would get posted after 10 seconds and others thought they *had* to click *post now*. Our timer nudge seemed to leave some participants feeling uneasy and afraid to move on until their massage got posted. We could design an interface where the post would appear on the user's screen as if it had been posted (perhaps by posting it to "only me") while the timer is counting down. Even though the nudge would still have the same functionality, a change in the visual display might make it more acceptable.

The idea of delays may be applied in other scenarios where people may benefit from some extra time to think about their actions. However, since the time delay interrupts primary tasks (in this case Facebook posting), it should be used selectively and with caution. Future research should explore other ways to slow people down and encourage them to think about their actions, as well as ways to introduce time delays more selectively.

9.6.4 Challenges and Limitations

Conducting our investigation as a longitudinal field trial allowed us to investigate the impact of our nudges under real life conditions, but that made our study more challenging to run and resulted in a number of limitations.

One source of limitations stemmed from implementation challenges. Since Facebook changes its interface frequently and unpredictably, we had to constantly monitor, test, and adapt our code to keep up with those changes. Despite this concerted effort, our nudges malfunctioned a few times during our field trials. For instance, some participants experienced some of their comments not being posted. In addition, since our nudge was not an integral part of the Facebook platform, we had to work around the Facebook UI to embed the nudge features. We also had to add logging functionality to capture all possible user-driven events. However, this extra logging slowed down our nudges and indirectly affected some participants' Facebook experience. These technical challenges made it difficult to run our field trial for an extended period of time. In addition, some participants disliked our nudges primarily because of these issues.

As with any research involving observation of participant behavior, one methodological concern is the Hawthorn effect: participants may change their behavior simply because they are in a study. To mitigate this we minimized our interactions with participants once the study began. In the mid-term survey, we explicitly asked them whether their posting behavior had changed. Some participants noted posting with different frequencies due to various reasons, but un-related to our study. Several participants also reported in the final survey that they thought our nudges were introduced by Facebook rather than us.

External factors beyond our control and observation likely affected participants' posting behavior, making it difficult to determine causality. Similarly, measuring the effectiveness of our nudges in preventing regret is also challenging because generally only a small fraction of the posts made by users may lead to regret, and arguably even fewer lead to the short-term regret we may detect in this study. In addition, it is often difficult to measure the effect of a nudge; users may not react to them in a noticeable way, or the reaction might be gradual.

While our combination of different nudge features might increase the chance that we detect some effect of our nudges, it makes it difficult for us to isolate the effects of individual nudge features or account for interactions between features.

9.6.5 Implications for Designing and Testing Privacy Nudges

We identified a set of key aspects to consider when designing and evaluating privacy nudges. First, designers should consider the intrusiveness of the nudges. We found that our less-intrusive audience nudge was better received by users than our more-intrusive timer nudge. On the other hand, we observed more direct benefits from the timer nudge. It would be useful to investigate whether the timer nudge could be improved by making it less visually intrusive — for example, showing a user's post actually posted but visible to "only me" until the end of the countdown.

Second, designers should keep in mind that some users will dislike the sense of being watched. Designers should look for ways to nudge people without making them feel that a new "big brother" is watching. In our second pilot study users disliked being judged by a nudge that provided subjective feedback on the sentiment of the users' posts [228]. The nudges we tested in this study, on the other hand, were not perceived as judgmental by participants.

Third, designers should consider the extent to which they should allow users to control or customize a nudge. In our system, we did not give users any ability to control the nudges except for the *post now* button that allowed them to skip the time delay. Some users wanted to be able to turn off the nudges or personalize them according to their needs and preferences. Controls could allow users to configure nudges such that they are enabled only under certain circumstances, such as at specific times, when certain people can see their posts, or when they type certain sensitive words.

Fourth, it is critical that the nudges function properly and do not interfere with the usability and reliability of the system in which they are embedded. Our nudges suffered from technical glitches that decreased their perceived value for some participants. However, without help from Facebook, we found it difficult to improve the reliability of our system.

Lastly, but importantly, nudges are difficult to evaluate both quantitatively and precisely when they are designed to impact behaviors that may occur only occasionally, or that may be hard to observe. And yet, when it comes to privacy, it could be precisely occasional, rare behaviors that end up causing the most damage — for example, a spur-of-the-moment status update that leaves a long and painful trail of unintended consequences. Collecting enough measurable, quantitative data to compute aggregate results from a small sample of users is difficult under such circumstances. Unless it is feasible to study a large number of users, an evaluation strategy including qualitative participant-level analysis is likely to provide more informative results than a quantitative analysis.

9.7 Conclusions

While the field study we presented in this paper should be considered exploratory, our results suggest that privacy nudges have the potential to be a powerful mechanism to assist users in avoiding unintended disclosures. Although our findings come from a Facebook case study, the principles underlying the privacy nudges we tested may be extended to similar services such as Twitter or to other types of services such as e-commerce, location sharing, and smart phone applications.

Chapter 10

Public Policy Implications

In this chapter, I put together insights gained through our research and turn them into considerations for policy making. I first discuss OBA-specific recommendations in Section 10.1, then discuss broader considerations in Section 10.2. These two sections are intended to assist both industry and government policy makers with the design of more effective notice and choice tools. Finally, in Section 10.3 I argue for the need for stricter government regulation to protect online privacy.

10.1 Improving OBA Notice and Choice

We have extensively studied Internet users' privacy expectations and information needs in the context of OBA. We have further evaluated the usability of various OBA notice and choice tools, and analyzed in detail the content of advertising companies' privacy policies. The following considerations are provided in the context of OBA; however, they can also inform the design of notice and choice mechanisms in other domains.

10.1.1 Different kinds of notice and choice methods are necessary

Industry and government have given much emphasis to privacy policies as a way to provide users with notice about companies' practices. However, research has shown that users don't read or understand privacy policies, and it has been estimated that if Internet users read them, it would cost more than the cost of accessing the Internet itself. Furthermore, notice and choice face important challenges in light of new technologies that allow automatic and seamless collection and analysis of users' information [229]. Online Behavioral Advertising (OBA) is an example of these new technologies. I argue that different audiences can benefit from different types of privacy notices. While traditional privacy policies with legalistic language are necessary to make companies accountable for their practices, more usable privacy notices can be used to empower users.

From a users' perspective, I recommend requiring advertising companies and websites to implement three types of privacy notices: privacy icons, privacy summaries, and interactive notices with privacy choices.

- **Privacy Icon.** A conspicuous privacy notice in the form of a meaningful icon should be provided on websites. The icon should convey at least one of the following four situations: No tracking exists on the website; tracking exists only for website customization and user analytics (no sharing); tracking exists for advertising purposes without involving users' personal information and with limitations on the use of sensitive inferences, sharing only with reliable entities, and limited retention period; and tracking exist for other purposes (e.g., advertising, marketing, or any other purpose) without explicit collection, sharing, and retention limitations. Such an icon should be placed in a consistent and salient place (e.g., at the top of the webpage) and have an appropriate size and shape, allowing users to notice the icon and realize they can click on it. Furthermore, the icon should not be placed in the boundaries or inside ads as not all tracking is necessary related to advertisement and it could also mislead users into thinking that the icon is part of the ad (as discussed in Chapter 7). A tooltip should be added to the icon, succinctly explaining its purpose and encouraging users to click on it to learn details. Consistent icon location and shape across websites are important to educate users gradually about its purpose and benefit.

- **Privacy Summaries.** When applicable (i.e., when tracking exists on the website), this notice should be linked from the privacy icon and should contain a concise summary to make it easy for users to quickly assess the risks and determine if they want to take any action. Based on our results from Chapter 8, the privacy summaries should inform about the following: what the purpose of tracking is; whether or not sensitive information (e.g., health conditions, income range, location, etc.) is being collected or inferred from users' activities; whether or not the information used or collected for tracking purposes can be linked to users' identity; whether or not that information is shared with non-affiliates; and whether or not those non-affiliates can link received information with users' identity. In addition, this notice should provide a link to a webpage where users can exercise their privacy choices. Such link should be labeled properly to communicate that users can benefit from clicking on it. For example, the label can say, "If you don't like any of these, change your privacy settings here." As in the case of the icons, it is also important for the design of privacy summaries to be standardized to gradually educate users about their purpose and benefit, and to facilitate comparison of websites' practices.

- **Interactive Notice with Choices.** The need for better choices was discussed in Chapters 5 and 8. When applicable (i.e., when tracking exists on the website), a third notice linked from the privacy summary should provide detailed information regarding what has been collected or inferred about the user. This third notice should also provide choice mechanisms to allow users to remove whatever information they don't want advertising companies to know about their online activities; provide the opportunity to express a preference to not be tracked at all; and provide the opportunity to express a preference to collect only certain information or make certain inferences, but not others. Providing users with access to the information collected or inferred about them is also important because it enables users to visualize the effect of data aggregation, enabling them to assess the risks more realistically.

While these recommendations are focused on the practice of OBA, similar notices can be used in other application domains. For example, interactive notices as described above can take the form of privacy dashboards for those companies that collect users' personal information, providing more effective notice and choice for users. Similar notices can also be used in smartphones to inform about apps' privacy practices and allow users to make privacy choices, if they want to. It is important to mention that specific notice designs and implementations need to be informed by usability testing as discussed in Section 10.1.3, and will likely take different forms for different kinds of devices. However, their design and implementation should be standardized across websites regardless of the different third-party advertising companies operating on those websites.

10.1.2 Notices are useless without meaningful choices

In the online advertising industry, efforts have focused on providing notices without presenting users with meaningful choices. If users believe that they don't have choices that allow them to protect their privacy, they will ignore privacy notices. Therefore, the notices themselves should communicate that consumer-relevant choices exist. Unfortunately, the main choice that the advertising industry is offering is not one that users are willing to make. Specifically, ad companies that follow self-regulation guidelines are required to offer users the opportunity to opt out of targeted ads, but not necessarily of being tracked. Nevertheless, users are more concerned about being tracked than they are about receiving targeted ads and, in fact, users prefer targeted ads to random ads. Advertising companies have argued that to prevent fraud and limit the number of times the same ad is shown to a particular user (i.e., frequency capping) they must keep collecting users' online activities even if users opt out. While it is reasonable that companies need to establish operational controls, they should be clear that if users express a desire to not be tracked, data collected for operational purposes will not be aggregated in any form that enables tracking across websites or sold to third-parties for any purpose. Based on our results from Chapters 5, 6, and 8, I recommend providing choices that empower users to:

- **Stop online tracking altogether.** Privacy concerned users who don't want their online activities to be monitored should be able to stop this practice.

- **Stop online tracking selectively.** Some users may perceive benefits from receiving customized ads based on some of the webpages they visit. However, most users are uncomfortable with their activities being tracked across every website they visit. In addition, even if users don't perceive any benefits from customized ads, most users understand that ads pay for free content, hence some users may be willing to let ad companies track them in some websites, but not others. In Section 10.1.1, I have proposed four different icons to represent four online tracking situations that pose different privacy risks. Users could further be empowered by allowing them to express their preferences to allow any or none of the icon-represented online tracking situations.

- **Know and correct what has been learned.** Users have expressed a need for transparency and believe that it is their right to know what is known about them and be able to edit it. Our results suggest that users are willing to edit their own advertising profiles to both receive relevant ads and protect their privacy. It is also important to mention that giving access to user profiles shouldn't require identifying users, as access can be offered based on computer identifiers.[1]

- **Express permanent preferences.** Current opt-out approaches based on cookies are problematic because users have learned that deleting cookies is "good" for privacy and users may unintentionally delete opt-out cookies when attempting to protect their privacy. Therefore, other mechanisms to permanently keep users' privacy preferences should be used.

For those users willing and able to express more granular preferences, I recommend further considering the following:

- **Express inference preferences.** Ad companies are able to create detailed users' profiles based on tracked users online activities. However, in many cases the profiles contain information that is perceived as personal and sensitive for users. Allowing users to specify what inferences tracking companies can and cannot make about them can alleviate some of the privacy concerns that we have identified. Furthermore, if a user has not expressed inference preferences, only non-sensitive information should be used for targeted ads.

- **Express PII linkage preferences.** In many cases, ad companies have the ability to link anonymous tracking data with users' identities. For example, some ad companies are also service providers and can collect contact and other personal information in a first-party context (e.g., when they offer a direct service to the user). Ad companies can further receive users' personal data from other third parties. Given the privacy implications of personally identifying tracked users, ad companies should not link anonymous tracking data with personal identifiers unless users grant their explicit consent.

- **Express data sharing preferences.** Particularly important is the sharing with entities that can link the received data with other data streams (including offline data) and with personal information. Data brokers are examples of such entities, whose practices can negatively impact users' privacy. Therefore, this kind of sharing should be prohibited unless the users have granted their explicit consent. Furthermore, users who have granted such consent should also be offered clear choices to limit it later if they want to.

The fact that the online advertising ecosystem has become complex with many participating players should not mean that users have to bear the burden of understanding the privacy risks. I argue that the ad industry should innovate and redesign its system in a way that it is possible to control how users' information is used and shared. This will not only enable users with the control they demand and increase their trust, but will also make the industry more efficient.

Moreover, implementing meaningful choices for users is challenging in the context of third-party tracking, where users don't have an explicit interaction with tracking companies. In our studies, users expressed a need to have privacy choices integrated in the tools that they already use, including web browsers and antivirus software. This suggests that advertising companies could work together with websites and other stakeholders (e.g.,service and software providers, web browser companies, etc.) to design and implement meaningful and usable privacy tools integrated in users' daily Internet experience.

[1] See for example the Exelate opt-out page at: `exelate.com/privacy/opt-in-opt-out/`

10.1.3 Users should be taken into account

In order to design effective notice and choice tools that the average Internet user can understand and benefit from, it is important to take into account users' information needs, skills, and expectations. Companies may be tempted to assume that non-expert Internet users will be able to understand and use the notice and choice tools that companies design in the same way that experts do. This is not true, not only in the privacy, but also in any other domain. Specifically, in Chapter 5, we have shown that users are often not aware that ad companies track their online activities and that they don't notice OBA notices. In Chapter 7, we have further shown that users misunderstand OBA notices and are afraid to click on them. Currently, those unnoticed and misunderstood notices are the only way users may have access to opt-out choices. Moreover, the usability study described in Chapter 6 demonstrated that current tools to limit online tracking are challenging to use and are often misused by Internet users interested in protecting their online privacy. Therefore, usability testing should be performed before any tool is claimed to provide effective notice and choice. The U.S. Government has proposed a multi-stakeholder process to develop privacy tools and codes of conduct; however, as of now, the multi-stakeholder processes have struggled to implement user testing [230]. An example of a more effective multi-stakeholder process in the financial domain was presented in Chapter 3, where we discussed how U.S. regulators, after recognizing that financial institutions' privacy policies were not empowering consumers, decided to engage industry and researchers to design standardized notices that users may understand and act upon. I believe that the same idea can be used to design effective notice and choice methods in other domains, including the online advertising one. In this work, I have demonstrated different methods including, in-person interviews, laboratory testing, field studies, and large-scale online studies. These methods are well-known by researchers and companies interested in user-interface design and usability. I recommend that ad companies and other interested entities use these methods to evaluate their privacy notice and choice tools.

10.1.4 Take it or leave it approaches won't work

Research has demonstrated that privacy preferences are contextual. In Chapter 8, we also demonstrated that users' privacy preferences in the context of OBA are shaped by different factors including advertising companies practices, the type of information that is collected and used to deliver ads, the expected benefits, and the perceived risks of being tracked. Furthermore, we showed in Chapters 5 and 6 that if users are asked to signal an opt-out preference or to limit online tracking on a per-company basis, they are not able to make meaningful choices because users are not familiar with the names of tracking companies or know their privacy practices. These results suggest that binary approaches like Do-Not-Track or Opt-Outs, which do not allow users to consider the context and ad companies' current disclosure practices, are less than optimal. Current binary approaches are further problematic because privacy-concerned users may end up attempting to block all tracking by using extreme measures (e.g., using tools that block all third party-content or opting out from all companies) giving up any potential benefits for themselves (e.g., customized content or relevant ads) and also affecting advertising companies' profits. Providing users with relevant information, allowing them to express nuanced choices based on types of collected information and companies' practices, and establishing limits to riskier practices (e.g., collection or inference of sensitive data types or linkage of tracking data with personal information) will better protect people's privacy, while giving the ad industry the opportunity to regain users' trust.

10.2 Additional considerations to improve notice and choice

I now discuss other non-OBA specific considerations to improve notice and choice; which can be applied to OBA and other domains.

10.2.1 Notice and choice have limitations

Notice and choice are important elements to protect privacy. They allow users to make privacy trade-offs that better align with their preferences. If properly implemented, they have the potential to shape companies' information privacy practices and lead the market to an equilibrium between users' expectations and companies' practices. Nevertheless, as discussed in Chapter 9, behavioral biases, cognitive limitations, and user-interface designs can prevent users from acting consistently with their stated privacy preferences. In particular, while Internet users are concerned about privacy, their primary goal is not to protect their privacy when they are online. Their goal is to benefit from accessing the Internet, may it be shopping, entertainment, searching for information, connecting with friends, etc. As a result, users have to rely on heuristics (i.e., rules of thumb) and contextual cues (e.g., feedback and user-interfaces) to make privacy decisions as they accomplish their primary goals. However, research has shown that those heuristics and contextual factors can negatively affect privacy decisions, leading privacy concerned users to behave against their stated preferences, a phenomenon ofter referred to as the privacy paradox. With these facts in mind, I recommend policy makers to consider the following:

- **Notices nudge users.** Research has shown that user-interface design plays an important role in shaping users' privacy behaviors. Therefore, notice designers should be aware that the notices themselves can lead users to take actions that don't match their privacy expectations. For example, notices can highlight the benefits of giving up personal information or allowing tracking, while downplaying the potential risks. Furthermore, notices can include cues that make users believe that their privacy is protected. For example, while conducting the study presented in Chapter 4, we found that many ad companies' privacy policies start with sentences like "we care about your privacy," or "your privacy is important for us," potentially leading users to believe that the company does actually protect users' privacy, even if that may not be the case. Furthermore, in Chapter 9, we designed notices with the goal to encourage Facebook users to make more thoughtful disclosures; however, research has shown that changes in user-interface design over time has led Facebook users to disclose more [223]. Therefore, it is important to consider what behavior would be generally best for users based on users' stated preferences and objective risks. The notices can then be designed to encourage such desirable behaviors, while allowing users to make different choices if they knowingly want to.

- **Defaults matter.** Behavioral research has shown that users normally stick to defaults. In Chapter 6 we found that most participants never changed the default settings of the tools they tested and believed that such settings were already limiting online tracking.

Therefore, defaults also play an important role in shaping users behaviors and should be properly selected to protect users' privacy.

- **Privacy controls don't always help.** Providing more privacy controls to users may not necessarily improve privacy as users may suffer from overconfidence biases and end up exhibiting riskier privacy behaviors [231]. Users can further become overwhelmed with excessive privacy controls leading them to make suboptimal decisions. In Chapter 6 we found that the large number of privacy settings offered by some of the tested tools confused participants, leading them to choose settings that did not match their stated preferences.

- **Feedback is important.** In Chapter 6, participants were wary of opt-out tools that did not provide any indication of protecting participants' privacy. To make notice and choice tools more effective, users should receive appropriate feedback informing about the consequences of their choices. Furthermore, indicators similar to the privacy icons described in Section 10.1.1 can provide timely and visual feedback communicating potential privacy risks.

- **Users' beliefs matter.** In Chapter 7, we found that participants were unwilling to click on OBA icons because they believed that such action would enable tracking or would signal a preference for the advertised product. Researchers have also found that Internet users mistakenly believe that websites that have privacy policies protect their privacy [21]. Therefore, privacy notice design should further consider users' mental models to prevent users from misusing or misinterpreting the notices they are shown.

Overall, presenting users with privacy notices and choices, even if those are conspicuous and clear, is not a guarantee that users would be able to act in their best interests. Subtle user-interface cues can impact users' behaviors in unintended ways and it is important to consider that the designers of notice and choice tools play an important role in shaping users' choices. Therefore, notice and choice mechanisms should not be overused and complementary privacy protections discussed in Section 10.3 should be put in place.

10.2.2 Standardized privacy notices are powerful

Traditionally, privacy notices take the form of long and difficult-to-read privacy policies. Those policies, while necessary to make companies accountable, don't provide transparency or empower users to compare companies' practices. In Chapters 2 and 3, we evaluated standardized notices from websites and financial institutions, respectively. There, we showed and discussed the benefits of machine-readable and human-readable standardized notices. Specifically, standardized notices have the following desirable characteristics:

- **Ease comparisons.** Human-readable standardized privacy notices that have a standard structure and terminology are easier for users to understand and compare.

- **Enable automatic checking.** Machine-readable privacy notices, if properly implemented, can relieve users from reading many long policy documents by using software agents that check those standard privacy notices against users' a priori stated preferences [37].

- **Enable large-scale evaluations.** Standardized notices improve transparency by enabling large-scale evaluations and comparisons of practices. Human-readable standardized notices allow users to compare companies' practices at a small scale; however, machine-readable standardized notices can take this step further, by allowing automatic evaluations at a large scale. Results of those evaluations can then be made publicly available, empowering users to find companies with practices that align with their privacy expectations.
- **Compliance with Notice Requirements.** Regardless of the regulatory regime (e.g., regulation or self-regulation), oversight organizations define the requirements that covered entities should follow, including disclosure of collection, sharing and opt-out practices. When standardized notices are in use, internal (e.g., internal auditors) and external oversight entities can use automatic tools to verify compliance with established notice and choice requirements. While notices don't necessary reflect companies' practices, a company using a sloppy notice, may deserve a deeper investigation of its practices. In addition, forcing companies to be complaint with notice requirements can also motivate them to be more diligent reviewing and understanding their actual practices.

Standardized notices have the potential to improve transparency, delivering benefits for users, companies with privacy-respectful practices, and oversight entities. Companies can further use semi-automatic notice generators, assisting companies with the creation of compliant standardized notices. Such generators can take input regarding companies' practices through standard questionnaires and therefore generate notices that align with actual practices. Machine-readable notices offer additional important scalability benefits over human-readable standard notices. Moreover, apart from providing better transparency, machine readable notices can further the development of current efforts to build semantic data flow protocols, such as OASIS XDI [232], which can be used to automatically enforce companies' practices.

10.3 Government regulation is necessary

Online privacy has been globally recognized as a matter of social importance. The risks of online tracking have been discussed for decades, but users' privacy concerns have increased over time and companies' practices have prioritized business needs over privacy. The U.S. Government has already recognized the privacy risks of the uncontrolled practices of online tracking and data aggregation companies, recommending that regulations should be created to protect users' privacy [233]. Our work also suggests that government regulations are necessary in the domain of online tracking and behavioral advertising and I recommend to consider the following aspects:

- **Incentives.** Although OBA self-regulation organizations have made efforts to protect users' privacy, those efforts have mostly being made to mitigate the risks of government regulation, and not necessarily to improve users' privacy. As a result, as shown in this work and elsewhere, those efforts have been largely ineffective. The ad industry does not have the right incentives to develop more privacy-respectful practices, including enhanced notice and choice methods. It has mostly focused on improving the financial benefits of online advertising, but has overlooked important privacy-related risks for society. I believe governments are in a better position to assess both benefits and risks at the society level and create the right incentives accordingly. Specifically, under the current self-regulatory regime, companies don't have any incentives to become members of self-regulatory organizations. In fact, companies sometimes have more incentives to not affiliate with those self-regulatory bodies as they won't need to follow any guidelines neither offer opt-out opportunities. Furthermore, companies seem to not face any negative consequences from not joining self-regulatory bodies. In Chapter 4, we used a publicly available database of online tracking companies to show that only about 20% of them had affiliations with the main online advertising self-regulation entities in the U.S. I further argue that regulations can also incentivize the development of innovative privacy-respectful technologies, for example by subsidizing companies that commit to design and implement those technologies. Overall, I argue that most companies would not have incentives to follow privacy-protective practices (as the ones discussed earlier in this chapter) or design their own ones unless there is legislation or civil liability to incentivize them.

- **Accountability and enforcement** are necessary to guarantee that companies operate as mandated. Even the best designed privacy guidelines are useless if we can't have certainty that companies comply with them to a reasonable extend. Unfortunately, the current self-regulatory environment doesn't provide reasonable accountability and enforcement guarantees. In Chapter 2, we showed that about 30% of more than thirty thousand evaluated websites used invalid P3P compact policies, and hundreds of them misused them by inaccurately representing their full P3P or human-readable policies. The P3P is a machine-readable standard proposed by the industry more than a decade ago. Importantly, when large website were confronted with these findings, they recognized that they were non-complaint and argued that the standard, which was created by the industry itself, was not suitable for their business' needs. In contrast, in Chapter 3, we found that only about 1.5% of the evaluated standardized financial privacy notices, which are government-regulated, had mistakes. And, when these financial institutions were confronted, several recognized that their notices had errors, but told us that their actual practices were compliant. In Chapter 4, we further showed that privacy policies of advertising companies don't fully comply with self-regulatory guidelines and other instances of non-compliance had been found previously [17]. In sum, it is unclear how non-compliant companies are punished or forced to abide by the rules currently. Therefore, I propose that government regulation should incorporate elements of accountability and enforcement that self-regulation has failed to incorporate.

- **Baseline privacy protections** are necessary to prevent unintended and negative consequences. The industry has argued that imposing collection or use limitations can affect companies's ability to operate efficiently and innovate. However, it is important to set clear limitations based on the potential risks. Potential risks identified by privacy researchers include, market manipulation, discrimination, and infringing on people's autonomy [101, 102, 103]. I further believe that online tracking can increase the probability of other financial-related risks such as identity-theft.

- **Guiding privacy principles.** Self-regulatory organizations have defined high-level privacy principles; however, those principles strongly favor companies' business priorities over users' privacy expectations. In particular, current OBA self-regulation principles allow practices that this research shows are problematic from a user's standpoint. Therefore, regulators can establish principles that better balance businesses' profit needs with society's needs. In addition to notice and choice requirements, government-guided principles can also encourage other FIPPs-based practices such as data security, data quality and access, recourse, remedies, accountability and enforcement.

- **Privacy standards.** Regulations should encourage the design and implementation of standardized notices and privacy protocols that can further both transparency and enforceability of companies' practices. In particular, mandated standardized privacy notices can improve transparency by empowering anyone to analyze companies' practices at large scale and making that information accessible for users. Current efforts are been made to analyze privacy policies using natural language processing algorithms and crowd sourcing [50, 98, 99]; however, as discussed in Chapter 4, significant challenges exist for these efforts to succeed. Regulators can help researchers and other stakeholders in improving transparency by mandating the design of both human-readable and machine-readable standardized notices.

I finalize my discussion on government regulation with an analogy between fossil fuels pollution and privacy injuries [234]. The trade-offs between privacy risks and benefits of personal data uses are difficult to assess in a similar way as are those trade-offs between environmental risks and benefits of using fossil fuels. Furthermore, personal data leaks are similar to pollution leaks in two ways: they can occur massively (i.e., a data breach versus a big oil spill) or at small, but constant rate (i.e., small pieces of information being constantly leaked versus CO_2 being constantly released to the environment). In the long term, both privacy and environmental harms have important negative consequences for society. That is, while privacy harms affect innovation, people's autonomy and democratic values, pollution affects humans' health. I believe that in the same way that the U.S. Government have come to realize that it is important to regulate and control fossil fuels pollution, it can also come to the conclusion that legal protections are needed to prevent undesired privacy harms. I further believe that the research presented in this thesis provides evidence that government-regulation is necessary in the domain of online privacy.

Chapter 11

Conclusions

I have presented a set of case studies covering different aspects of privacy notice and choice in four domains: online behavioral advertising (OBA), online social networks (OSN), financial institutions' standardized privacy notices, and websites' machine-readable privacy notices. I sought to answer three general research questions: 1) What is the effect of industry self-regulation and government-regulation on companies' implementations of privacy notices?, 2) What are the strengths, weaknesses, and opportunities for improvement of current online notice and choice mechanisms? and 3) Can contextual, in-time notices mitigate users' regrettable information disclosures? We investigated the first of these questions in Chapters 2, 3, and 4, concluding that websites and advertising companies, which operate in self-regulatory environments, have fewer incentives to be transparent and comply with notice requirements than financial institutions, which operate in a government-regulated environment. The second of these questions was explored within the scope of three of the studied domains: OBA, financial institutions' notices, and websites' privacy policies. In Chapter 10, I extensively discussed how to improve notice and choice mechanisms in general. In the next section, I further discuss those improvements within the scope of each application domain. We investigated the third and last of these questions in Chapter 9, concluding that contextual and in-time notices have the potential to assist Facebook users with their information disclosures and mitigate unintended consequences.

I now present my conclusions, followed by a set of research questions to be considered for future work.

11.1 Domain-specific conclusions

Investigating notice and choice mechanisms in different application domains allowed us to identify opportunities for improvement in each of these domains, as well as to provide recommendations to improve notice and choice mechanisms in general. In this section I discuss the main findings and conclusions of each of the domains studied.

11.1.1 Websites' Machine-readable Privacy Disclosures

In Chapter 2, we investigated websites' usage of Platform for Privacy Preferences (P3P) compact policies (CPs). We found that about a third of more than thirty thousand evaluated websites had invalid CPs. A large fraction of those were using exactly the same P3P CPs that had been recommended in Internet blogs to bypass the Internet Explorer browser's CPs-based cookie filtering feature. Many other websites, while using semantically and syntactically valid CPs were misrepresenting their human-readable policies. We concluded that P3P as a voluntary notice standard is ineffective because of two potential reasons. First, the lack of an appropriate privacy-governance structure may lead to misalignments between the technical implementation of privacy protections such as P3P CPs and companies' privacy protection strategy (if any). And second, the lack of enforcement may allow websites to misused an industry privacy standard without facing any legal consequences.

11.1.2 Financial Standardized Privacy Disclosures

In Chapter 3, we evaluated standardized privacy notices from more than six thousand U.S. financial institutions. We found that information sharing practices were statistically correlated with three main factors: company size, geographic location, and the type of company. We also found that even among similar types of companies, there were relevant differences with respect to consumer information sharing practices. For example, while large banks located in the north-eastern region of the U.S. were the ones that overall shared the most for marketing and other purposes, there were large banks in the same region that had more privacy-protective practices. We concluded that standardized privacy notices have an enormous potential to improve transparency and empower users to select companies that better align with their privacy preferences. We further made a set of recommendations to improve the implementation of online financial standardized notices and incentivize their use.[1]

11.1.3 OBA

In the context of OBA, we conducted laboratory (i.e., usability testing and interviews) and large-scale online user studies to investigate Internet users' privacy expectations and ability to understand and use notice and choice mechanisms. We further evaluated online advertising companies' implementation of human-readable notices on which Internet users rely to make privacy decisions. In Chapter 4, we conducted a detailed analysis of 75 online tracking companies' privacy policies from three sets: 1) the largest tracking companies (e.g., those that track the most); 2) companies that are members of either of the two largest online advertising self-regulatory organizations in the U.S. (Digital Advertising Alliance and Network Advertising Initiative); and 3) non-member companies. We found that very few companies (20%) of a total of 2,750 in Evidon's database listed self-regulation affiliations and that not all studied member companies comply with self-regulatory principles. We further found that overly generic OBA self-regulatory guidelines allowed member companies to be compliant with many of the principles, without offering real transparency or privacy protections for users. All three sets of companies were silent about consumer-relevant practices including, the sharing of users' collected data with third parties that can link users' tracking data with users' identity, and the use of sensitive data categories for targeting advertising. We concluded that, regardless of self-regulation mem-

[1] We submitted comments on the Consumer Financial Protection Bureau (CFPB)'s recent proposal to change GLBA.

bership status, online tracking companies are neither transparent nor do they offer meaningful choices to users. We discussed how to make these policies more transparent and usable.

In Chapter 5, we conducted 48 semi-structured interviews to study Internet users' perceptions and awareness of online behavioral advertising (OBA) and understanding of OBA privacy icons (i.e., AdChoices and Interest-based Ads). We found that participants were mostly unaware of online tracking for advertising purposes. After learning about it, participants expressed concerns regarding the lack of transparency and control. Many participants also feared the collection of personal data for advertising purposes. Furthermore, participants were unaware of tools to limit OBA and expected their web browser or antivirus software to provide the privacy protections needed to limit it. We concluded that while users might not dislike targeted ads, better user education and opportunities to control OBA are important to both protect users' privacy and build trust in this practice.

In Chapter 6, we evaluated the usability of nine popular tools to limit OBA. We tested three general types of tools: 1) tools that block online tracking; 2) tools that allow users to opt out of targeted ads; and 3) privacy settings of two popular browsers: IE9 and Firefox. We found several usability problems with all of the tested tools, including inappropriate defaults, lack of feedback, excessive use of technical jargon, and confusing user-interfaces. We also found that participants' mental models of cookies conflict with opt-out methods based on cookies, leading participants to believe that they would further protect their privacy if they deleted their web browser cookies after having opted out. We identified weakness and strengths of these tools and provided recommendations for improvement.

In Chapter 7, we conducted a large-scale online study to evaluate Internet users' reactions to and understanding of OBA privacy disclosures and OBA opt-out pages. We evaluated the two taglines widely used by the advertising industry (i.e., AdChoices and Interest-based Ads) in addition to five other alternatives. We also tested five opt-out pages from popular ad companies. We found that while half of the participants remembered the ads that were shown to them, only 12% of the participants correctly remembered the accompanying OBA taglines. The AdChoices tagline was particularly ineffective at providing notice about OBA, compared with the alternatives tested. In general, participants were afraid to click on OBA disclosure icons, believing that if they clicked on them, they would be shown more ads or would be tracked. We further found that after reading opt-out pages, participants were confused about the meaning of opting out. We concluded that a main challenge to the effectiveness of OBA disclosures is that users don't understand what OBA is to begin with, suggesting that OBA disclosures should first educate users about OBA and then provide choices that users can understand and use. Furthermore, OBA disclosures should not be placed inside or nearby ads that users are afraid to click, but in a websites' consistent and salient location.

In Chapter 8, we discussed two large-scale online studies investigating how different advertising companies' practices affect users' willingness to share information. After showing explicit disclosures summarizing different advertising companies' practices regarding scope of collection and sharing, purpose, data retention, and access to study participants, we collected participants' willingness to share different data types under the shown conditions. We found that, with the exception of access, the outlined practices affected, to different extents, participants' willingness to share information. In the second of these studies, we further looked into participants' decision making process, asking them about the reasons they would or would not share specific data types. Finally, we investigated participants' reactions to a sample (but realistic) advertising profile, showing the types of information and inferences that ad companies can create about Internet users. We found that participants were not opposed to targeted ads, but were concerned about the collection of sensitive and personal information that was, according to them, irrelevant for targeted ads. Participants were concerned about the extend of inferences and personal details included in the sample profiles. Participants also perceived various benefits from having access to their own profiles, including the ability to, know what is known about them, remove information they don't want others to know, and make profiles more accurate to receive relevant ads. We concluded that current ad companies' practices don't align with users' privacy expectations in at least two ways: ad companies often make inferences about users related to information that participants did not want to disclose; and ad companies are able to personally identify tracked users, which, with a few exceptions, participants didn't want. The results also suggest that binary approaches like Do-Not-Track or opt-outs, which do not consider the context of collection or ad companies' practices are less than optimal.

11.1.4 Online Social Networks

In the context of OSN, we conducted field studies with active Facebook users to explore the effect of nudging notices designed to encourage more thoughtful disclosures. In Chapter 9, we designed and evaluated a new type of privacy notice, which we refer to as a soft-paternalistic or "nudging" notice. This was motivated by research in the fields of behavioral economics and privacy decision making, which has shown that humans' cognitive limitations and behavioral biases can lead people to make decisions against their own privacy preferences. We studied these nudging notices in the Facebook domain instead of the OBA domain because we wanted to study their effects in situations where users could make actual disclosure decisions as opposed to hypothetical ones. Currently, users can't make meaningful choices about OBA. Furthermore, we wanted to disambiguate the effect of the nudging notice from the novelty effect of OBA, which as discussed above, many users are still not aware of. In this study, we modified the Facebook user-interface to add feedback about the audience of each post and also add a 10-second delay to each post participants made, allowing users to edit or cancel their posts during the delay, if they wanted. We collected study participants behavioral data and attitudes about the nudging notice. We found that reminders about the audience of posts can prevent unintended disclosures without major burden; however, introducing a delay before publishing users' posts can be perceived as both beneficial and annoying. We concluded that nudges have the potential to mitigate unintended disclosures, but it is important to align their design with users' particular needs and privacy preferences.

11.2 Overarching conclusions

Providing users with privacy notices and the opportunity to make privacy choices is important to allow users to make the privacy trade-offs that best align with their preferences. A necessary condition for targeted transparency to work in privacy and others domain is that users should be able to make meaningful choices based on companies' disclosures. Information should empower users with information that is relevant for them and should give users the ability to compare companies' practices. Therefore, I investigated users' privacy preferences, information needs and ability to exercise choices in the OBA domain. This allowed me to provide recommendations to improve the design of notice and choice methods currently in use in this application domain.

I have also demonstrated how standardized notices can improve transparency and benefit users, privacy-respectful companies, and oversight entities. Specifically, machine-readable notices enable large-scale evaluations and comparisons of companies' stated practices, allowing users to find companies with better practices and regulators (and companies themselves) to verify compliance with transparency requirements. Furthermore, machine-readable notices have the potential to improve enforcement and integrate with current efforts to build standard semantic data flow protocols that can allow for automatic enforcement.

I have further argued that given the limitations of notice and choice in practice, users should not be charged with full responsibility to protect their online privacy. These limitations are particularly relevant to situations where users are not aware that their privacy may be at stake, as is the case with third-party online tracking. Additional FIPPs-based protections including data security, quality and access, as well as redress, accountability and enforcement should be considered.

Finally, I have provided empirical evidence of the need for stricter government regulation in the domain of online tracking, suggesting recommendations for improving the designing of notice and choice mechanisms, both in the OBA and other domains.

11.3 Opportunities for future research

We have extensively explored notice and choice methods in the OBA domain and selectively studied websites' and financial institutions implementations of notice and choice mechanisms. Moving forward, the following research questions could be explored:

- **Financial privacy tradeoffs.** What is the relationship between financial institutions' privacy practices and the quality and cost of services offered to consumers? How can we incorporate privacy information with other consumer-relevant information? How would users make privacy versus cost/quality of service tradeoffs?

- **Standardized notices.** Is it possible to design usable and standardized (human-readable and machine-readable) notices for other domains with important privacy implications, including OBA, websites, health, mobile and smart home devices?

- **Users' privacy agents.** Can we apply statistical techniques to model users' privacy preferences under different contexts and configure devices' or services' privacy settings accordingly?

- **Privacy dashboards.** In the context of the Internet of things and big data, can we design privacy dashboards to disclose personal data uses in meaningful ways for users? How can we design effective user-interfaces to collect users' contextual preferences?

- **Parsing human-readable policies.** Can we extract user-relevant practices from long privacy polices? How salient and clear websites' privacy notices impact users' decision making?

Appendix A

The Misrepresentation of Website Privacy Policies Through the Misuse of P3P Compact Policy Tokens

A.1 Description of P3P Compact Policy Tokens

Element	Token	Full P3P Vocabulary	Plain Language Translation of P3P Policy Element [35]
Access	NOI	<nonident/>	We do not keep any information identified with you
	ALL	<all/>	We give you access to all of our information identified with you
	CAO	<contact-and-other/>	We give you access to your contact information and some of our other information identified with you
	IDC	<ident-contact/>	We give you access to only your contact information in our records
	OTI	<other-ident/>	We allow you to access some of our information identified with you, but not your contact information
	NON	<none/>	We do not give you access to our information about you
Disputes	DSP	There are some disputes	There are ways to resolve privacy-related disputes with us
Remedies	COR	<correct/>	We will correct any errors we make related to the commitments in our privacy policy
	MON	<money/>	We will compensate individuals if it is determined that we have violated our privacy policy
	LAW	<law/>	Our privacy policy references a law that may determine remedies for breaches of our policy

Continued on next page . . .

Continued from previous page

Element	Token	Full P3P Vocabulary	Description
Non-Identifiable	NID	<NON-IDENTIFIABLE/>	We do not keep any information that could be used to identify you personally
Purpose	CUR	<current/>	To provide the service you requested
	ADM[attr]	<admin/>	To perform web site and system administration
	DEV[attr]	<develop/>	For research and development, but without connecting any information to you
	TAI[attr]	<tailoring/>	To customize the site for your current visit only
	PSA[attr]	<pseudo-analysis/>	To do research and analysis in which your information may be linked to an ID code but not to your personal identity
	PSD[attr]	<pseudo-decision/>	To make decisions that directly affect you without identifying you, for example to display content or ads based on links you clicked on previously
	IVA[attr]	<individual-analysis/>	To do research and analysis that uses information about you
	IVD[attr]	<individual-decision/>	To make decisions that directly affect you using information about you, for example to recommend products or services based on your previous purchases
	CON[attr]	<contact/>	To contact you through means other than telephone (for example, email or postal mail) to market services or products
	HIS[attr]	<historical/>	To aid in historical preservation as governed by a law or policy described in this privacy policy
	TEL[attr]	<telemarketing/>	To contact you by telephone to market services or products
	OTP[attr]	<other-purpose/>	For other uses described in the site's human readable policy
Recipient	OUR	<ours/>	Companies that help us fulfill your requests (for example, shipping a product to you), but these companies must not use your information for any other purpose
	DEL[attr]	<delivery/>	D elivery companies that help us fulfill your requests and who may also use your information in other ways
	SAM[attr]	<same/>	Companies that have privacy policies similar to ours
	UNR[attr]	<unrelated/>	Other companies whose privacy policies are unknown to us
	PUB[attr]	<public/>	People who may access your information from a public area, such as a bulletin board, chat room, or directory
	OTR[attr]	<other-recipient/>	Companies that are accountable to us, though their privacy policies may be different from ours

Continued on next page . . .

Continued from previous page

Element	Token	Full P3P Vocabulary	Description
Retention	NOR	<no-retention/>	We do not keep your information beyond your current online session
	STP	<stated-purpose/>	We keep your information only long enough to perform the activity for which we collected it
	LEG	<legal-requirement/>	We keep your information only as long as we need to for legal purposes
	BUS	<business-practices/>	Our full privacy policy explains how long we keep your information
	IND	<indefinitely/>	We may keep your information indefinitely
Categories	PHY	<physical/>	Name, address, phone number, or other physical contact information
	ONL	<online/>	Email address or other online contact information
	UNI	<uniqueid/>	Website login IDs and other identifiers (excluding government IDs and financial account numbers)
	PUR	<purchase/>	Information about your purchases, including payment methods
	FIN	<financial/>	Financial information such as accounts, balances, and transaction history
	COM	<computer/>	Information about the computer you are using, such as its hardware, software, or Internet address
	NAV	<navigation/>	Which pages you visited on this web site and how long you stayed at each page
	INT	<interactive/>	Activities you engaged in at this web site, such as your searches and transactions
	DEM	<demographic/>	Information about social and economic categories that might apply to you, such as your gender, age, income, or where you are from
	CNT	<content/>	Messages you send to us or post on this site, such as email, bulletin board postings, or chat room conversations
	STA	<state/>	Cookies and mechanisms that perform similar functions
	POL	<political/>	Which groups you might be a member of such as religious organizations, trade unions, and political parties
	HEA	<health/>	Health information such as information about your medical condition or your interest in health-related topics, services, or products
	PRE	<preference/>	Information about your tastes or interests
	LOC	<location/>	Information about an exact geographic location, such as data transmitted by your GPS-enabled device
	GOV	<government/>	Government-issued identifiers such as social security numbers

Continued on next page . . .

Continued from previous page

Element	Token	Full P3P Vocabulary	Description
	OTC	<other-category/>	Other types of data described in the site's human readable policy
Test	TST	<test/>	The CP is under test

Attributes [attr]: a = always, i = opt-in, o = opt-out

A.2 Observed frequency of tokens

Access	NOI	28%	NA	NA
	ALL	6%	NA	NA
	CAO	44%	NA	NA
	IDC	12%	NA	NA
	OTI	<1%	NA	NA
	NON	5%	NA	NA
Disputes	DSP	61%	NA	NA
Remedies	COR	54%	NA	NA
	MON	<1%	NA	NA
	LAW	3.4%	NA	NA
Non-identifiable	NID	6%	NA	NA
Purpose	CUR/CURa	58%	NA	NA
	ADM	73%	1%	1%
	DEV	70%	1%	1%
	TAI	37%	2%	<1%
	PSA	52%	16%	2%
	PSD	32%	1%	1%
	IVA	11%	27%	2%
	IVD	2%	23%	3%
	CON	1%	30%	12%
	HIS	3%	<1%	<1%
	TEL	<1%	1%	22%
	OTP	1%	22%	<1%
Recipient	OUR	96%	NA	NA
	DEL	3%	22%	<1%
	SAM	1%	22%	2%
	UNR	<1%	21%	<1%
	PUB	<1%	21%	<1%
	OTR	2%	23%	15%
Retention	NOR	7%	NA	NA
	STP	20%	NA	NA
	LEG	1%	NA	NA
	BUS	13%	NA	NA
	IND	67%	NA	NA
Categories	PHY	39%	NA	NA
	ONL	40%	NA	NA
	UNI	56%	NA	NA
	PUR	31%	NA	NA
	FIN	23%	NA	NA
	COM	61%	NA	NA
	NAV	61%	NA	NA
	INT	39%	NA	NA
	DEM	49%	NA	NA
	CNT	25%	NA	NA
	STA	45%	NA	NA
	POL	21%	NA	NA
	HEA	21%	NA	NA
	PRE	31%	NA	NA
	LOC	15%	NA	NA
	GOV	21%	NA	NA

A.3 Evaluation of full P3P and human-readable policies for web sites using the CP suggested by the O'Reilly blog

URL	Valid Full P3P Policy?	Location of Human-readable Policy	Comments
alleghenyinstitute.org	NO	Not found	**No policies found to compare with CP**
bordellfuehrer.de	NO	Not found	**No policies found to compare with CP**
caidep.com	NO	*caidep.com*	**Policies do not match** - privacy policy does not mention any information associated with the *NOI*, *ADM*, *DEV*, *PSAi*, *COM*, *OTRo*, *STP*, *IND* or *DEM* tokens included in the CP; privacy policy mentions the use of cookies to store preferences and to perform customization but CP does not include *PRE* or *TAI* tokens
cakephp.org	NO	*cakephp.org/pages/ privacy*	**Policies do not match** - privacy policy does not mention any information associated with the *NOI*, *ADM*, *DEV*, *PSAi*, *STP*, *IND* or *DEM* tokens included in its CP; privacy policy mentions the use of cookies to store preferences but CP does not include *PRE* token
campbell.house.gov	NO	Not found	**No policies found to compare with CP**
condusef.gob.mx	NO	Not found	**No policies found to compare with CP**
creditolo.de	NO	Not found	**No policies found to compare with CP**
dme.kerala.gov.in	NO	Not found	**No policies found to compare with CP**
economics.harvard.edu	NO	Not found	**No policies found to compare with CP**
equestrian.com.my	NO	*equestrian.com.my/ privacy-policy*	**Human-readable policy does not mention cookies**
gilldivers.com	NO	Not found	**No policies found to compare with CP**
gss.ucsb.edu	NO	Not found	**No policies found to compare with CP**
honor.unc.edu	NO	Not found	**No policies found to compare with CP**
itech-ny.com	NO	*itech-ny.com/ privacy-policy.html*	**Policies do not match** - privacy policy does not mention any information associated with *NOI*, *ADM*, *DEV*, *STP*, *IND* or *DEM* tokens included in its CP; privacy policy mentions the use of cookies to store preferences but CP does not include the *PRE* token
joomla.org	NO	*joomla.org/ privacy-policy.html*	**Human-readable policy does not mention cookies**
komodorock.com	NO	*www.komodorock.com/ privacy-policy/*	**Policies do not match** - privacy policy does not mention any information associated with *NOI*, *ADM*, *DEV*, *STP*, *IND* or *DEM* tokens included in CP; privacy policy mentions the use of cookies to store preferences and customize advertising but CP does not include the *PRE* or *TAI* tokens

Continued on next page ...

Continued from previous page

URL	Valid Full P3P Policy?	Location of Human-readable Policy	Comments
laser.org	NO	Not found	**No policies found to compare with CP**
majorleague.com.au	NO	Not found	**No policies found to compare with CP**
megasearch.net	NO	*megasearch.net/ PrivacyPolicy.html*	**Policies do not match** - the use of cookies is not well detailed in the human-readable policy.
navicat.com	NO	*navicat.com/en/ privacy.html*	**Human-readable policy does not mention cookies**
ocean.tamu.edu	NO	*geosciences.tamu.edu/ about-us/ geonet-information-hub/ web-site-policies/ 677-site-privacy-and-security-policy*	**Policies do not match** - privacy policy does not mention any information associated with *NOI*, *DEV*, *PSAi*, *OTRo*, *STP*, *IND* or *DEM* tokens included in CP
orange-pocket.com	NO	Not found	**No policies found to compare with CP**
parktrust.org	NO	Not found	**No policies found to compare with CP**
rcn.com	NO	*rcn.com/dc-metro/ privacy-policy*	**Policies do not match** - the use of cookies is not well detailed in the human-readable policy - privacy policy does not mention any information associated with *NOI*, *ADM*, *DEV*, *PSAi*, *STP*, *IND* or *DEM* tokens included in CP; privacy policy mentions that cookies are used to provide seamless visit and expedite customer login but CP does not include *CUR* token
relevantmagazine.com	NO	*relevantmagazine.com/ privacy-policy*	**Policies do not match** - the use of cookies is not well detailed in the human-readable policy
themacstore.com	NO	*themacstore.com/ privacy/*	**Human-readable policy does not mention cookies**
theories.com	NO	*theories.com/index.php/ Privacy-Policy.html*	**Policies do not match** - the use of cookies is not well detailed in the human-readable policy
topcities.com	NO	Not found	**No policies found to compare with CP**
womensmedia.com	NO	*womensmedia.com/new/ privacy-policy.shtml*	**Human-readable policy does not mention cookies**
wsashow.com	NO	*wsashow.com/homepage /privacy_policy*	**Policies do not match** - privacy policy does not mention any information associated with *NOI*, *DEV*, *STP* or *IND* tokens included in its CP; privacy policy mentions that cookies are used for log-in, enable personalization, analytics, shopping cart, personalized service, and targeted advertisement, but CP does not include *CUR*, *IVD*, *IVA*, *TAI* or *INT* tokens

A.4 Evaluation of full P3P and human-readable policies for Websites with CP errors in top 50 most-visited list

URL	Valid Full P3P Policy?	Location of Human-readable Policy	Comments
facebook.com	NO	*facebook.com/policy.php*	**Policies do not match** - CP contains only two tokens (*DSP* and *LAW*) but privacy policy mentions that cookies are used for several purposes, including the provision of services, advertising, easy log-in, etc., and that cookies are stored for an extended period
msn.com	NO	*privacy.microsoft.com/ en-us/fullnotice.mspx*	**Slight differences between CP and privacy policy** - privacy policy mentions that cookies may be used to collect demographic information but CP does not include *DEM* token
safety.live.com	NO	*privacy.microsoft.com/ en-us/fullnotice.mspx*	**Slight differences between CP and privacy policy** - privacy policy mentions that cookies may be used to collect demographic information but CP does not include *DEM* token
amazon.com	NO	*amazon.com/gp/help/ customer/display.html /ref=footer_privacy/ 191-3583711-6331321? ie=UTF8&nodeId=468496*	**Invalid CP, unable to compare**
microsoft.com	YES	*privacy.microsoft.com/ en-us/fullnotice.mspx*	**Slight differences between CP, full P3P policy, and privacy policy** - privacy policy mentions that cookies may be used to collect demographic information but CP and full P3P policy do not include *DEM* token
reference.aol.com	NO	*about.aol.com/aolnetwork/ aol_pp*	**Policies do not match** - privacy policy mentions cookies are used to remember preferences, measure ad effectiveness, customize site, store demographic information, share info with ad networks and service providers, but CP does not include any *PRE*, *ONL*, *TAI*, *DEM* or *SAM* tokens.
atlas.mapquest.com	NO	*about.aol.com/aolnetwork/ aol_pp*	**Policies do not match** - privacy policy mentions cookies are used to remember preferences, measure ad effectiveness, customize site, store demographic information, share info with ad networks and service providers, but CP does not include any *PRE*, *ONL*, *TAI*, *DEM* or *SAM* tokens.

Continued on next page . . .

Continued from previous page

URL	Valid Full P3P Policy?	Location of Human-readable Policy	Comments
godaddy.com	NO	*godaddy.com/Agreements/ShowDoc.aspx?pageid=PRIVACY&ci=20803&app_hdr=0*	**Policies do not match** - privacy policy mentions the collection of name, address, credit card numbers, government IDs, and collected information might be used to contact the user and to present co-branded offers on opt-in basis, but CP does not contain PHY, DEM, GOV, CON or $SAMo$
imdb.com	NO	*imdb.com/privacy*	**Invalid CP, unable to compare**
windows.com	NO	*privacy.microsoft.com/en-us/fullnotice.mspx*	**Slight differences between CP and privacy policy** - privacy policy mentions that cookies may be used to collect demographic information but CP does not include DEM
hulu.com	NO	*hulu.com/privacy*	**Policies do not match** - privacy policy mentions targeted advertising based on user's activity but CP does not include IVA; policy states: "We may use cookies and similar technologies to relate your use of the Hulu Services to personally identifiable information," yet CP includes the NID token, claiming that they do not collect PII

A.5 Top visited domains using CPs

Domain	Valid Full P3P Policy?	Errors found
about.com	YES	None
amazon.com	NO	Invalid tokens; Missing tokens
angelfire.com	NO	Invalid tokens; IVA and CON conflicting tokens
aol.com	NO	Missing tokens
apple.com	NO	None
att.com	YES	None
bing.com	NO	None
bizrate.com	NO	Invalid tokens
blogspot.com	NO	None
careerbuilder.com	YES	IVA conflicting token
causes.com	NO	IVD, IVA and CON conflicting tokens
cnet.com	NO	None
cnn.com	NO	None
comcast.net	NO	None
dailymotion.com	NO	None
examiner.com	NO	NID conflicting token
facebook.com	NO	Missing tokens
flickr.com	NO	None
go.com	NO	Invalid tokens; Missing tokens
godaddy.com	NO	IVD and IVA conflicting tokens
google.com	NO	None
hulu.com	NO	Invalid tokens
ign.com	NO	None
imdb.com	NO	Invalid tokens; Missing tokens
latimes.com	YES	None
linkedin.com	NO	None
live.com	NO	Missing tokens
mapquest.com	NO	Missing tokens
match.com	NO	None
metacafe.com	NO	None
microsoft.com	YES	CON conflicting token
monster.com	NO	None
msn.com	NO	Invalid tokens
mybloglog.com	NO	None
nytimes.com	NO	Missing tokens
people.com	NO	None
simplyhired.com	NO	None
target.com	NO	None
thefind.com	YES	IVD and IVA conflicting token
tripod.com	NO	Invalid tokens; IVA and CON conflicting tokens
tumblr.com	NO	None
twitter.com	NO	None
washingtonpost.com	NO	TEL, IVD, IVA, and CON conflicting tokens
weatherbug.com	NO	None
wikipedia.org	NO	None
windows.com	NO	Invalid tokens
yahoo.com	YES	None
yellowpages.com	NO	None
Total	7/48	21/48

A.6 Network advertising domains using CPs

Domain	Valid Full P3P Policy?	Errors found in CP
247realmedia.com	YES	None
adsfac.sg	YES	None
atdmt.com	YES	None
casalemedia.com	NO	None
imiclk.com	YES	None
intellitxt.com	NO	None
navegg.com	NO	Invalid tokens
realmedia.com	YES	None
vizu.com	YES	None
weborama.fr	NO	None
zedo.com	NO	None
Total	6/11	1/11

A.7 Domains holding TRUSTe seals using CPs

Domain	Valid Full P3P Policy?	Errors found in CP
10kscholarship.com	NO	None
1800mobiles.com	NO	None
192.com	YES	None
1choice4yourstore.com	NO	None
247realmedia.com	YES	None
2fixyourtrafficticket.com	YES	None
3dcart.com	NO	None
abc.com	NO	None
abcnews.com	YES	None
activeinternational.ca	NO	IVD, IVA, and CON conflicting tokens
activeinternational.com	NO	IVD, IVA, and CON conflicting tokens
adt.com	NO	None
agilent.com	NO	Missing tokens
alladvertisingagencies.com	NO	None
aloharents.com	NO	None
alvenda.com	NO	None
amiastri.com	NO	Invalid tokens
angelfire.com	NO	Invalid tokens; IVA and CON conflicting tokens
aol.com	NO	Missing tokens
apothica.com	YES	Missing tokens; TEL and CON conflicting tokens
appexchange.com	NO	Missing tokens
apple.com	NO	None
asksanta.ca	NO	Missing tokens
att.com	YES	None
att.net	YES	Invalid tokens
attinteractive.com	NO	None
automationcontrols.com	NO	None
autonation.com	NO	None
avaline.com	NO	None
aviationarthangar.com	NO	None

Continued on next page . . .

Continued from previous page

Domain	Valid Full P3P Policy	Errors found in CP
bellsouth.com	NO	None
bic-gsa.com	NO	None
bicgsa.com	NO	None
bicwarehouse.com	NO	None
bidezone.com	NO	Missing tokens
billhighway.com	NO	Invalid tokens; Missing tokens
billiardsaddiction.com	NO	None
billshrink.com	NO	IVD, IVA, and CON conflicting tokens
bing.com	NO	None
bizrate.com	NO	Invalid tokens
bizrate.de	NO	Invalid tokens
bluerazor.com	NO	IVD and IVA conflicting tokens
bodymedia.com	NO	None
boostflow.com	NO	None
boston.com	NO	None
burstnet.com	NO	None
buyingadvice.com	NO	None
buysafe.com	YES	IVD, IVA, and CON conflicting tokens
buysafeshopping.com	YES	IVD and IVA conflicting tokens
caliberlocal.com	NO	None
calibex.com	NO	None
candlewoodsuites.com	NO	Missing tokens
candywarehouse.com	NO	None
caoh.org	NO	None
careonecredit.com	YES	None
carid.com	NO	None
casalemedia.com	NO	None
caspio.com	NO	Missing tokens
caspio.net	NO	Missing tokens
cataloglink.com	NO	NID conflicting token
catchfirefunding.com	NO	None
cellstores.com	NO	None
ceu4u.com	NO	None
chatthreads.com	NO	Missing tokens
cheaptickets.com	NO	IVD and IVA conflicting tokens
chefsresource.com	NO	None
chegg.com	NO	None
chipin.com	NO	None
christmastreeforme.com	NO	None
cjhomeandoffice.com	NO	None
classmates.com	YES	IVD, IVA, and CON conflicting tokens
clcleather.net	NO	None
clubbing.com	NO	Missing tokens
code7contest.com	YES	None
comcast.net	NO	None
conair-store.com	NO	None
concreteexchange.com	NO	None
controlscan.com	YES	None
coremetrics.com	YES	None
costumecity.com	YES	None
couponbug.com	NO	None
coupons.com	NO	None
couponsinc.com	NO	None
cpp.com	NO	Missing tokens

Continued on next page . . .

Continued from previous page

Domain	Valid Full P3P Policy	Errors found in CP
crafta.com	NO	None
credit.com	NO	None
criteo.com	NO	None
crowneplaza.com	NO	Missing tokens
cufflinksdepot.com	NO	None
datepad.com	NO	Missing tokens
dealsonhotels.com	NO	Missing tokens
debtgoal.com	NO	None
depositagift.com	NO	None
dexclusive.com	NO	None
digicert.com	YES	CON conflicting token
digilifestudios.com	NO	None
digitalimaginghq.com	NO	None
digitallanding.com	NO	IVD, IVA, and CON conflicting tokens
digitalspyders.com	NO	None
directfix.com	NO	None
directtextbook.com	YES	None
dreamlandweddingshoppe.com	NO	None
drugs.com	YES	None
duiattorney.com	NO	None
dynamiclogic.com	YES	None
e-miles.com	NO	None
e-rewards.com	NO	None
e-rewards.de	NO	None
e-rewards.fr	NO	None
e-rewards.nl	NO	None
ea.com	NO	Missing tokens
earnmydegree.com	NO	None
ebates.com	NO	None
ebooks.com	YES	Missing tokens; CON conflicting token
ecampustours.com	NO	None
echosign.com	NO	IVD, IVA, and CON conflicting tokens
educadium.com	NO	None
educationconnection.com	YES	IVD, IVA, and CON conflicting tokens
emeraldisland.com	NO	None
emergingmed.com	NO	None
enjoycpr.com	NO	None
epals.com	NO	None
eprooft.com	NO	None
espn.com	YES	None
ether.com	NO	None
eversave.com	NO	None
facebook.com	NO	Missing tokens
familyfun.com	NO	None
fansnap.com	NO	None
federaldebtreduction.com	NO	None
firstagain.com	NO	CON conflicting token
flemingoutdoors.com	NO	None
forzieri.com	NO	None
freeshop.com	NO	NID conflicting token
genealogytoday.com	NO	None
getaroom.com	NO	None
getinsurancequotes.ca	NO	None
globesmart.com	NO	None

Continued on next page . . .

Continued from previous page

Domain	Valid Full P3P Policy	Errors found in CP
go.com	YES	Invalid tokens; Missing tokens
godaddy.com	NO	IVD and IVA conflicting tokens
gotomypc.com	NO	Missing tokens
gowearfit.com	NO	None
greenfieldonline.com	NO	None
greensherpa.com	NO	None
greenwayuniversity.com	NO	None
grovesite.com	NO	None
healthscout.com	YES	None
healthsquare.com	YES	None
hiexpress.com	NO	Missing tokens
higherone.com	NO	None
holiday-inn.com	NO	Missing tokens
homedecorhardware.com	NO	None
homegain.com	NO	None
hotbot.com	NO	CON conflicting token
hotelindigo.com	NO	Missing tokens
houstontexans.com	NO	None
htmlgear.com	NO	Invalid tokens; IVA and CON conflicting tokens
hyperstreet.com	NO	None
ibm.com	YES	None
ichotelsgroup.com	NO	Missing tokens
ideascale.com	NO	None
ifriends.net	YES	None
ifriendsv2.net	YES	None
ihg.com	NO	Missing tokens
ihgarmyhotels.com	NO	Missing tokens
importedblankets.com	NO	None
inksell.com	NO	None
inoutcash.com	NO	None
insightexpress.com	NO	None
intelius.com	NO	Missing tokens
intercontinental.com	NO	Missing tokens
intuit.com	NO	Missing tokens
itech-ny.com	NO	None
itwixie.com	NO	None
jackpotrewards.com	NO	Invalid tokens
jaman.com	NO	Missing tokens
jameslimousines.com	NO	None
jewelrywonder.com	NO	Missing tokens
jobtarget.com	NO	Invalid tokens; Missing tokens
justasktoday.com	NO	None
kanetix.ca	NO	None
kanetix.com	NO	None
karmacar.com	YES	None
keen.com	NO	None
keysurvey.com	NO	Missing tokens
kinglinen.com	NO	None
largestmall.com	NO	Missing tokens
legalmatch.com	NO	None
letstalk.com	NO	None
life360.com	NO	None
lifequote.com	NO	None
linkedin.com	NO	None

Continued on next page . . .

Continued from previous page

Domain	Valid Full P3P Policy	Errors found in CP
listyourdebt.com	YES	None
lithium.com	NO	None
live.com	NO	Missing tokens
livemeeting.com	NO	CON conflicting token
loanio.com	NO	None
logcap4jobs.com	NO	None
lycos.com	NO	CON conflicting token
maghound.com	NO	None
mail2world.com	NO	Missing tokens; IVA and CON conflicting tokens
mailchimp.com	NO	Invalid tokens; Missing tokens
market2lead.com	YES	None
mate1.com	NO	None
maven.net	YES	None
mba.com	NO	None
mcmobileaccessories.com	NO	None
medelita.com	NO	None
medlink.com	YES	None
medsurvey.com	NO	None
mercedsystems.com	NO	None
mesh.com	NO	Missing tokens
microsoft-hohm.com	NO	None
microsoft.com	YES	CON conflicting token
microsoftfinancing.com	NO	None
microsofthohm.com	YES	None
mitto.com	NO	CON conflicting token
mndigital.com	NO	Invalid tokens
moneybookers.com	YES	None
monster.ch	NO	None
monster.com	NO	None
moversdeal.com	NO	None
msn.at	NO	Missing tokens
msn.be	NO	Missing tokens
msn.com	YES	Invalid tokens
msn.de	NO	Missing tokens
msn.dk	NO	Missing tokens
msn.es	NO	Missing tokens
msn.fi	NO	Missing tokens
msn.fr	NO	Missing tokens
msn.it	NO	Missing tokens
msn.nl	NO	Missing tokens
msn.no	NO	Missing tokens
msn.pt	NO	Missing tokens
msn.se	NO	Missing tokens
mybarstools.com	NO	None
myfreepaysite.com	NO	NID conflicting token
myhomepage.com	YES	CON conflicting token
mynewplace.com	YES	Missing tokens; CON conflicting token
napster.com	YES	None
napster.de	NO	None
nationalgamecity.com	NO	None
nextag.ca	NO	None
nextag.com	NO	None
nflflag.com	NO	None
nupplegal.com	NO	None

Continued on next page . . .

Continued from previous page

Domain	Valid Full P3P Policy	Errors found in CP
nytimes.com	NO	Missing tokens
oakcitygallery.com	NO	None
officedrop.com	NO	None
omniture.com	NO	None
onebagoneearth.com	NO	None
onesky.com	NO	None
onetravel.com	NO	None
onetravelindia.com	NO	None
onewayfurniture.com	NO	None
opinion-central.com	NO	None
orbitz.com	NO	IVD and IVA conflicting tokens
orbitzforbusiness.net	NO	IVD and IVA conflicting tokens
paybycash.com	NO	IVD, IVA, and CON conflicting tokens
paycycle.com	NO	Missing tokens
payscale.com	NO	IVD, IVA, and CON conflicting tokens
pch.com	NO	None
pcicomplianceguide.org	YES	None
pensxpress.com	NO	None
peopleclick.com	NO	None
perfectmatch.com	NO	None
periogen.com	YES	None
permuto.com	NO	None
photosynth.net	NO	Invalid tokens
pictureyoursunique.com	NO	None
pinnaclesys.com	NO	None
platinumgalleria.com	YES	None
popularmedia.com	NO	None
posonlinestore.com	NO	None
pospaper.com	NO	None
precharge.com	YES	None
predictiveresponse.com	NO	None
press8.com	NO	None
priorityclub.com	NO	Missing tokens
priortax.com	YES	Invalid tokens; Missing tokens
prixmoinscher.fr	NO	None
prodebtsupport.com	NO	None
prosperitypublications.net	NO	None
qualityhealth.com	YES	None
quickenbillpay.com	NO	None
quikcondoms.com	NO	None
racingusa.com	NO	None
rapidrefund.net	NO	Invalid tokens; Missing tokens
rapidrepair.com	NO	None
rapidtax.com	YES	Invalid tokens; Missing tokens
rednel.com	NO	None
remington-store.com	NO	None
rent.com	NO	None
repequity.com	NO	None
rewardtv.com	NO	IVD and IVA conflicting tokens
rixty.com	YES	None
roblox.com	NO	None
rockstargames.com	NO	Missing tokens
rockyou.com	NO	None
rozee.pk	NO	None

Continued on next page . . .

Continued from previous page

Domain	Valid Full P3P Policy	Errors found in CP
safecount.net	NO	None
salesforcefoundation.org	NO	Missing tokens
sharefile.com	NO	Missing tokens
shermanstravel.com	NO	None
shop.com	NO	NID conflicting token
shopbrita.com	NO	None
shopcompanion.com	NO	NID conflicting token
shopdeck.com	NO	None
shopiogear.com	NO	None
shopkitchenaid.com	NO	None
shopzilla.com	NO	Invalid tokens
shopzilla.de	NO	Invalid tokens
shopzilla.fr	NO	Invalid tokens
shustir.com	NO	None
simplifi.net	NO	None
simplybabyfurniture.com	NO	None
simplykidsfurniture.com	NO	None
sixcontinentsclub.com	NO	Missing tokens
skincarerx.com	YES	Missing tokens; TEL and CON conflicting tokens
skintreatment.com	NO	None
smartsourceonline.com	NO	None
snaglo.com	NO	None
snapfish.com	NO	None
soccernet.com	NO	None
spardeingeld.de	NO	Invalid tokens
spendgrowgive.com	NO	None
spiceworks.com	NO	None
spoke.com	NO	IVD, IVA, and CON conflicting tokens
spokesoftware.com	NO	IVD, IVA, and CON conflicting tokens
sportingnews.com	NO	None
spytown.com	NO	None
starfieldtech.com	NO	IVD and IVA conflicting tokens
starwars.com	NO	Missing tokens
staybridge.com	NO	Missing tokens
strands.com	NO	None
suresource.com	NO	None
surveillance-video.com	NO	None
sweatmonkey.org	NO	None
talentfilter.biz	NO	IVD, IVA, and CON conflicting tokens
taxact.com	YES	IVA and CON conflicting tokens
taxactonline.com	YES	IVA and CON conflicting tokens
taxcut.com	NO	None
taxpack.com	YES	Invalid tokens; Missing tokens
techbargains.com	NO	IVD, IVA, and CON conflicting tokens
techcctv.com	YES	None
theblueriverbabyshoppe.com	NO	None
theopenskyproject.com	YES	None
thesims2.com	NO	Missing tokens
thumbplay.com	NO	IVD, IVA, and CON conflicting tokens
toluna.com	NO	None
topdjgear.com	NO	None
toponeshop.com	NO	None
torbalscales.com	NO	None
treadmilldoctor.com	NO	None

Continued on next page . . .

Continued from previous page

Domain	Valid Full P3P Policy	Errors found in CP
tripit.com	YES	Missing tokens
tripod.com	NO	Invalid tokens; IVA and CON conflicting tokens
tycoonu.com	NO	Missing tokens
ultimatepay.com	NO	IVD, IVA, and CON conflicting tokens
unbeatablesale.com	YES	None
unique-egifts.com	NO	None
us-appliance.com	NO	None
uscretailproducts.com	NO	None
verisign.com	NO	None
verizon.net	NO	IVD, IVA, and CON conflicting tokens
vermontgear.com	NO	None
viewpoint.com	YES	None
vitadigest.com	NO	None
vitamaker.com	NO	None
voice123.com	YES	None
w3i.com	NO	Invalid tokens; Missing tokens
wallpapers.com	NO	Invalid tokens; Missing tokens
waterpik-store.com	NO	None
weatherbug.com	YES	None
webtv.net	NO	Missing tokens
westfloridacomponents.com	NO	None
whitakertaylor.com	YES	None
whitesmoke.com	YES	Missing tokens
whowhere.com	NO	CON conflicting token
wildwestdomains.com	NO	IVD and IVA conflicting tokens
windowsmedia.net	NO	None
wine.com	YES	TEL and CON conflicting tokens
winferno.com	NO	None
wirefly.com	NO	None
wirelessground.com	NO	None
wisemanfinance.com	NO	None
wondertime.com	NO	None
wooddashexperts.com	NO	None
xbox.com	NO	Invalid tokens
yahoo.com	YES	None
yellowpages.com	NO	None
zocdoc.com	NO	Missing tokens
Total	63/391	134/391

Appendix B

A Large-Scale Evaluation of U.S. Financial Institutions' Standardized Privacy Notices

B.1 Automatic retrieval of privacy notices

B.1.1 Searching for standardized notices

To search for a standardized notice from an institution without exhaustively crawling all parts of each institution's website, we chose to perform an automated Google query. To minimize the chance of accidentally retrieving another institution's standardized notice, particularly in light of the large number of financial institutions with similar names, we restricted each query to a financial institution's website domain using Google's *as_sitesearch* parameter. Among the 6,781 institutions in the FDIC list, 6,409 institutions listed a website URL. For these institutions, we considered the domain of this URL to be that institution's only official domain. The remaining financial institutions, as well as all of the credit unions, did not include a website URL among the metadata we retrieved from regulators. To determine the website domain for that institution, we performed an automated Google query of the string "*Institution name*, *City*, *State*" and took the domain of the first result to be that institution's domain. This heuristic is imperfect, yet we believe it conservatively minimizes false associations (incorrectly attributing a standardized notice to the wrong institution) at the expense of increasing the number of false negatives (not finding notices for institutions that have them available).

Armed with a website domain for an institution, we performed an automated Google Query using the search string "What does *institution name* do with your personal information," inserting the institution's name. This string was the header of the model privacy form [49], leading us to use it as the query. We disabled query autocomplete and the geographic localization of search results using Google's *complete* and *pws* parameters, respectively. For each Google query, we recorded the first page of results, containing between zero and ten links for each institution.

We then automatically downloaded these zero to ten items linked from the first page of Google results for each institution. In our pilot testing, we found standardized notices in both HTML (webpage) and PDF formats. We therefore supported both filetypes. To provide a consistent input for our parser and to record the formatting for future display to consumers, we automatically saved both types of files in the PDF format. We downloaded each webpage using the wkhtmltopdf utility running on Ubuntu Linux.[1] The wkhtmltopdf utility renders a webpage using the webkit engine and then saves this output to PDF. In practice, we found that some links redirected automatically to PDF files, which would cause wkhtmltopdf to return a "failed loading page" error. If our program received this message, or if the URL itself ended in the extension *.pdf*, we instead fetched the PDF using the Linux utility Wget.[2] To prevent the crawler from stalling for long periods of time, we instituted a 60-second timeout that abandoned downloading a page if the download took more than 60 seconds.

B.1.2 Identification of standardized notices

From the 10 or fewer files downloaded for each financial institution, we next chose the single file that had the largest number of features of the model privacy form and considered that to be the institution's standardized notice. If none of the files downloaded matched a substantial fraction of features of the model privacy form, we concluded that we did not have a standardized notice for that institution.

Our first step in making this determination was to extract the text from each PDF file using the Linux utility pdftotext[3] to convert PDF files to plaintext. This utility attempts to maintain the relative layout of text. Because the spacing is not always maintained perfectly, particularly for tables, we designed our parser to be robust to text from different columns of a table flowing together. Furthermore, to eliminate false negatives in parsing caused by unexpected whitespace being inserted in the conversion from PDF to plaintext, we removed all whitespace and non-ASCII characters before parsing the document.

The next step involved selecting at most one file per institution. We selected 25 phrases that should always appear in the model disclosure [49], spread approximately evenly throughout the document. For each file, we searched for all 25 phrases and recorded the number of phrases found as the file's "score." To weed out files that did not appear to be based on the model privacy form, we set a cutoff score of 21, thereby eliminating all files missing 20% or more of these expected keywords and phrases. For each institution, we chose the remaining file with the highest score, if any, to give preference to the most complete disclosure that we found for each institution. In the case of a tie, we chose the file that appeared first in the Google results.

B.2 Verification of parsing

This section provides more detail on our manual verification of our parser's accuracy. We also provide greater detail about our parsing of the disclosure table.

The bank name and the list of six types of personal information an institution collects were both parsed correctly for all 50 institutions we manually verified (100% accuracy). We correctly parsed the document's revision date for 48 of 50 institutions (96%). One of the remaining two institutions used an unexpected hyphen in its revision date (05-2011), which we had not accounted

[1] wkhtmltopdf. `http://wkhtmltopdf.org/`

[2] GNU Wget. `https://www.gnu.org/software/wget/`

[3] Pdftotext. `http://linux.die.net/man/1/pdftotext`

for, while the other institution simply included a bare date in the corner of the form without the required "Rev." or similar text. We correctly identified the "who we are" section for 49 of 50 institutions (98%), missing an institution who reworded this section's header as "who are we?"

We correctly parsed the "to limit sharing" section for 50 of 50 institutions (100%), but we encountered two problems when parsing mail-in forms. Although we correctly parsed 48 of 50 institutions' mail-in forms (96%), or lack thereof, we did not recognize one mail-in form that was embedded as an image file, foiling our conversion from pdf to text. We did not recognize a second mail-in form that lacked a header, instructions, or indication that the form was detachable; instead, the form simply included fields for the consumer to fill in, as well as a series of checkboxes for limiting sharing.

We parsed other sections with slightly lower accuracy. For instance, our parser correctly identified how the institution collects information for 46 of 50 institutions (92%). All errors, however, were caused by the financial institutions deviating in small or large ways from the model privacy form. For instance, one bank rewrote "your investment or retirement portfolio" as "your investments or retirement portfolio," while another bank rewrote "pay your bills" as "pay bills online."

In our manual verification of 50 notices, we parsed 45 of 50 institutions' complete disclosure tables with perfect accuracy across all 6–7 rows (90%). For the five remaining institutions, we correctly parsed all except one or two of the rows of the disclosure table. In four of the five cases, we reported as missing one or two sections that were actually included. In three cases the errors were due to differences in spacing. In two cases, the company unexpectedly omitted a required row of the table, and in another case the company centered a column of the table vertically. In one other case we had a subtle error in our regular expression that lead to a mismatch in text, and in the final case, the institution rewrote "for our everyday business purposes" to read "for your everyday business purposes."

We also correctly parsed the "definitions" section for 45 of the 50 institutions we examined (90%). In three cases, institutions' nonstandard use of the model privacy form caused the incorrect parsing. One institution reworded the specified "doesn't have" as "don't have," another embedded the phrase "we have no affiliates" as an image even though the rest of the section was written as text, and the third institution omitted the definition of "joint marketing" entirely. Vertical centering in tables caused the remaining two errors.

Some individual elements were parsed at a lower rate; manual inspection reveals, however, that these missing elements were often missing from the standardized notice. For instance, we parsed the name of the bank from the header "What does *institution name* do with your personal information?" for 5,973 of notices. Many of the policies for which this section was not recognized seemed to omit this section, often replacing it with the institution's logo. The "Who we are...Who is providing this notice?" section was observed at an even lower rate; our parser found 3,405 of notices to contain this section. The specification for the model privacy form notes that "an institution may omit this FAQ only when one financial institution is providing the notice and that institution is identified in the title" [49]. We did not attempt to verify that this case applied for all institutions that omitted this section.

Similarly, a revision date was recognized for only 4,530 of the policies, even though we accepted a number of different phrasings for this section based on manual inspection of policies that seemed to lack revision dates. The model privacy form [49] included *Rev.* for the revision date. We also accepted the following text: *Revised*, *Privacy Notice:*, and *Revision Date*. All of these matches were case insensitive, and we treated all punctuation as optional. We supported a wide range of formats for dates, including YY/MM/DD and MM/DD/YY formats. We allowed the year to be specified with either two or four digits, we permitted only the month and year to be specified, we allowed either forward slashes or periods as delimiters, and we also recognized dates where the month was written out in words and spaces were used as the delimiter.

B.3 Institutions that Appear to Violate FCRA and GLBA

"**For our affiliates' everyday business purposes** – information about creditworthiness. This reason incorporates sharing information pursuant to section 603(d)(2)(A)(iii) of the FCRA. An institution that shares for this reason must provide an opt-out" [49]. The following institutions stated that they shared for this purpose, yet said that consumers cannot limit this sharing:

Credit Unions:

- Interstate Unlimited Credit Union (http:iufcu.org)
- Keystone Credit Union (keystonecu.com)
- L And N Credit Union (www.lnfcu.com)
- L and N Credit Union (lnfcu.com)
- St Agnes Empls Credit Union (stagnescu.com)
- St Jules Credit Union (stjcu.com)
- The Florist Credit Union (thefloristfcu.org)
- West Branch Valley Credit Union (wbvfcu.org)
- 1st Financial Credit Union (1stfinancialfcu.org)
- Acadiana Medical Credit Union (mylcu.net)
- American Partners Credit Union (apfcu.com)
- Capstone Credit Union (capstonefcu.coop)
- Cherokee County Credit Union (cherokeecountyfcu.com)
- City Employees Credit Union (cecuknox.com)
- Clarkston Brandon Community Credit Union (cbccu.org)
- Clearance Community And Schools Credit Union (ccsfcu.com)
- Community Financial Credit Union (yourlocalcreditunion.com)
- Coors Credit Union (coorscu.org)
- Credit Union South Credit Union (creditunionsouth.com)
- Destinations Credit Union Credit Union (destinationscu.org)
- Family Horizons Credit Union (familyhorizons.com)
- First Credit Union Of Scranton Credit U (firstcu.org)

- Fond Du Lac Credit Union (fdlcu.com)
- Fort Worth Community Credit Union (ftwccu.org)
- Gr Consumers Credit Union (grccu.com)
- Greater Pittsburgh Police Credit Union (pittsburghpolicefcu.com)
- Greenville Credit Union (greenvillefcu.com)
- Hartford Healthcare Credit Union (hhcu.org)
- Highmark Credit Union (highmarkfcu.com)
- Homeport Credit Union (homeportfcu.com)
- Honor Credit Union (honorcu.com)
- Honor Credit Union (honorcu.com)
- Horizons North Credit Union (hncu.org)
- Houston Metropolitan Credit Union (hmefcu.org)
- Jersey Central Credit Union (jerseycentralfcu.com)
- Jersey Shore Credit Union (jerseyshorefcu.org)
- Maryvale Schools Credit Union (maryvaleschoolsfcu.com)
- Nebraska Energy Credit Union (ne-fcu.org)
- Nuvista Credit Union (nuvista.org)
- Oshkosh Community Credit Union (oshkoshcommunitycu.com)
- Pbc Credit Union (pbccu.coop)
- Pelican State Credit Union (pelicanstatecu.com)
- Penobscot County Credit Union (penobscotfcu.com)
- Pinnacle Credit Union (pinnaclecu.org)
- Premier Community Credit Union (premierccu.org)
- Premier Community Credit Union (premierccu.org)
- Proponent Credit Union (proponentfcu.org)
- Sisters Hospital Employees Credit Union (shefcu.org)
- Southern Credit Union (southernfederalcu.org)

Other financial institutions:

- A.J. Smith Federal Savings Bank (ajsmithbank.com)
- A.J.S. Bancorp Inc. (ajsmithbank.com)
- Aquesta Bank (aquesta.com)
- Citizens State Bank of Loyal (csbloyal.com)
- Community Development Bank FSB (comdevbank.com)
- Community State Bank (bankcommunitystate.com)
- First County Bank (firstcountybank.com)
- Hometrust Bancshares Inc (hometrustbanking.com)
- Hometrust Bank (hometrustbanking.com)
- Hyperion Bank (hyperionbank.com)
- Midwest Independent Bancshares Inc (mibanc.com)
- Midwest Independent Bank (mibanc.com)
- SunMark Community Bank (sunmarkbank.com)
- The Bank of Star Valley (bosv.com)
- The Biltmore Bank of Arizona (biltmorebankaz.com)
- The Kansas State Bank (mykansasstatebank.com)
- West One Bank (westonebank.com)

"**For our affiliates to market to you**. This reason incorporates sharing information specified in section 624 of the FCRA. Institutions that include this reason must provide an opt-out of indefinite duration. An institution that is required to provide an affiliate marketing opt-out, but does not include that opt-out in the model form under this part, must comply with section 624 of the FCRA and 12 CFR Part 717, Subpart C, with respect to the initial notice and opt-out and any subsequent renewal notice and opt-out." The following institutions stated that they shared for this purpose, yet said that consumers cannot limit this sharing:

Credit Unions:

- Interstate Unlimited Credit Union (iufcu.org)
- State Employees Credit Union (secufl.org)
- Acadiana Medical Credit Union (mylcu.net)

- Credit Union Of Denver Credit Union (cudenver.com)
- Family Horizons Credit Union (familyhorizons.com)
- Hartford Healthcare Credit Union (hhcu.org)
- Healthcom Credit Union (healthcomfcu.org)
- Mead Employees Credit Union (meadecu.com)
- Mountain America Credit Union (macu.com)
- Nebraska Energy Credit Union (ne-fcu.org)
- North Alabama Educators Credit Union (naecu.org)
- Pbc Credit Union (pbccu.coop)
- Proponent Credit Union (proponentfcu.org)
- Velocity Community Credit Union (velocitycommunity.org)
- Winsouth Credit Union (winsouthcu.com)

Other financial institutions:

- Aquesta Bank (aquesta.com)
- Carolina Premier Bank (carolinapremierbank.com)
- Citizens State Bank of Loyal (csbloyal.com)
- Crest Savings Bank (crestsavings.com)
- Elmira Savings Bank (elmirasavingsbank.com)

"**For nonaffiliates to market to you**. This reason incorporates sharing described in section 6802(b)(1) of GLBA. An institution that shares personal information for this reason must provide an opt-out." The following institutions stated that they shared for this purpose, yet said that consumers cannot limit this sharing:

Credit Unions:

- Brownfield Credit Union (brownfieldfcu.com)
- Chevron Valley Credit Union (chevronvalley.com)
- Financial Center Credit Union (fccuburt.org)
- Franklin First Credit Union (franklinfirst.org)
- Goetz Credit Union (goetzcu.com)

- Harbor Credit Union (harborfcu.org)
- Hartford Healthcare Credit Union (hhcu.org)
- Heritage Valley Credit Union (heritagevalleyfcu.org)
- Lanier Credit Union (lanierfcu.org)
- Lower Columbia Longshoremen Credit Unio (lclfcu.org)
- Lubrizol Employees Credit Union (lzecu.org)
- Marisol Credit Union (marisolcu.org)
- North County Credit Union (northcountycu.org)
- Northwoods Community Credit Union (northwoodscu.com)
- Onomea Credit Union (onomeafcu.org)
- Perry Point Credit Union (perrypointfcu.com)
- Piedmont Credit Union Credit Union (piedmontcu.org)
- Priority One Credit Union (priorityonecu.org)
- Proponent Credit Union (proponentfcu.org)
- Queen Of Peace Arlington Credit Union (qpafcu.com)
- Reno City Employees Credit Union (rcefcu.com)
- San Mateo Credit Union (smcu.org)
- Sd Medical Credit Union (sdmfcu.org)

Other financial institutions:

- Bank of Delight (bankofdelight.com)
- Cornerstone Bancorp (cornerstonenatlbank.com)
- Cornerstone National Bank (cornerstonenatlbank.com)
- Northern Trust Company of New York (northerntrust.com)
- Northern Trust Corporation (northerntrust.com)
- The Northern Trust Company (northerntrust.com)
- The First National Bank of Pontotoc (1stnbpontotoc.com)

B.4 Sharing practices of large banks

Reason for sharing personal information	Does not share		Offers opt-out		No opt-out		(Missing)	
For our everyday business purposes– such as to process your transactions, maintain your account(s), respond to court orders and legal investigations, or report to credit bureaus	0	0.0%	0	0.0%	73	100.0%	0	0.0%
For our marketing purposes– to offer our products and services to you	6	8.2%	7	9.6%	59	80.8%	1	1.4%
For joint marketing with other financial companies	26	35.6%	9	12.3%	38	52.1%	0	0.0%
For our affiliates' everyday business purposes– information about your transactions and experiences	13	17.8%	8	11.0%	51	69.9%	1	1.4%
For our affiliates' everyday business purposes– information about your creditworthiness *[Opt-out mandatory]*	24	32.9%	48	65.8%	0	0.0%	1	1.4%
For our affiliates to market to you *[Opt-out mandatory when sharing; row may be omitted in certain cases]*	5	6.8%	47	64.4%	0	0.0%	21	28.8%
For nonaffiliates to market to you *[Opt-out mandatory when sharing]*	59	80.8%	13	17.8%	1	1.4%	0	0.0%

Table B.1: A summary of data-sharing practices among the 73 of Forbes' 100 largest banks for which we found a standardized notices [51].

Institution name	Our marketing	Joint marketing	Affiliates: Transactions	Affiliates: Credit-worth.	Affiliates' marketing	Nonaffiliates' marketing
1st Source	No opt-out	No opt-out	Don't share	Don't share	Don't share	Don't share
Associated Banc-Corp	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
BancFirst	Don't share	Don't share	Don't share	Don't share	Missing	Don't share
BancorpSouth	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out
Bank of America	No opt-out	No opt-out	No opt-out	Opt-out	Missing	Opt-out
Bank of Hawaii	Opt-out	Opt-out	No opt-out	Opt-out	Opt-out	Don't share
Beneficial Bank	No opt-out	Don't share	No opt-out	Opt-out	Opt-out	Don't share
BOK Financial	No opt-out	No opt-out	No opt-out	Don't share	Opt-out	Opt-out
Brookline Bank	No opt-out	No opt-out	No opt-out	Don't share	Don't share	Don't share
Capital Bank	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out
Capital One	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out

Continued from previous page

Institution name	Our marketing	Joint marketing	Affiliates: Transactions	Affiliates: Credit-worth.	Affiliates' marketing	Nonaffiliates' marketing
Chase	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out
Chemical Bank	No opt-out	Don't share	Don't share	Don't share	Missing	Don't share
Citi	No opt-out	No opt-out	No opt-out	Opt-out	Missing	Opt-out
Cole Taylor Bank	No opt-out	Don't share	Opt-out	Opt-out	Opt-out	Don't share
Columbia State Bank	No opt-out	No opt-out	No opt-out	Don't share	Missing	Don't share
Comerica	No opt-out	Opt-out	No opt-out	Opt-out	Opt-out	Don't share
Commerce Bank	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out	Don't share
Community Bank	Don't share	Don't share	Don't share	Don't share	Missing	Don't share
Doral	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
East West Bank	No opt-out	Don't share	No opt-out	Don't share	Missing	Don't share
Farmers & Merchants	No opt-out	Don't share	Don't share	Don't share	Don't share	Don't share
Fifth Third Bank	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
First Bancorp	No opt-out	Opt-out	No opt-out	Opt-out	Opt-out	Don't share
First Citizens Bancshares	Opt-out	Opt-out	No opt-out	Opt-out	Opt-out	Don't share
First Financial Bank	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out	Don't share
First Horizon	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
First Interstate Bank	No opt-out	No opt-out	No opt-out	Don't share	Missing	Don't share
FirstMerit	Opt-out	Opt-out	No opt-out	Opt-out	Opt-out	Opt-out
First Niagara	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
First Republic Bank	Missing	Don't share	Missing	Missing	Missing	Don't share
Frost	No opt-out	Don't share	No opt-out	Opt-out	Opt-out	Don't share
Glacier Bancorp	No opt-out	No opt-out	No opt-out	Don't share	Missing	Don't share
Hancock Holding	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out	Don't share
Huntington	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
Iberia Bank	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out
Independent Bank	Don't share	Don't share	Don't share	Opt-out	Missing	Don't share
Investors Bank	No opt-out	Opt-out	No opt-out	Opt-out	Opt-out	Don't share
Keycorp	No opt-out	No opt-out	No opt-out	Opt-out	Missing	Don't share
M&T Bank Corporation	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out
MB Financial	Opt-out	Opt-out	No opt-out	Don't share	Opt-out	Don't share
National Penn Bancshares	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
National Bank Holding	No opt-out	Don't share	Don't share	Don't share	Missing	Don't share
NBT Bank	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out	Don't share
Northern Trust	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	No opt-out
N.Y. Community Bancorp	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out
Old National	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out	Don't share
Pinnacle Bank	Opt-out	Don't share	No opt-out	Opt-out	Opt-out	Don't share
PNC Bank	No opt-out	Opt-out	No opt-out	Opt-out	Opt-out	Don't share

Continued from previous page

Institution name	Our marketing	Joint marketing	Affiliates: Transactions	Affiliates: Credit-worth.	Affiliates' marketing	Nonaffiliates' marketing
Popular	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
Private Bancorp	No opt-out	No opt-out	No opt-out	Don't share	Don't share	Don't share
Regions Financial	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
Signature Bank	Don't share	Don't share	Don't share	Don't share	Missing	Don't share
State Street Bank	Don't share	Don't share	Don't share	Don't share	Missing	Don't share
Sterling Bank	No opt-out	Don't share	Don't share	Don't share	Missing	Don't share
Suntrust	No opt-out	Don't share	No opt-out	Opt-out	Opt-out	Don't share
Susquehanna Bank	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out	Opt-out
Synovus Financial	No opt-out	Don't share	No opt-out	Opt-out	Opt-out	Don't share
TCF Financial	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out
Texas Capital Bank	Don't share	Don't share	Don't share	Don't share	Missing	Don't share
West America	No opt-out	Don't share	No opt-out	Don't share	Missing	Don't share
Western Alliance Bancorp	No opt-out	Don't share	Don't share	Don't share	Don't share	Don't share
TFS	No opt-out	Don't share	No opt-out	Opt-out	Opt-out	Don't share
Trustmark	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
U.S. Bank	No opt-out	Don't share	No opt-out	Opt-out	Missing	Don't share
UMB Financial	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
United Community Bank	No opt-out	Don't share	Don't share	Don't share	Missing	Don't share
Valley National Bancorp	No opt-out	Don't share	No opt-out	Don't share	Missing	Don't share
Webster Bank	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
Wells Fargo	No opt-out	Don't share	No opt-out	Opt-out	Opt-out	Don't share
WesBanco	Opt-out	Opt-out	No opt-out	Don't share	Opt-out	Don't share
Valley National Bancorp	No opt-out	Don't share	No opt-out	Don't share	Missing	Don't share
Wintrust Financial	No opt-out	Don't share	Opt-out	Opt-out	Opt-out	Don't share
Zions First National Bank	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share

Table B.2: The detailed sharing practices of each of the 73 financial institutions on Forbes' list of "100 best banks" [51] for which we found a standardized notice.

B.5 Sharing practices of credit card companies

Institution name	Our marketing	Joint marketing	Affiliates: Transactions	Affiliates: Credit-worth.	Affiliates' marketing	Nonaffiliates' marketing
Capital One; Chase; Discover Bank; HSBC	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Opt-out
Bank of America; Citi	No opt-out	No opt-out	No opt-out	Opt-out	Missing	Opt-out
American Express	No opt-out	No opt-out	No opt-out	Opt-out	Opt-out	Don't share
Barclays Bank	No opt-out	Opt-out	No opt-out	Opt-out	Missing	Don't share
GE Capital	No opt-out	Don't share	No opt-out	Don't share	Missing	Don't share
U.S. Bank	No opt-out	Don't share	No opt-out	Opt-out	Missing	Don't share
Wells Fargo	No opt-out	Don't share	No opt-out	Opt-out	Opt-out	Don't share

Table B.3: Sharing practices for reasons other than "our everyday business purposes" of credit card companies that appear on a J.D. Power & Associates list [52]. Capital One, Chase, Discover Bank, and HSBC are listed in a group because they have the same sharing practices. Similarly, Bank of America and Citi have the same sharing practices. We note that institutions differ in their sharing practices. For instance, GE Capital says that it does not share data for three of the purposes listed, whereas other institutions say they share for all purposes listed in the disclosure table.

B.6 Logistic Regression Models

The OCC districts as used in our logistic regression models are: **Northeastern**: Connecticut, Delaware, DC, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Puerto Rico, Rhode Island, U.S. Virgin Islands, Vermont, Virginia, and West Virginia; **Southern**: Alabama, Arkansas, Florida, Georgia, Louisiana, Mississipi, Oklahoma, Tennessee and Texas; **Central**: Illinois, Indiana, Kentucky, Michigan, Minnesota, Ohio, and Wisconsin; and **Western**: Alaska, American Samoa, Arizona, California, Colorado, Guam, Hawaii, Idaho, Iowa, Kansas, Missouri, Montana, Nebraska, Nevada, New Mexico, Oregon, South Dakota, States of Micronesia, Utah, Washington, and Wyoming.

Independent variable	β	**Std. Err.**	**P>|Z|**	β **95% CI**
For our marketing purposes				
Size: Small	0.43	0.10	<0.001	[0.24, 0.62]
Size: Medium	0.74	0.10	<0.001	[0.54, 0.93]
Size: Large	1.46	0.13	<0.001	[1.21, 1.70]
Size: Very large	2.53	0.20	<0.001	[2.14, 2.92]
OCC District (Northeastern)	-0.14	0.12	0.25	[-0.39, 0.10]
OCC District (Central)	-0.23	0.10	0.02	[-0.42, -0.40]
OCC District (Southern)	-0.46	0.10	<0.001	[-0.66, -0.27]
Type: Commercial Bank (OCC)	0.02	0.11	0.88	[-0.20, 0.23]
Type: Savings Association (OTS)	0.34	0.15	0.03	[0.04, 0.63]
Type: Savings Bank (FDIC)	0.26	0.17	0.13	[-0.08, 0.61]
Type: Commercial Bank (FED)	0.11	0.11	0.31	[-0.10, 0.33]
For joint marketing with other financial companies				
Size: Small	0.56	0.14	<0.001	[0.30, 0.83]
Size: Medium	0.80	0.13	<0.001	[0.54, 1.06]
Size: Large	1.52	0.14	<0.001	[1.25, 1.80]
Size: Very large	2.39	0.16	<0.001	[2.08, 2.70]
Trust powers	0.35	0.09	<0.001	[0.17, 0.52]
OCC District (Northeastern)	0.34	0.12	0.01	[0.10, 0.58]
OCC District (Central)	0.22	0.11	0.05	[0.00, 0.45]
OCC District (Southern)	0.08	0.11	0.46	[-0.14, 0.31]
For our affiliates' everyday business purposes– transactions and experiences				
Size: Small	0.41	0.15	0.01	[0.12, 0.69]
Size: Medium	0.77	0.14	<0.001	[0.49, 1.04]
Size: Large	1.50	0.15	<0.001	[1.21, 1.79]
Size: Very large	2.37	0.17	<0.001	[2.04, 2.69]
Trust powers	0.23	0.09	0.01	[0.05, 0.42]
OCC District (Northeastern)	0.003	0.13	0.98	[-0.25,0.25]
OCC District (Central)	0.10	0.12	0.40	[-0.13, 0.33]
OCC District (Southern)	-0.41	0.12	0.001	[-0.65, -0.17]

Continued from previous page

Independent variable	β	Std. Err.	P>|Z|	β 95% CI
For our affiliates' everyday business purposes– creditworthiness				
Size: Small	0.18	0.23	0.45	[-0.28, 0.64]
Size: Medium	0.74	0.21	0.001	[0.32, 1.15]
Size: Large	1.45	0.21	<0.001	[1.03, 1.86]
Size: Very large	2.54	0.21	<0.001	[2.14, 2.95]
Ownership: No stock	-0.85	0.35	0.02	[-1.54, -0.15]
For our affiliates to market to you				
Size: Small	0.51	0.27	0.06	[-0.02, 1.02]
Size: Medium	0.84	0.25	0.001	[0.35, 1.34]
Size: Large	1.59	0.26	<0.001	[1.09, 2.10]
Size: Very large	2.58	0.27	<0.001	[2.06, 3.09]
OCC District (Northeastern)	0.72	0.20	<0.001	[0.33, 1.11]
OCC District (Central)	0.09	0.19	0.63	[-0.29, 0.47]
OCC District (Southern)	0.17	0.19	0.37	[-0.20, 0.54]
Type: Commercial Bank (OCC)	0.06	0.21	0.79	[-0.36,0.47]
Type: Savings Association (OTS)	0.002	0.27	0.99	[-0.52, 0.53]
Type: Savings Bank (FDIC)	-0.03	0.29	0.93	[-0.59, 0.53]
Type: Commercial Bank (FED)	0.38	0.18	0.04	[0.02, -1.86]
For nonaffiliates to market to you				
Size: Small	0.49	0.34	0.15	[-0.18, 1.16]
Size: Medium	0.77	0.33	0.02	[0.13, 1.42]
Size: Large	1.51	0.33	<0.001	[0.87, 2.15]
Size: Very large	1.88	0.33	<0.001	[1.23, 2.53]
OCC District (Northeastern)	0.24	0.30	0.43	[-0.35, 0.82]
OCC District (Central)	0.62	0.26	0.02	[0.11, 1.13]
OCC District (Southern)	0.44	0.27	0.10	[-0.08, 0.95]
Type: Commercial Bank (OCC)	0.73	0.23	0.001	[0.28, 1.17]
Type: Savings Association (OTS)	0.31	0.33	0.348	[-0.34, 0.96]
Type: Savings Bank (FDIC)	0.36	0.36	0.32	[-0.34, 1.05]
Type: Commercial Bank (FED)	0.21	0.27	0.43	[-0.31, 0.72]

Table B.4: Results from the logistic regression models corresponding to the different types of sharing practices. The control categories for each variable are: Size (Very small), OCC District (Western), Type (Commercial Bank - FDIC), Trust Powers (No powers), and Ownership (Shareholders). Only those variables significant at α=0.05 are shown.

B.7 Detailed Sharing Practices

Sharing Practice	Very small		Small		Medium		Large		Very large	
Financial institutions' own marketing purposes (N = 3,552)*										
Don't Share	509	57.6%	423	47.2%	354	39.8%	126	23.6%	33	9.4%
Share, Opt-Out	6	0.7%	15	1.7%	21	2.4%	17	3.2%	27	7.7%
Share, No Opt-Out	368	41.7%	457	51.1%	515	57.9%	390	73.2%	291	82.9%
Joint marketing with other financial companies (N = 3,564)*										
Don't Share	784	88.3%	714	80.3%	678	75.6%	316	59.1%	129	36.3%
Share, Opt-Out	11	1.2%	12	1.4%	19	2.1%	17	3.2%	33	9.3%
Share, No Opt-Out	93	10.5%	163	18.3%	200	22.3%	202	37.8%	193	54.4%
For affiliates' everyday business purposes – transactions and experiences – (N = 3,537)*										
Don't Share	785	89.6%	752	85.1%	711	80.0%	349	65.1%	150	42.6%
Share, Opt-Out	6	0.7%	8	0.9%	14	1.6%	20	3.7%	17	4.8%
Share, No Opt-Out	85	9.7%	124	14.0%	164	18.5%	167	31.2%	185	52.6%
For affiliates' everyday business purposes – creditworthiness – (N = 3,530)*										
Don't Share	835	96.0%	841	95.2%	819	92.0%	455	85.1%	229	65.1%
Share, Opt-Out	31	3.6%	38	4.3%	65	7.3%	79	14.8%	119	33.9%
Share, No Opt-Out	4	0.5%	4	0.5%	6	0.7%	1	0.2%	7	1.1%
For affiliates to market to you (N = 1,284)*										
Don't Share	218	89.7%	232	82.9%	256	76.2%	129	59.5%	72	34.6%
Share, Opt-Out	25	10.3%	47	16.8%	77	22.9%	87	40.1%	136	65.4%
Share, No Opt-Out	0	0.0%	1	0.4%	3	0.9%	1	0.5%	0	0.0%
For non-affiliates to market to you (N = 3,508)*										
Don't Share	857	98.4%	852	97.4%	845	96.5%	499	93.1%	316	90.3%
Share, Opt-Out	12	1.4%	23	2.6%	30	3.4%	37	6.9%	32	9.1%
Share, No Opt-Out	2	0.2%	0	0.0%	1	0.1%	0	0.0%	2	0.6%

Table B.5: Sharing practices by institution's size (assets). Assets' brackets are as follows: Very small: $x <$ 25% percentile; Small: 25% percentile $< x <$ 50% percentile; Medium: 50% percentile $< x <$ 75% percentile; Large: 75% percentile $< x <$ 90% percentile; Very large: 90% percentile $< x$; Smaller institutions share consistently less than larger ones for each of the purposes. * denotes statistical significant at α=0.05 using a χ^2 proportionality test.

Sharing practice	Southern		Central		Western		Northeastern	
Financial institutions' own marketing purposes (N = 3,552)*								
Don't Share	460	47.2%	428	43.9%	364	37.6%	193	30.3%
Share & Opt-Out	23	2.4%	21	2.2%	20	2.1%	22	3.5%
Share & No Opt-Out	491	50.4%	525	53.9%	583	60.3%	422	66.3%
Joint marketing with other financial companies (N = 3,564)*								
Don't Share	747	75.8%	729	75.1%	745	76.7%	400	62.9%
Share & Opt-Out	18	1.8%	24	2.5%	23	2.4%	27	4.3%
Share & No Opt-Out	220	22.3%	218	22.5%	204	21%	209	32.9%
For affiliates' everyday business purposes – transactions and experiences – (N = 3,537)*								
Don't Share	817	83.4%	753	77.4%	737	77.3%	440	69.7%
Share & Opt-Out	11	1.1%	27	2.8%	9	0.9%	18	2.9%
Share & No Opt-Out	151	15.4%	193	19.8%	208	21.8%	173	27.4%
For affiliates' everyday business purposes – creditworthiness – (N = 3,530)*								
Don't Share	901	92.1%	883	90.9%	869	91.10%	526	83.9%
Share & Opt-Out	76	7.8%	83	8.9%	78	8.2%	95	15.2%
Share & No Opt-Out	1	0.1%	5	0.5%	7	0.7%	6	1.0%
For affiliates to market to you (N = 1,284)*								
Don't Share	231	73.1%	267	77.8%	277	75.1%	132	51.6%
Share & Opt-Out	85	26.9%	75	21.9%	92	24.9%	120	46.9%
Share & No Opt-Out	0	0.0%	1	0.3%	0	0.0%	4	1.6%
For non-affiliates to market to you (N = 3,508)								
Don't Share	921	95.8%	934	95.5%	927	97.4%	587	95.1%
Share & Opt-Out	38	3.4%	41	4.2%	25	2.6%	30	4.9%
Share & No Opt-Out	2	0.2%	3	0.3%	0	0.0%	0	0.0%

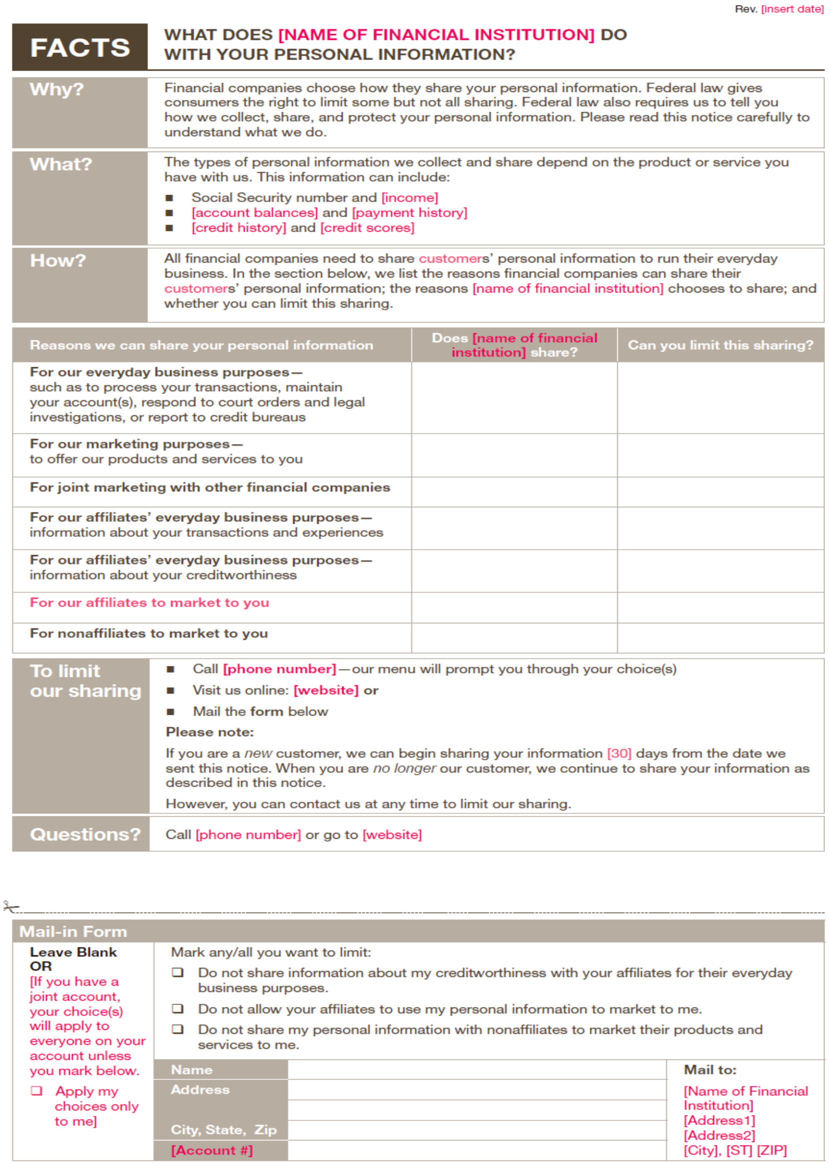
Table B.6: Sharing practices by the OCC District where the institution is physically located. Overall, institutions in the Southern OCC Region shared for the fewest different reasons. Institutions in the Western and Northeastern OCC Regions shared for the largest number of reasons. * denotes statistical significant at α=0.05 using a χ^2 proportionality test.

	Commercial		Commercial		Commercial		Savings		Savings	
Sharing Practice	bank,		bank,		bank,		association,		bank	
Sharing Practice	FDIC		OCC		FED		OTS		FDIC	
Financial institutions' own marketing purposes (N = 3,552)*										
Don't Share	918	43.9%	203	41.9%	180	35.7%	81	32.1%	63	28.8%
Share, Opt-Out	55	2.6%	6	1.2%	15	3.0%	5	2.1%	5	2.3%
Share, No Opt-Out	1,120	53.1%	275	56.8%	309	61.3%	166	65.9%	151	69.0%
Joint marketing with other financial companies (N = 3,564)*										
Don't Share	1,609	76.4%	359	73.9%	340	67.9%	180	71.2%	133	61.3%
Share, Opt-Out	49	2.3%	7	1.4%	18	3.6%	11	4.4%	7	3.2%
Share, No Opt-Out	449	21.1%	120	24.7%	143	28.5%	62	24.5%	77	35.5%
For affiliates' everyday business purposes – transactions and experiences – (N = 3,537)*										
Don't Share	1,664	79.8%	378	77.5%	356	71.5%	185	74.3%	164	76.0%
Share, Opt-Out	34	1.6%	13	2.7%	10	2.0%	4	1.6%	4	1.9%
Share, No Opt-Out	388	18.6%	97	19.9%	132	26.5%	60	24.1%	48	22.2%
For affiliates' everyday business purposes – creditworthiness – (N = 3,530)*										
Don't Share	1,908	91.5%	425	87.5%	429	86.3%	225	91.5%	192	89.3%
Share, Opt-Out	166	8.0%	61	12.6%	65	13.1%	18	7.3%	22	10.2%
Share, No Opt-Out	12	0.6%	0	0.0%	3	0.6%	3	1.2%	1	0.5%
For affiliates to market to you (N = 1,284)*										
Don't Share	551	75.6%	115	68.9%	133	60.5%	62	68.9%	46	58.2%
Share, Opt-Out	174	23.9%	52	31.1%	86	39.1%	28	31.1%	32	40.5%
Share, No Opt-Out	3	0.4%	0	0.0%	1	0.5%	0	0.0%	1	1.3%
For non-affiliates to market to you (N = 3,508)*										
Don't Share	2,016	97.0%	448	93.3%	468	95.7%	234	95.1%	203	94.4%
Share, Opt-Out	60	2.9%	31	6.5%	20	4.1%	11	4.5%	12	5.6%
Share, No Opt-Out	2	0.1%	1	0.2%	1	0.2%	1	0.4%	0	0.0%

Table B.7: Sharing practices by type of institution. Relative to other types of institutions, commercial banks supervised by the FDIC most frequently did not share data. Savings banks supervised by the FDIC share more for joint marketing and own marketing than all other institutions. * denotes statistical significant at α=0.05 using a χ^2 proportionality test.

B.8 Model privacy form

This page and the one that follows contain a screenshot of the most comprehensive version of the model privacy form. Institutions that do not offer opt-outs may use a reduced version that omits the “mail-in form” and “to limit sharing” section [49]. Text in pink is meant to be replaced with information, and the cells of the disclosure table (“reasons we can share your personal information”) must be populated with the institution’s practices.



Rev. [insert date]

FACTS	WHAT DOES [NAME OF FINANCIAL INSTITUTION] DO WITH YOUR PERSONAL INFORMATION?
Why?	Financial companies choose how they share your personal information. Federal law gives consumers the right to limit some but not all sharing. Federal law also requires us to tell you how we collect, share, and protect your personal information. Please read this notice carefully to understand what we do.
What?	The types of personal information we collect and share depend on the product or service you have with us. This information can include: ■ Social Security number and [income] ■ [account balances] and [payment history] ■ [credit history] and [credit scores]
How?	All financial companies need to share customers’ personal information to run their everyday business. In the section below, we list the reasons financial companies can share their customers’ personal information; the reasons [name of financial institution] chooses to share; and whether you can limit this sharing.

Reasons we can share your personal information	Does [name of financial institution] share?	Can you limit this sharing?
For our everyday business purposes— such as to process your transactions, maintain your account(s), respond to court orders and legal investigations, or report to credit bureaus		
For our marketing purposes— to offer our products and services to you		
For joint marketing with other financial companies		
For our affiliates’ everyday business purposes— information about your transactions and experiences		
For our affiliates’ everyday business purposes— information about your creditworthiness		
For our affiliates to market to you		
For nonaffiliates to market to you		

To limit our sharing	■ Call **[phone number]**—our menu will prompt you through your choice(s) ■ Visit us online: **[website]** or ■ **Mail the form** below **Please note:** If you are a *new* customer, we can begin sharing your information [30] days from the date we sent this notice. When you are *no longer* our customer, we continue to share your information as described in this notice. However, you can contact us at any time to limit our sharing.
Questions?	Call [phone number] or go to [website]

✂

Mail-in Form

Leave Blank
OR
[If you have a joint account, your choice(s) will apply to everyone on your account unless you mark below.
❑ Apply my choices only to me]

Mark any/all you want to limit:

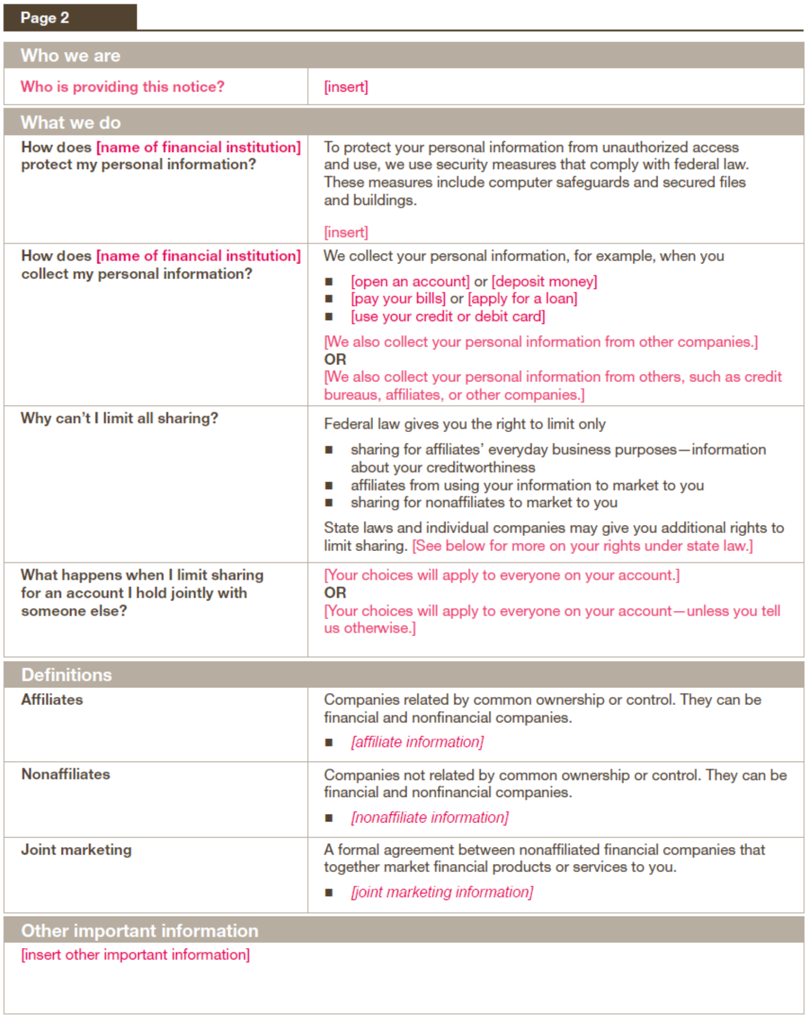
❑ Do not share information about my creditworthiness with your affiliates for their everyday business purposes.

❑ Do not allow your affiliates to use my personal information to market to me.

❑ Do not share my personal information with nonaffiliates to market their products and services to me.

Name	
Address	
City, State, Zip	
[Account #]	

Mail to:
[Name of Financial Institution]
[Address1]
[Address2]
[City], [ST] [ZIP]



Page 2

Who we are

Who is providing this notice?	[insert]

What we do

How does [name of financial institution] protect my personal information?	To protect your personal information from unauthorized access and use, we use security measures that comply with federal law. These measures include computer safeguards and secured files and buildings. [insert]
How does [name of financial institution] collect my personal information?	We collect your personal information, for example, when you ■ [open an account] or [deposit money] ■ [pay your bills] or [apply for a loan] ■ [use your credit or debit card] [We also collect your personal information from other companies.] **OR** [We also collect your personal information from others, such as credit bureaus, affiliates, or other companies.]
Why can't I limit all sharing?	Federal law gives you the right to limit only ■ sharing for affiliates' everyday business purposes—information about your creditworthiness ■ affiliates from using your information to market to you ■ sharing for nonaffiliates to market to you State laws and individual companies may give you additional rights to limit sharing. [See below for more on your rights under state law.]
What happens when I limit sharing for an account I hold jointly with someone else?	[Your choices will apply to everyone on your account.] **OR** [Your choices will apply to everyone on your account—unless you tell us otherwise.]

Definitions

Affiliates	Companies related by common ownership or control. They can be financial and nonfinancial companies. ■ *[affiliate information]*
Nonaffiliates	Companies not related by common ownership or control. They can be financial and nonfinancial companies. ■ *[nonaffiliate information]*
Joint marketing	A formal agreement between nonaffiliated financial companies that together market financial products or services to you. ■ *[joint marketing information]*

Other important information

[insert other important information]

B.9 Developed Codes

Information collected or inferred	Entities with which info may be shared*	Retention
I: Information is collected II: Information is inferred III: Information is collected and inferred IV: The policy doesn't mention this V: Information is explicitly not collected or inferred VI: Information is collected or inferred, but not merged with tracking data VII: Unclear if collected	I: Non-PII (only non-sensitive) II: Non-PII (sensitive and non-sensitive) III: PII IV: Both PII and non-PII V: Information is shared (not clear which) VI: Information is explicitly not shared VII: The policy doesn't mention this VIII: Unclear if shared	0: Company doesn't collect this information I: Limited retention period II: Unlimited retention period III: As required by law IV: The policy doesn't mention this V: Unclear
Purposes*	**Consent Model (Can users limit?)**	**Policy Changes**
0: Company doesn't engage in this practice I: Non-PII (non-sensitive) is used II: Non-PII (sensitive and non-sensitive) III: PII is used IV: Both PII and non-PII V: Information is used, but not clear which VI: The policy doesn't mention this VII: Unclear if it does	0: Company doesn't engage in this practice I: User cannot limit this practice II: Opt-out III: Opt-in IV: The policy doesn't mention this V: This use is not mentioned in policy, hence choices don't apply	I: No notice will be provided II: Notice will be posted in the policy III: Notice will be posted in the policy if major changes IV: Notice will be posted in the policy and email sent if major changes
Mergers and Acquisitions	**Contact means**	**Contact recipient**
I: Notice given (no user choices mentioned) II: Notice is not given (no user choices mentioned) III: Notice is given (user choices mentioned) IV: Notice is not given (user choices mentioned) V: The policy doesn't mention this VI: Unclear	I: Email II: Telephone III: Postal address IV: Web form V: Email and telephone VI: Email and postal address VII: Telephone and postal address VIII: Web form and other IX: More than two of the above X: None	0: No contact information provided I: CPO or similar II: Company customer service or similar III: Legal department IV: Industry organization (e.g., BBB, NAI, DAA, TRUSTe) V: Government entity (FTC) VI: Other VII: Unclear
Access	**Access options**	**Portability and deletion**
I: Authentication-required website II: Anonymous website III: Both anonymous and authenticated website IV: Other V: No access is provided	0: No access is provided I: View II: View and edit	0: No access is provided I: User data can be exported II: User data can be wiped out from company's databases III: User data can be exported and wiped out from company's databases IV: No portability or deletion options mentioned
Security and other practices exist	**Choice method exist**	**Affiliates and Non-affiliates**
I: Yes II: No	I: Yes II: No	I: Mentioned and defined II: Mentioned, but not defined III: Not mentioned

Table B.8: The answer choices for each group of aspects we investigated. *Choices denote the data types (if any) that are used for each of the investigated purposes or shared with each of the investigated entities

B.10 Collection Disclosures

Company	Type of Business	Collect Non-PII (Non-sensitive)	Collect Non-PII (Sensitive)	Collect PII (Non-sensitive)	Collect PII (Sensitive)	Collect Location
Large Members						
AddThis	Analytics, Data Aggregator/Supplier, Social Media	Collect	Collect	Collect, no merge	Don’t mention	Collect
Adobe Advertising	Advertiser, Analytics, Marketing Solutions	Collect	Don’t mention	Collect	Don’t mention	Collect
Adobe Analytics	Analytics, Tag Manager	Collect	Don’t mention	Collect	Don’t mention	Collect
AppNexus	Ad Exchange, Data Management Platform	Collect	Don’t mention	Don’t collect	Don’t mention	Collect
Atlas	Ad Network, Ad Server	Collect	Don’t mention	Collect, no merge	Don’t mention	Collect
Audience Science	Data Management Platform, Demand Side Platform	Collect	Collect	Collect, no merge	Don’t mention	Collect
BlueKai	Data Aggregator/Supplier, Data Management Platform	Collect	Collect	Don’t collect	Don’t collect	Don’t mention
Chango	Data Aggregator/Supplier, Retargeter	Collect	Don’t collect	Collect, no merge	Don’t mention	Don’t mention
Criteo	Ad Network, Retargeter	Collect	Don’t collect	Don’t collect	Don’t collect	Collect
eXelate	Data Aggregator/Supplier, Data Management Platform	Collect	Don’t mention	Don’t collect	Don’t mention	Don’t mention
Facebook Exchange	Ad Exchange, Social Media	Collect	Collect	Collect	Collect	Collect
Google AdSense	Supply Side Platform	Collect	Don’t collect	Collect	Don’t mention	Collect
Lotame	Analytics, Data Management Platform	Collect	Don’t mention	Don’t collect	Don’t mention	Collect
Neustar	Data Aggregator/Supplier	Collect	Collect	Collect	Don’t collect	Collect, no merge
Quantcast	Data Management Platform	Collect	Don’t collect	Collect, no merge	Don’t mention	Collect
Rubicon	Ad Exchange, Supply Side Platform	Collect	Don’t mention	Don’t collect	Don’t mention	Collect
ShareThis	Social Media	Collect	Collect	Collect	Don’t mention	Collect
ValueClick Mediaplex	Ad Network, Ad Server	Collect	Collect	Don’t collect	Don’t collect	Don’t mention
Xaxis	Ad Network	Collect	Collect	Don’t collect	Don’t mention	Don’t mention
Large Non-members						
Disqus	Social Media	Collect	Don’t mention	Collect	Don’t mention	Don’t mention
Gemius	Ad Server, Analytics	Collect	Don’t mention	Don’t collect	Don’t mention	Collect
Histats	Analytics	Collect	Don’t mention	Don’t collect	Don’t mention	Collect
Nielsen	Analytics, Optimizer, Research Provider	Collect	Don’t mention	Don’t collect	Don’t mention	Collect
OpenX	Ad Exchange	Collect	Don’t mention	Don’t mention	Don’t mention	Collect
Optimizely	Website Optimization	Collect	Don’t mention	Don’t collect	Don’t mention	Collect
Right Media	Ad Exchange, Ad Server	Collect	Don’t mention	Don’t collect	Don’t mention	Collect
Statcounter	Analytics	Collect	Don’t mention	Don’t collect	Don’t mention	Don’t mention
Twitter	Publisher, Social Media	Collect	Don’t mention	Collect	Don’t mention	Collect
Tynt	Analytics, Website Optimization	Collect	Don’t mention	Don’t collect	Don’t mention	Don’t mention
VoiceFive	Business Intelligence, Data Aggregator/Supplier	Unclear	Don’t mention	Collect	Don’t mention	Don’t mention
whos.amung.us	Analytics	Collect	Don’t collect	Don’t collect	Don’t mention	Collect
WordPress	Other	Collect	Don’t mention	Collect, no merge	Don’t mention	Don’t mention
Yandex	Ad Network, Publisher, Website Optimization	Collect	Don’t mention	Don’t collect	Don’t mention	Collect
Random Members						
Acxiom	Data Aggregator/Supplier	Collect	Collect	Collect	Collect	Don’t mention
AOL	Ad Network, Ad Server	Collect	Don’t mention	Collect	Collect	Collect
APT from Yahoo!	Ad Exchange	Collect	Collect	Collect	Collect	Collect
Bazaarvoice	Ad Network	Collect	Don’t mention	Collect	Don’t mention	Don’t mention
Media Innovation Group	Marketing Solutions	Collect	Don’t collect	Don’t collect	Don’t mention	Collect
Pulsepoint Audience	Data Management Platform	Collect	Collect	Don’t collect	Don’t mention	Collect
Rocket Fuel	Ad Network	Collect	Collect	Don’t collect	Don’t collect	Don’t mention
Sizmek	Ad Server, Optimizer	Collect	Don’t mention	Collect, no merge	Don’t mention	Collect
Specific Media	Ad Network	Collect	Collect	Collect	Collect	Collect
Vibrant Media	Ad Network, Ad Server	Collect	Don’t mention	Don’t collect	Don’t mention	Don’t mention
Random Non-members						
Ad Magnet	Ad Network, Ad Server	Collect	Collect	Collect, no merge	Don’t mention	Collect
AdGear	Ad Server, Ad Exchange, Analytics	Collect	Don’t mention	Don’t collect	Don’t mention	Collect
Advanse	Analytics	Collect	Don’t mention	Don’t mention	Don’t mention	Don’t mention
Apple	Ad Network, Advertiser, Mobile, Publisher	Collect	Don’t mention	Collect	Collect	Collect
AT&T AdWorks	Ad Network, Data Management Platform	Collect	Don’t mention	Collect	Collect	Collect
CBS Interactive	Ad Network, Publisher	Collect	Don’t mention	Collect	Collect	Collect
ChineseAN	Ad Network	Collect	Don’t mention	Don’t collect	Don’t mention	Don’t mention
Digg	Social Media	Collect	Don’t mention	Collect	Don’t mention	Collect
Dow Jones	Advertiser, Research Provider	Collect	Don’t collect	Collect	Collect	Collect
Essence	Agency	Don’t mention	Don’t mention	Don’t collect	Don’t mention	Don’t mention
ForeSee Results	Analytics, Research Provider	Collect	Collect	Collect	Don’t mention	Don’t collect

Gay Ad Network	Ad Network	Collect	Don’t mention	Collect	Don’t mention	Don’t mention
Httpool	Ad Network	Collect	Don’t mention	Collect	Collect	Collect
MdotM	Ad Network, Demand Side Platform, Mobile	Collect	Don’t mention	Don’t collect	Don’t mention	Collect
News Distribution Network	Ad Network	Collect	Collect	Collect	Don’t mention	Collect
Open Amplify	Data Aggregator/Supplier, Data Management Platform	Collect	Don’t mention	Don’t collect	Don’t mention	Don’t mention
Red Loop Media	Ad Network, Mobile	Collect	Don’t mention	Collect	Don’t mention	Don’t mention
RGM Alliance	Ad Network	Collect	Don’t mention	Don’t collect	Don’t collect	Don’t mention
SET Media	Ad Server, Analytics	Collect	Don’t collect	Don’t collect	Don’t collect	Don’t mention
Smowtion	Ad Network	Collect	Don’t mention	Collect	Don’t mention	Don’t mention
Sojern	Data Aggregator/Supplier	Collect	Don’t mention	Don’t collect	Don’t collect	Don’t mention
Star Media	Ad Network	Collect	Don’t mention	Collect	Don’t mention	Collect
SymphonyAM	Analytics, Research Provider	Collect	Don’t mention	Collect	Don’t mention	Collect
Tapjoy	Creative/Ad Format Technology, Mobile	Collect	Don’t mention	Collect	Don’t mention	Collect
Traffiq	Agency	Collect	Don’t mention	Collect	Collect	Don’t mention
Twelvefold Media	Ad Server, Analytics, Optimizer	Collect	Don’t mention	Don’t collect	Don’t mention	Don’t mention
Unite	Agency	Collect	Don’t mention	Don’t collect	Don’t mention	Don’t mention
Usability Sciences	Analytics, Website Optimization	Collect	Don’t mention	Collect, no merge	Don’t mention	Don’t mention
UserReport	Analytics	Collect	Don’t mention	Don’t collect	Don’t mention	Don’t mention
Verizon	Advertiser, Mobile, Publisher	Collect	Don’t mention	Collect	Collect	Collect
VisibleBrands	Ad Network	Collect	Don’t mention	Don’t mention	Don’t mention	Don’t mention
WildTangent Games	Ad Network	Collect	Don’t mention	Collect	Don’t mention	Don’t mention

Table B.9: Collection practices by companies that have an English-language privacy policy for tracked users. While most of the companies mention collection of device identifiers and general non-PII, they don’t explicitly mention the collection (or lack of) of sensitive non-PII (e.g., race, religion, sexual orientation, health conditions, income bracket, or credit score). A small number of companies that collect PII also indicate that they don’t link PII with tracking data.

B.11 Sharing Disclosures

Company	Affiliates	Non affiliates	Web Publishers	Ad companies	Entity that links with offline	Entity that links with PII	Law Enforcement
Large Members							
AddThis	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Don't mention	Yes
Adobe Advertising	PII	Shared-not clear which	Don't mention	Don't mention	Non-PII and PII	Non-PII and PII	Yes
Adobe Analytics	PII	Don't mention	Don't mention	Don't mention	Non-PII and PII	Non-PII and PII	Yes
AppNexus	Don't mention	Non-PII	Don't mention	Don't mention	Don't mention	Don't mention	Yes
Atlas	Non-PII	Non-PII	Unclear	Non-PII	Don't mention	Don't mention	Yes
Audience Science	Don't mention	Non-PII	Non-PII	Non-PII	Non-PII	Non-PII	Yes
BlueKai	Don't mention	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
Chango	Non-PII	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
Criteo	Non-PII	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Don't mention
eXelate	Non-PII	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
Facebook Exchange	Non-PII and PII	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
Google AdSense	PII	Non-PII	Non-PII	Non-PII	Unclear	Unclear	Yes
Lotame	Non-PII	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
Neustar	Non-PII	Non-PII	Don't mention	Non-PII	Don't mention	Don't share	Yes
Quantcast	Non-PII and PII	Non-PII	Non-PII	Non-PII	Unclear	Unclear	Yes
Rubicon	Unclear	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
ShareThis	Don't mention	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
ValueClick Mediaplex	Non-PII	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
Xaxis	Don't mention	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
Large Non-members							
Disqus	Non-PII and PII	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
Gemius	Non-PII	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Yes
Histats	Don't mention	Non-PII	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention
Nielsen	Non-PII	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
OpenX	Don't mention	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Don't mention
Optimizely	Don't mention	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
Right Media	Non-PII	Non-PII	Non-PII	Non-PII	Unclear	Unclear	Yes
Statcounter	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention
Twitter	Don't mention	Non-PII	Don't mention	Non-PII	Don't mention	Non-PII	Yes
Tynt	Don't mention	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
VoiceFive	PII	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Yes
whos.amung.us	Non-PII	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Don't mention
WordPress	PII	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Yes
Yandex	Non-PII	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
Random Members							
Acxiom	Non-PII and PII	Non-PII and PII	Don't mention	Non-PII and PII	Unclear	Unclear	Yes
AOL	Non-PII and PII	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
APT from Yahoo!	PII	Shared-not clear which	Don't mention	Non-PII	Don't mention	Don't mention	Yes
Bazaarvoice	Non-PII and PII	Non-PII and PII	Don't mention	Non-PII and PII	Don't mention	Don't mention	Yes
Media Innovation Group	Non-PII	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
Pulsepoint Audience	Non-PII	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
Rocket Fuel	Don't mention	Non-PII	Don't mention	Don't mention	Don't mention	Don't share	Yes
Sizmek	Don't mention	Non-PII	Non-PII	Non-PII	Don't mention	Non-PII	Don't mention
Specific Media	Non-PII and PII	Non-PII and PII	Non-PII and PII	Non-PII and PII	Don't mention	Don't mention	Yes
Vibrant Media	Don't mention	Non-PII	Don't mention	Don't mention	Don't mention	Don't mention	Yes
Random Non-members							
Ad Magnet	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Yes
AdGear	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention
Advanse	Non-PII	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
Apple	PII	Don't mention	Don't mention	Don't share	Don't mention	Don't mention	Yes
AT&T AdWorks	PII	Non-PII and PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
CBS Interactive	Non-PII and PII	Non-PII and PII	Don't mention	Non-PII and PII	Don't mention	Don't mention	Yes
ChineseAN	Don't mention	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Don't mention
Digg	PII	Non-PII	Don't mention	Don't mention	Don't mention	Don't mention	Yes
Dow Jones	PII	PII	Don't mention	Don't mention	PII	PII	Yes
Essence	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention
ForeSee Results	PII	Non-PII	Don't mention	Don't mention	Don't share	Don't share	Yes
Gay Ad Network	Don't mention	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
Httpool	Non-PII and PII	Non-PII and PII	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention
MdotM	Non-PII and PII	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
News Distribution Network	Non-PII and PII	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
Open Amplify	Non-PII	Non-PII	Don't mention	Don't mention	Don't mention	Don't mention	Yes
Red Loop Media	Non-PII and PII	Non-PII and PII	Unclear	Don't mention	Don't mention	Don't mention	Yes
RGM Alliance	Non-PII	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
SET Media	Non-PII	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
Smowtion	Don't mention	Non-PII	Non-PII	Non-PII	Non-PII	Non-PII	Yes
Sojern	Don't mention	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
Star Media	Non-PII and PII	Non-PII	Don't mention	Don't mention	Don't mention	Don't mention	Yes
SymphonyAM	PII	Non-PII	Don't mention	Don't mention	Don't mention	Don't mention	Yes
Tapjoy	PII	Non-PII and PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes
Traffiq	PII	Non-PII	Don't mention	Non-PII	Don't mention	Don't mention	Yes
Twelvefold Media	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention
Unite	Don't mention	Non-PII	Non-PII	Non-PII	Don't mention	Don't mention	Yes

Usability Sciences	Don’t mention	Non-PII	Don’t mention	Don’t mention	Don’t mention	Don’t mention	Yes
UserReport	Non-PII	Non-PII	Don’t mention	Don’t mention	Don’t mention	Don’t mention	Yes
Verizon	Non-PII and PII	Non-PII	Don’t mention	Non-PII	Don’t mention	Don’t mention	Yes
VisibleBrands	Don’t mention	Don’t share	Don’t share	Don’t share	Don’t share	Don’t share	Yes
WildTangent Games	PII	Don’t mention	Don’t mention	Don’t mention	Don’t mention	Don’t mention	Don’t mention

Table B.10: Sharing practices by companies that have an English-language privacy policy for tracked users. The cells show the types of information shared with each of the listed entities. Companies share extensively non-PII with non affiliates, but they don’t mention with which particular affiliates the information is shared with. Most companies are particularly silent about sharing information with entities that can link online tracking data with offline data or PII. Only four companies (TargusInfo, VisibleBrands, RocketFuel, and ForeSee Results) explicitly say they don’t share with entities that can link online tracking data with PII.

B.12 Purpose Disclosures

Company	Targeted Ads	Marketing	User Analytics	Ad Analytics	Customize content	Enforcement	Other purposes
Large Members							
AddThis	Non-PII	Don't engage	Non-PII	Don't mention	Non-PII	Yes	Non-PII
Adobe Advertising	Non-PII and PII	PII	Non-PII and PII	Non-PII and PII	Don't mention	Yes	Non-PII and PII
Adobe Analytics	Don't mention	PII	Non-PII and PII	Don't mention	Non-PII and PII	Yes	Non-PII
AppNexus	Non-PII	Don't engage	Don't mention	Non-PII	Don't mention	Don't mention	Non-PII
Atlas	Non-PII	PII	Non-PII	Non-PII	Don't mention	Yes	Don't mention
Audience Science	Non-PII	Don't engage	Don't mention	Non-PII	Don't mention	Yes	Don't mention
BlueKai	Non-PII	Don't engage	Non-PII	Don't mention	Don't mention	Don't mention	Non-PII
Chango	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Don't mention
Criteo	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Don't mention
eXelate	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Don't mention	Non-PII
Facebook Exchange	Non-PII and PII	Don't mention	Don't mention	Non-PII	Non-PII and PII	Yes	Non-PII and PII
Google AdSense	Non-PII and PII	Don't mention	Non-PII and PII	Don't mention	Non-PII and PII	Yes	PII
Lotame	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Don't mention	Non-PII
Neustar	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Non-PII
Quantcast	Non-PII	PII	Non-PII	Don't mention	Don't mention	Yes	Non-PII
Rubicon	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Non-PII
ShareThis	Non-PII	PII	Non-PII	Non-PII	Non-PII and PII	Yes	Non-PII and PII
ValueClick Mediaplex	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Non-PII
Xaxis	Non-PII	Don't engage	Don't mention	Non-PII	Don't mention	Yes	Don't mention
Large Non-members							
Disqus	Non-PII	PII	Non-PII	Don't mention	PII	Yes	Non-PII
Gemius	Don't mention	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Don't mention
Histats	Don't mention	Don't engage	Non-PII	Don't mention	Don't mention	Don't mention	Non-PII
Nielsen	Don't mention	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Non-PII
OpenX	Non-PII	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	Non-PII
Optimizely	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Non-PII
Right Media	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Non-PII
Statcounter	Don't mention	Don't engage	Don't mention	Don't mention	Don't mention	Don't mention	Non-PII
Twitter	Non-PII and PII	Don't mention	Non-PII and PII	Non-PII and PII	Non-PII and PII	Yes	Don't mention
Tynt	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Non-PII
VoiceFive	Unclear if engage	Don't engage	Don't mention	Don't mention	Don't mention	Don't mention	Unclear which info
whos.amung.us	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Don't mention	Don't mention
WordPress	Don't mention	Don't engage	Non-PII	Don't mention	Don't mention	Yes	Don't mention
Yandex	Non-PII	Don't engage	Unclear if engage	Unclear if engage	Non-PII	Yes	Non-PII
Random Members							
Acxiom	Non-PII and PII	Unclear if it does	Non-PII and PII	Non-PII	Non-PII and PII	Don't mention	Non-PII and PII
AOL	Non-PII and PII	PII	Non-PII and PII	Non-PII and PII	Non-PII and PII	Yes	Non-PII and PII
APT from Yahoo!	Non-PII and PII	Don't mention	Don't mention	Don't mention	Don't mention	Yes	Non-PII and PII
Bazaarvoice	Non-PII	PII	Non-PII	Non-PII	Non-PII	Yes	Don't mention
Media Innovation Group	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Don't mention
Pulsepoint Audience	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Don't mention
Rocket Fuel	Non-PII	Don't engage	Don't mention	Non-PII	Don't mention	Yes	Non-PII
Sizmek	Non-PII	Don't engage	Don't mention	Non-PII	Don't mention	Don't mention	Don't mention
Specific Media	Non-PII and PII	Don't mention	Non-PII and PII	Non-PII and PII	Don't mention	Yes	Non-PII and PII
Vibrant Media	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Non-PII
Random Non-members							
Ad Magnet	Non-PII	Don't mention	Non-PII	Non-PII	Don't mention	Yes	Don't mention
AdGear	Non-PII	Don't engage	Don't mention	Non-PII	Don't mention	Don't mention	Don't mention
Advanse	Non-PII	Don't mention	Don't mention	Non-PII	Non-PII	Don't mention	Don't mention
Apple	Non-PII	PII	Non-PII and PII	Non-PII and PII	PII	Yes	Non-PII and PII
AT&T AdWorks	Non-PII and PII	PII	Non-PII and PII	Non-PII and PII	Non-PII and PII	Yes	Non-PII and PII
CBS Interactive	Non-PII and PII	PII	Non-PII	Non-PII and PII	Non-PII and PII	Yes	Non-PII and PII
ChineseAN	Don't mention	Don't engage	Don't mention	Non-PII	Don't mention	Don't mention	Don't mention
Digg	Don't mention	Don't mention	Non-PII	Don't mention	Don't mention	Yes	PII
Dow Jones	Non-PII and PII	PII	Don't mention	Don't mention	Non-PII and PII	Yes	Non-PII and PII
Essence	Non-PII	Don't engage	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention
ForeSee Results	Don't engage	Don't mention	Non-PII and PII	Don't engage	Don't engage	Yes	Don't engage
Gay Ad Network	Non-PII	PII	Don't mention	Non-PII	Non-PII	Don't mention	Don't mention
Httpool	Non-PII and PII	Don't mention	Don't mention	Don't mention	Don't mention	Don't mention	PII
MdotM	Non-PII	Don't mention	Don't mention	Don't mention	Non-PII	Yes	Non-PII
News Distribution Network	Non-PII and PII	PII	Non-PII	Non-PII	Non-PII and PII	Yes	Non-PII and PII
Open Amplify	Don't mention	Don't mention	Unclear if engage	Don't mention	Don't mention	Yes	Non-PII
Red Loop Media	Non-PII	Don't engage	Don't engage	Unclear if engage	Don't engage	Don't mention	Don't mention
RGM Alliance	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Non-PII
SET Media	Non-PII	Don't engage	Non-PII	Non-PII	Don't mention	Yes	Non-PII
Smowtion	Non-PII and PII	Don't mention	Non-PII and PII	Don't mention	Don't mention	Yes	Non-PII and PII
Sojern	Non-PII	Don't engage	Don't mention	Non-PII	Don't mention	Yes	Non-PII
Star Media	Non-PII	Don't mention	Non-PII	Non-PII	Don't mention	Yes	Don't mention
SymphonyAM	Don't mention	Don't engage	Non-PII and PII	Don't mention	Don't mention	Yes	PII
Tapjoy	Unclear which info	PII	Don't mention	Don't mention	Don't mention	Yes	PII
Traffiq	Non-PII and PII	PII	Non-PII and PII	Unclear if engage	Don't mention	Yes	Non-PII and PII
Twelvefold Media	Don't engage	Don't engage	Non-PII	Non-PII	Non-PII	Don't mention	Non-PII
Unite	Non-PII	Don't engage	Don't mention	Non-PII	Don't mention	Yes	Non-PII

Usability Sciences	Don’t mention	Don’t mention	Non-PII	Don’t mention	Don’t mention	Don’t mention	Don’t mention
UserReport	Don’t engage	Don’t engage	Non-PII	Don’t mention	Don’t mention	Yes	Non-PII
Verizon	Non-PII	PII	Non-PII	Non-PII	Don’t mention	Yes	Non-PII and PII
VisibleBrands	Don’t mention	Don’t engage	Non-PII	Don’t mention	Don’t mention	Don’t mention	Don’t mention
WildTangent Games	Non-PII	Don’t mention	Non-PII	Non-PII	Don’t mention	Don’t mention	Unclear which info

Table B.11: Uses by companies that have an English-language privacy policy for tracked users. Cells show the types of information used for the listed purposes. Most of the companies use non-PII to deliver targeted ads. We defined “Marketing” as the practice of using contact information to offer products.“Don’t engage” means the company explicitly say it does not use information for that practice, with the exception of marketing where we entered “Don’t engage” if the company either explicitly says so or it does not collect PII.

B.13 Retention and Access Disclosures

Company	Retention of Non PII	Retention of PII	Type of Access	Data Format (if access provided)	Options (if access provided)	Portability and Data Deletion
Large Members						
AddThis	Limited	Unspecified	No Access	No Access	No Access	No Access
Adobe Advertising	Unspecified	Unspecified	No Access	No Access	No Access	No Access
Adobe Analytics	Unspecified	Unspecified	No Access	No Access	No Access	No Access
AppNexus	Limited	Don't collect	No Access	No Access	No Access	No Access
Atlas	Limited	Unspecified	No Access	No Access	No Access	No Access
Audience Science	Limited	Don't collect	No Access	No Access	No Access	No Access
BlueKai	Limited	Don't collect	Anonymous Access	Profile	View and Edit	Delete
Chango	Limited	Don't collect	No Access	No Access	No Access	None
Criteo	Limited	Don't collect	No Access	No Access	No Access	No Access
eXelate	Limited	Don't collect	Anonymous Access	Profile	View and Edit	None
Facebook Exchange	Limited	Limited	No Access	No Access	No Access	Delete
Google AdSense	Unspecified	Unspecified	Both Anonymous and aunthenticated Access	Profile	View and Edit	Export and Delete
Lotame	Limited	Don't collect	Anonymous Access	Profile	View and Edit	None
Neustar	Unclear	Unspecified	Anonymous Access	Profile	View and Edit	None
Quantcast	Limited	Unspecified	Both Anonymous and aunthenticated Access	Profile	View and Edit	Delete
Rubicon	Limited	Don't collect	No Access	No Access	No Access	No Access
ShareThis	Limited	Limited	No Access	No Access	No Access	No Access
ValueClick Mediaplex	Limited	Don't collect	No Access	No Access	No Access	No Access
Xaxis	Limited	Don't collect	No Access	No Access	No Access	No Access
Large Non-members						
Disqus	Unspecified	Unclear	Authenticated Access	Profile and PII	View and Edit	Delete
Gemius	Unclear	Don't collect	No Access	No Access	No Access	No Access
Histats	Unspecified	Don't collect	No Access	No Access	No Access	No Access
Nielsen	Unclear	Don't collect	No Access	No Access	No Access	No Access
OpenX	Unspecified	Don't collect	No Access	No Access	No Access	No Access
Optimizely	Unlimited	Don't collect	No Access	No Access	No Access	None
Right Media	Unspecified	Don't collect	No Access	No Access	No Access	None
Statcounter	Unspecified	Don't collect	No Access	No Access	No Access	No Access
Twitter	Unspecified	Limited	Authenticated Access	Profile and PII	View and Edit	Delete
Tynt	Limited	Don't collect	No Access	No Access	No Access	No Access
VoiceFive	Unspecified	Unspecified	No Access	Unspecified	No Access	No Access
whos.amung.us	Limited	Don't collect	No Access	No Access	No Access	No Access
WordPress	Unspecified	Unspecified	No Access	No Access	No Access	No Access
Yandex	Unspecified	Don't collect	No Access	No Access	No Access	No Access
Random Members						
Acxiom	Unspecified	Unspecified	Authenticated Access	Profile and PII	View and Edit	None
AOL	Unspecified	Unspecified	Authenticated Access	Unspecified	View and Edit	None
APT from Yahoo!	Unclear	Unclear	Both Anonymous and aunthenticated Access	Profile and PII	View and Edit	Delete
Bazaarvoice	Limited	Unclear	Other Access	Unspecified	View and Edit	None
Media Innovation Group	Limited	Don't collect	No Access	No Access	No Access	No Access
Pulsepoint Audience	Limited	Don't collect	No Access	No Access	No Access	No Access
Rocket Fuel	Limited	Don't collect	No Access	No Access	No Access	No Access
Sizmek	Limited	Unclear	No Access	No Access	No Access	No Access
Specific Media	Limited	Unspecified	No Access	No Access	No Access	No Access
Vibrant Media	Limited	Don't collect	No Access	No Access	No Access	No Access
Random Non-members						
Ad Magnet	Unclear	Unclear	No Access	No Access	No Access	No Access
AdGear	Unspecified	Unspecified	No Access	No Access	No Access	No Access
Advanse	Limited	Don't collect	No Access	No Access	No Access	No Access
Apple	Unspecified	Unclear	Authenticated Access	Unspecified	View and Edit	Delete
AT&T AdWorks	Unspecified	Unclear	Authenticated Access	Unspecified	View and Edit	None
CBS Interactive	Unspecified	Unspecified	Authenticated Access	PII	View and Edit	Delete
ChineseAN	Unspecified	Don't collect	No Access	No Access	No Access	No Access
Digg	Unspecified	Unspecified	Both Anonymous and aunthenticated Access	Profile	View and Edit	Delete
Dow Jones	Unspecified	Unspecified	Both Anonymous and aunthenticated Access	Profile	View and Edit	None
Essence	Unspecified	Don't collect	No Access	No Access	No Access	No Access
ForeSee Results	Unclear	Unclear	Other Access	Unspecified	View	None
Gay Ad Network	Unspecified	Unspecified	No Access	No Access	No Access	No Access
Httpool	Unspecified	Unspecified	No Access	No Access	No Access	No Access
MdotM	Unspecified	Unspecified	No Access	No Access	No Access	No Access
News Distribution Network	Unspecified	Unspecified	Other Access	Unspecified	View and Edit	None
Open Amplify	Unspecified	Unspecified	Authenticated Access	Unspecified	View and Edit	None
Red Loop Media	Unspecified	Unspecified	No Access	No Access	No Access	No Access
RGM Alliance	Unclear	Don't collect	No Access	No Access	No Access	No Access
SET Media	Limited	Don't collect	No Access	No Access	No Access	No Access
Smowtion	Unlimited	Unspecified	Authenticated Access	Unspecified	View and Edit	Delete
Sojern	Limited	Don't collect	No Access	No Access	No Access	No Access
Star Media	Limited	Unclear	No Access	No Access	No Access	None
SymphonyAM	Unspecified	Unspecified	No Access	No Access	No Access	No Access
Tapjoy	Unspecified	Unspecified	Authenticated Access	PII	View and Edit	Delete
Traffiq	Unspecified	Unspecified	Authenticated Access	Unspecified	View and Edit	None
Twelvefold Media	Unspecified	Don't collect	No Access	No Access	No Access	None

Unite	Limited	Don't collect	No Access	No Access	No Access	No Access
Usability Sciences	Unspecified	Unspecified	No Access	No Access	No Access	No Access
UserReport	Unspecified	Don't collect	No Access	No Access	No Access	No Access
Verizon	Unspecified	Unclear	Authenticated Access	PII	View and Edit	None
VisibleBrands	Unspecified	Unspecified	No Access	No Access	No Access	No Access
WildTangent Games	Unspecified	Unspecified	Authenticated Access	PII	View and Edit	None

Table B.12: Retention and access practices by companies that have an English-language privacy policy for tracked users. A large fraction of companies don't disclose the retention period of either non-PII or PII. Disclosed retention periods ranged from 20 days (whos.amung.us) to 2 years (Sojern). Only 28% of the companies offered access to collected data. ForeSee Results requires users to send a written request for access.

B.14 Choice Options

Company	Non-PII for ads	Sensitive non-PII for ads	PII for ads	Collection of non-PII	Merge of non-PII w/PII	Merge w/Offline	Merge Across devices
Large Members							
AddThis	Opt-out	Opt-out	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Adobe Advertising	Opt-out	N/A	Opt-out	Unspecified	Unspecified	Unspecified	Unspecified
Adobe Analytics	N/A	N/A	N/A	Opt-out	Unspecified	N/A	N/A
AppNexus	Opt-out	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Atlas	Opt-out	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Audience Science	Opt-out	Opt-out	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
BlueKai	Opt-out	Opt-out	Don’t engage	Opt-out	Don’t engage	Don’t engage	Unspecified
Chango	Opt-out	Don’t engage	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Criteo	Opt-out	Don’t engage	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
eXelate	Opt-out	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Facebook Exchange	Opt-out	Opt-out	Opt-out	Unspecified	Unspecified	Unspecified	Unspecified
Google AdSense	Opt-out	Don’t engage	Opt-out	Unspecified	Opt-in	N/A	Unspecified
Lotame	Opt-out	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Neustar	Opt-out	Don’t engage	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Quantcast	Opt-out	N/A	Don’t engage	Opt-out	Don’t engage	Don’t engage	Unspecified
Rubicon	Opt-out	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
ShareThis	Opt-out	Opt-in	Don’t engage	Opt-out	Opt-out	N/A	N/A
ValueClick Mediaplex	Opt-out	Opt-out	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Xaxis	Opt-out	Opt-out	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Large Non-members							
Disqus	Unspecified	N/A	Don’t engage	Unspecified	Unspecified	Unspecified	N/A
Gemius	N/A	N/A	Don’t engage	Opt-out	Don’t engage	Don’t engage	N/A
Histats	N/A	N/A	Don’t engage	Opt-out	Don’t engage	Don’t engage	N/A
Nielsen	N/A	N/A	Don’t engage	Opt-out	Don’t engage	Unspecified	N/A
OpenX	Opt-out	N/A	N/A	Unspecified	Unspecified	Unspecified	N/A
Optimizely	Opt-out	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Right Media	Opt-out	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Statcounter	N/A	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Twitter	Opt-out	N/A	Opt-out	Unspecified	Opt-out	Unspecified	Unspecified
Tynt	Opt-out	N/A	Don’t engage	Opt-out	Don’t engage	Don’t engage	N/A
VoiceFive	N/A	N/A	N/A	Opt-out	N/A	N/A	N/A
whos.amung.us	Opt-out	Don’t engage	Don’t engage	Opt-out	Don’t engage	Don’t engage	N/A
WordPress	N/A	N/A	Don’t engage	N/A	N/A	N/A	N/A
Yandex	Unspecified	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Random Members							
Acxiom	Opt-out	Opt-in	Opt-out	Opt-out	Unspecified	Unspecified	Unspecified
AOL	Opt-out	N/A	Opt-out	Unspecified	Unspecified	Unspecified	N/A
APT from Yahoo!	Opt-out	Opt-out	Opt-out	Opt-out	Unspecified	N/A	N/A
Bazaarvoice	Opt-out	N/A	N/A	Opt-out	Don’t engage	Don’t engage	N/A
Media Innovation Group	Opt-out	Don’t engage	Don’t engage	Opt-out	Don’t engage	Don’t engage	N/A
Pulsepoint Audience	Opt-out	Opt-out	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Rocket Fuel	Opt-out	Opt-out	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Sizmek	Opt-out	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Specific Media	Opt-out	Opt-out	Opt-out	Unspecified	Unspecified	Unspecified	Unspecified
Vibrant Media	Opt-out	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Random Non-members							
Ad Magnet	Opt-out	Opt-out	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
AdGear	Opt-out	N/A	Don’t engage	N/A	Don’t engage	Don’t engage	N/A
Advanse	Unspecified	N/A	N/A	Unspecified	Don’t engage	Don’t engage	N/A
Apple	Opt-out	N/A	Unspecified	Unspecified	Unspecified	Unspecified	Unspecified
AT&T AdWorks	Opt-out	N/A	Opt-out	Opt-out	Unspecified	N/A	N/A
CBS Interactive	Opt-out	N/A	Opt-out	Unspecified	N/A	N/A	Unspecified
ChineseAN	N/A	N/A	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Digg	N/A	N/A	N/A	Unspecified	Unspecified	N/A	N/A
Dow Jones	Unspecified	Don’t engage	Unspecified	Unspecified	Unspecified	Unspecified	N/A
Essence	Opt-out	N/A	Don’t engage	N/A	Don’t engage	Don’t engage	N/A
ForeSee Results	Don’t engage	Don’t engage	Don’t engage	Opt-in	Unspecified	Unspecified	N/A
Gay Ad Network	Unspecified	N/A	Unspecified	Unspecified	Unspecified	Unspecified	N/A
Httpool	Unspecified	N/A	Unspecified	Unspecified	Unspecified	N/A	N/A
MdotM	Unspecified	N/A	Don’t engage	Unspecified	Don’t engage	N/A	N/A
News Distribution Network	Opt-out	Opt-out	Unspecified	Unspecified	N/A	N/A	Unspecified
Open Amplify	N/A	N/A	N/A	Unspecified	Don’t engage	Don’t engage	N/A
Red Loop Media	Unspecified	N/A	N/A	N/A	N/A	N/A	N/A
RGM Alliance	Opt-out	N/A	Don’t engage	Unspecified	Don’t engage	N/A	N/A
SET Media	Opt-out	Don’t engage	Don’t engage	Opt-out	Don’t engage	Don’t engage	N/A
Smowtion	Opt-out	N/A	Opt-out	Unspecified	Opt-out	Opt-out	N/A
Sojern	Unspecified	Unspecified	Don’t engage	Unspecified	Don’t engage	N/A	N/A
Star Media	Unspecified	N/A	Don’t engage	Unspecified	N/A	N/A	N/A
SymphonyAM	N/A	N/A	Don’t engage	Unspecified	N/A	N/A	N/A
Tapjoy	Opt-out	N/A	N/A	Unspecified	Unspecified	N/A	N/A
Traffiq	Unspecified	N/A	Unspecified	Unspecified	Unspecified	N/A	N/A
Twelvefold Media	Don’t engage	Don’t engage	Don’t engage	Opt-out	Don’t engage	Don’t engage	N/A
Unite	Opt-out	Opt-out	Don’t engage	Unspecified	Don’t engage	Don’t engage	N/A
Usability Sciences	N/A	N/A	Don’t engage	Opt-out	Don’t engage	Don’t engage	N/A
UserReport	Don’t engage	Don’t engage	Don’t engage	Opt-out	Don’t engage	Don’t engage	N/A
Verizon	Opt-out	N/A	Opt-in	Unspecified	N/A	N/A	Unspecified

VisibleBrands	N/A	N/A	N/A	Unspecified	Don't engage	Don't engage	N/A
WildTangent Games	Unspecified	N/A	N/A	Unspecified	Unspecified	N/A	N/A

Table B.13: User consent practices by companies that have an English-language privacy policy for tracked users. Cells show the choices offered to users for each of the listed data uses. "N/A" means the company does not mention that practice (i.e., we don't know if it does it or not) and therefore no consent options are applicable. While most of the companies offer the opportunity to opt out of targeted ads they don't mention any options to limit online tracking. Nevertheless, there are 18 companies (Gemius, BlueKai, Tynt, Adobe Analytics, VoiceFive, Nielsen, Histats, ShareThis, whos.amung.us, Axciom, Yahoo, Bazaarvoice, Media Innovation Group, AT&T AdWorks, Twelvefold Media, SET Media, Usability Sciences, and UserReport) that state users can opt out of online tracking. The reason why ForSee results says "opt-in" for collection of non-PII is because users voluntarily participate in online surveys implemented by this company. This company also doesn't link data across surveys in a way that survey takers are uniquely identified. While most companies don't engage in merging non-PII with PII or off-line data, those that do engage don't specify consent options for that practice. None of the companies that mention tracking across devices offer any option to limit it.

B.15 Choice methods and affiliations

Company	Link to DAI/NAI home page	Link to DAA/NAI opt-out page	Opt-out button in policy	Opt-out button somewhere else	Other choice method	Membership with DAA/NAI?*
Large Members						
AddThis	Yes	Yes	No	Yes	No	Y/ Y
Adobe Advertising	No	Yes	No	Yes	Yes	Y/ N
Adobe Analytics	No	Yes	No	Yes	No	Y/ N
AppNexus	Yes	Yes	Yes	No	Yes	N/ Y
Atlas	Yes	Yes	No	No	No	Y/ Y
Audience Science	Yes	Yes	Yes	No	No	Y/ Y
BlueKai	Yes	Yes	No	Yes	Yes	Y/ Y
Chango	No	No	No	Yes	No	Y/ Y
Criteo	Yes	Yes	Yes	Yes	No	Y/ Y
eXelate	Yes	Yes	No	Yes	No	Y/ Y
Facebook Exchange	Yes	Yes	No	Yes	Yes	Y/ N
Google AdSense	No	No	No	Yes	No	Y/ Y
Lotame	Yes	Yes	Yes	No	No	Y/ Y
Neustar	Yes	No	No	Yes	No	Y/ Y
Quantcast	Yes	Yes	No	Yes	No	Y/ Y
Rubicon	Yes	Yes	No	No	No	Y/ Y
ShareThis	Yes	Yes	Yes	No	No	Y/ Y
ValueClick Mediaplex	Yes	No	No	Yes	No	Y/ Y
Xaxis	No	No	Yes	No	No	Y/ Y
Large Non-members						
Disqus	No	No	No	No	Yes	N/ N
Gemius	No	No	No	Yes	No	N/ N
Histats	No	No	No	Yes	No	N/ N
Nielsen	No	No	Yes	No	No	N/ N
OpenX	No	Yes	No	Yes	No	N/ N
Optimizely	No	No	Yes	No	Yes	N/ N
Right Media	No	No	Yes	No	No	N/ N
Statcounter	No	No	No	No	No	N/ N
Twitter	No	No	No	Yes	No	N/ N
Tynt	No	No	No	Yes	No	N/ N
VoiceFive	No	No	Yes	No	Yes	N/ N
whos.amung.us	No	Yes	Yes	No	No	N/ N
WordPress	No	No	No	No	No	N/ N
Yandex	No	No	No	No	Yes	N/ N
Random Members						
Acxiom	Yes	No	No	Yes	No	Y/ N
AOL	Yes	Yes	No	No	Yes	Y/ Y
APT from Yahoo!	Yes	Yes	No	Yes	No	Y/ Y
Bazaarvoice	Yes	Yes	Yes	No	Yes	Y/ Y
Media Innovation Group	Yes	Yes	Yes	No	No	Y/ Y
Pulsepoint Audience	No	Yes	Yes	No	Yes	Y/ Y
Rocket Fuel	Yes	Yes	Yes	No	No	Y/ Y
Sizmek	Yes	Yes	Yes	No	No	Y/ Y
Specific Media	Yes	No	Yes	No	Yes	Y/ Y
Vibrant Media	Yes	Yes	No	Yes	No	Y/ Y
Random Non-members						
Ad Magnet	No	No	Yes	No	No	N/ N
AdGear	No	No	Yes	No	No	N/ N
Advanse	Yes	No	No	No	No	N/ N
Apple	No	No	No	Yes	Yes	N/ N
AT&T AdWorks	No	Yes	No	Yes	Yes	N/ N
CBS Interactive	No	Yes	No	No	Yes	N/ N
ChineseAN	No	No	No	No	No	N/ N
Digg	No	No	No	No	Yes	N/ N
Dow Jones	No	No	No	No	No	N/ N
Essence	No	No	No	No	Yes	N/ N
ForeSee Results	No	No	No	No	No	N/ N
Gay Ad Network	No	Yes	No	No	No	N/ N
Httpool	No	No	No	No	No	N/ N
MdotM	No	No	No	No	No	N/ N
News Distribution Network	No	Yes	No	No	No	N/ N
Open Amplify	No	No	No	No	No	N/ N
Red Loop Media	No	No	No	No	No	N/ N
RGM Alliance	No	Yes	No	No	No	N/ N
SET Media	No	Yes	No	Yes	No	N/ N
Smowtion	No	No	No	Yes	No	N/ N
Sojern	Yes	No	No	No	No	N/ N
Star Media	No	No	No	No	No	N/ N
SymphonyAM	No	No	No	No	Yes	N/ N
Tapjoy	No	No	No	No	Yes	N/ N
Traffiq	No	No	No	No	Yes	N/ N
Twelvefold Media	No	Yes	No	No	No	N/ N
Unite	No	No	No	No	No	N/ N
Usability Sciences	No	No	No	No	Yes	N/ N
UserReport	No	No	Yes	No	Yes	N/ N
Verizon	Yes	No	No	No	Yes	N/ N
VisibleBrands	No	No	No	No	No	N/ N

WildTangent Games	No	Yes	No	No	No	N/ N

Table B.14: Choice Methods by companies that have an English-language privacy policy for tracked users. The most popular way to implement an opt-out choice is to provide a link to the DAA or NAI opt-out pages. *Last column indicates whether the DAA or NAI websites list the company as member as of June 2014.

B.16 Contact methods

Company	Contact Method	Contact Name
Large Members		
AddThis	Email and Postal	Privacy team
Adobe Advertising	Web form	Unclear
Adobe Analytics	Web Form and other	Unclear
AppNexus	Web form	Unclear
Atlas	Email	Unclear
Audience Science	Email	Privacy team
BlueKai	Email and Postal	Privacy team
Chango	Web Form and other	Unclear
Criteo	Email and Postal	Unclear
eXelate	More than two	Privacy team
Facebook Exchange	Web Form and other	Unclear
Google AdSense	Web Form and other	Customer Service
Lotame	Email and Postal	Privacy team
Neustar	Email and Postal	Privacy team
Quantcast	Email and Postal	Legal Department
Rubicon	More than two	Other
ShareThis	Postal	Privacy team
ValueClick Mediaplex	Web Form and other	Privacy team
Xaxis	Email and Postal	Legal Department
Large Non-members		
Disqus	Email	Privacy team
Gemius	More than two	Privacy team
Histats	Email	Unclear
Nielsen	Web form	Unclear
OpenX	Web Form and other	Privacy team
Optimizely	Email and Postal	Unclear
Right Media	Postal	Privacy team
Statcounter	More than two	Unclear
Twitter	Email	Privacy team
Tynt	Email and Postal	Privacy team
VoiceFive	Email and Postal	Privacy team
whos.amung.us	Web form	Unclear
WordPress	None	No contact (NA)
Yandex	Web form	Unclear
Random Members		
Acxiom	Email and Phone	Customer Service
AOL	Email	Privacy team
APT from Yahoo!	Web Form and other	Customer Service
Bazaarvoice	Email and Postal	Privacy team
Media Innovation Group	Email and Postal	Legal Department
Pulsepoint Audience	Email	Privacy team
Rocket Fuel	Email and Postal	Privacy team
Sizmek	Web Form and other	Customer Service
Specific Media	Email and Postal	Privacy team
Vibrant Media	Web Form and other	Unclear
Random Non-members		
Ad Magnet	Email	Unclear
AdGear	Email and Phone	Customer Service
Advanse	None	No contact (NA)
Apple	Web Form and other	Customer Service
AT&T AdWorks	Email and Postal	Privacy team
CBS Interactive	Web Form and other	Legal Department
ChineseAN	None	No contact (NA)
Digg	Email	Unclear
Dow Jones	Email	Privacy team
Essence	Email and Phone	Unclear
ForeSee Results	More than two	Unclear
Gay Ad Network	Postal	Privacy team
Httpool	None	No contact (NA)
MdotM	Email	Unclear
News Distribution Network	Email	Customer Service
Open Amplify	More than two	Customer Service
Red Loop Media	Email	Privacy team
RGM Alliance	Email and Postal	Privacy team
SET Media	Email	Unclear
Smowtion	Email and Postal	Unclear
Sojern	More than two	Unclear
Star Media	Email	Unclear
SymphonyAM	More than two	Unclear
Tapjoy	More than two	Privacy team
Traffiq	More than two	Legal Department
Twelvefold Media	Postal	Unclear
Unite	Email and Postal	Privacy team
Usability Sciences	More than two	Unclear
UserReport	Email and Postal	Unclear
Verizon	Email and Postal	Privacy team
VisibleBrands	Web Form and other	Unclear
WildTangent Games	Web Form and other	Privacy team

Table B.15: Contact details by companies that have an English-language privacy policy for tracked users. "Privacy team" is used when a company provides an email with the word "privacy" in it or otherwise gives an indication that a privacy-related person (e.g., CPO or similar) is the recipient of the communication.

B.17 Policy updates, mergers, and definitions

Company	How Company Informs of Policy Changes	Last Policy Update	Merge and Acquisition	Define Affiliates	Define Non-affiliates
Large Members					
AddThis	No notice	4/7/14	No notice (No choices mentioned)	Mentioned and defined	Mentioned and defined
Adobe Advertising	Notice	12/20/13	No notice (No choices mentioned)	Mentioned	Mentioned and defined
Adobe Analytics	Notice	Don't mention	Not mentioned	Mentioned	Mentioned and defined
AppNexus	No notice	2/21/14	No notice (No choices mentioned)	Not mentioned	Mentioned and defined
Atlas	No notice	2/6/14	No notice (No choices mentioned)	Mentioned and defined	Mentioned
Audience Science	Notice	12/4/13	Not mentioned	Not mentioned	Mentioned and defined
BlueKai	Notice	2/27/14	No notice (No choices mentioned)	Not mentioned	Mentioned and defined
Chango	No notice	8/1/11	No notice (No choices mentioned)	Mentioned	Mentioned
Criteo	Notice	11/29/13	Not mentioned	Mentioned	Mentioned
eXelate	No notice	6/15/13	No notice (No choices mentioned)	Not mentioned	Mentioned and defined
Facebook Exchange	Notice	11/15/13	No notice (No choices mentioned)	Mentioned and defined	Mentioned and defined
Google AdSense	Notice + Email	12/20/13	Notice (No choices mentioned)	Mentioned	Mentioned
Lotame	Notice	1/1/12	No notice (No choices mentioned)	Mentioned	Mentioned and defined
Neustar	Notice	10/1/13	No notice (No choices mentioned)	Mentioned	Mentioned
Quantcast	Notice + Email	2/7/14	No notice (No choices mentioned)	Mentioned and defined	Mentioned and defined
Rubicon	Notice	10/28/13	No notice (No choices mentioned)	Not mentioned	Mentioned and defined
ShareThis	Notice + Email	9/20/13	Notice (No choices mentioned)	Not mentioned	Mentioned and defined
ValueClick Mediaplex	Notice	8/12/13	No notice (No choices mentioned)	Mentioned	Mentioned
Xaxis	Notice	1/21/14	No notice (No choices mentioned)	Not mentioned	Mentioned
Large Non-members					
Disqus	No notice	6/5/12	No notice (No choices mentioned)	Mentioned	Mentioned and defined
Gemius	Notice	10/19/11	No notice (No choices mentioned)	Mentioned	Not mentioned
Histats	No notice	Don't mention	Not mentioned	Not mentioned	Mentioned and defined
Nielsen	Notice	3/2/12	No notice (No choices mentioned)	Mentioned and defined	Mentioned and defined
OpenX	Notice	Don't mention	Not mentioned	Not mentioned	Mentioned and defined
Optimizely	Notice + Email	12/16/13	Notice (No choices mentioned)	Not mentioned	Mentioned
Right Media	No notice	11/21/13	No notice (No choices mentioned)	Mentioned and defined	Mentioned and defined
Statcounter	No notice	Don't mention	Not mentioned	Not mentioned	Not mentioned
Twitter	Notice + Email	10/21/13	No notice (No choices mentioned)	Not mentioned	Mentioned and defined
Tynt	No notice	8/8/12	No notice (No choices mentioned)	Not mentioned	Mentioned and defined
VoiceFive	No notice	12/19/13	Not mentioned	Mentioned	Mentioned
whos.amung.us	Notice	12/12/13	No notice (No choices mentioned)	Mentioned and defined	Mentioned and defined
WordPress	No notice	Don't mention	Not mentioned	Mentioned and defined	Not mentioned
Yandex	Notice	9/18/12	Not mentioned	Mentioned	Mentioned and defined
Random Members					
Acxiom	Notice	9/24/13	No notice (No choices mentioned)	Mentioned	Mentioned
AOL	Notice	6/28/13	Notice (opt-out offered)	Mentioned	Mentioned and defined
APT from Yahoo!	Notice + Email	1/7/13	Notice (No choices mentioned)	Mentioned and defined	Mentioned and defined
Bazaarvoice	Notice + Email	1/23/13	No notice (No choices mentioned)	Mentioned and defined	Mentioned
Media Innovation Group	Notice	9/6/11	No notice (No choices mentioned)	Mentioned	Mentioned
Pulsepoint Audience	Notice	4/3/14	No notice (No choices mentioned)	Mentioned and defined	Mentioned
Rocket Fuel	Notice	11/3/12	No notice (No choices mentioned)	Not mentioned	Mentioned
Sizmek	No notice	Don't mention	Not mentioned	Not mentioned	Mentioned and defined
Specific Media	Notice	11/4/13	No notice (No choices mentioned)	Mentioned	Mentioned
Vibrant Media	No notice	4/24/14	No notice (No choices mentioned)	Not mentioned	Mentioned
Random Non-members					
Ad Magnet	Notice	Don't mention	No notice (No choices mentioned)	Not mentioned	Mentioned
AdGear	No notice	Don't mention	Not mentioned	Not mentioned	Not mentioned
Advanse	No notice	Don't mention	Not mentioned	Mentioned	Mentioned and defined
Apple	Notice	3/1/14	No notice (No choices mentioned)	Mentioned	Mentioned
AT&T AdWorks	Notice	9/16/13	No notice (No choices mentioned)	Mentioned and defined	Mentioned and defined
CBS Interactive	Notice + Email	1/2/14	No notice (No choices mentioned)	Mentioned and defined	Mentioned and defined
ChineseAN	No notice	Don't mention	Not mentioned	Not mentioned	Mentioned and defined
Digg	Notice + Email	6/25/13	No notice (No choices mentioned)	Mentioned	Mentioned
Dow Jones	Notice	10/26/13	No notice (No choices mentioned)	Mentioned and defined	Mentioned and defined
Essence	No notice	Don't mention	Not mentioned	Not mentioned	Not mentioned
ForeSee Results	Notice + Email	5/15/13	No notice (No choices mentioned)	Mentioned	Mentioned
Gay Ad Network	No notice	7/24/12	No notice (No choices mentioned)	Not mentioned	Mentioned and defined
Httpool	No notice	Don't mention	Not mentioned	Mentioned	Mentioned
MdotM	Notice	1/16/11	No notice (opt-out offered)	Not mentioned	Mentioned
News Distribution Network	Notice + Email	9/6/13	No notice (No choices mentioned)	Mentioned and defined	Mentioned
Open Amplify	No notice	Don't mention	Not mentioned	Mentioned	Mentioned
Red Loop Media	No notice	Don't mention	No notice (No choices mentioned)	Mentioned	Mentioned
RGM Alliance	Notice	6/28/13	No notice (No choices mentioned)	Mentioned	Mentioned and defined
SET Media	Notice	2/12/13	No notice (No choices mentioned)	Mentioned	Mentioned and defined
Smowtion	Notice + Email	10/17/13	No notice (No choices mentioned)	Not mentioned	Mentioned and defined
Sojern	Notice	Don't mention	No notice (No choices mentioned)	Not mentioned	Mentioned
Star Media	No notice	Don't mention	Not mentioned	Mentioned	Mentioned and defined
SymphonyAM	Notice + Email	2/5/14	No notice (No choices mentioned)	Mentioned	Mentioned
Tapjoy	Notice + Email	2/18/14	No notice (No choices mentioned)	Mentioned and defined	Mentioned and defined
Traffiq	Notice + Email	Don't mention	No notice (No choices mentioned)	Mentioned	Mentioned
Twelvefold Media	No notice	11/3/11	Not mentioned	Not mentioned	Not mentioned
Unite	Notice	7/1/12	No notice (No choices mentioned)	Not mentioned	Mentioned and defined
Usability Sciences	Notice + Email	Don't mention	Not mentioned	Not mentioned	Mentioned

UserReport	No notice	Don't mention	Notice (opt-out offered)	Mentioned	Mentioned
Verizon	Notice + Email	1/1/14	No notice (No choices mentioned)	Mentioned and defined	Mentioned
VisibleBrands	No notice	Don't mention	Not mentioned	Not mentioned	Not mentioned
WildTangent Games	No notice	Don't mention	Not mentioned	Not mentioned	Not mentioned

Table B.16: For policy updates, "Notice" means the company indicates that it will post a notice in the privacy policy indicating that it has changed. While several companies mention and define affiliates and non-affiliates, those definitions are vague and not consistent across companies.

B.18 Other disclosures

Company	Mention EU provisions	Mention children's provisions	Claim self-regulation affiliation*	Mask IP Address	Stores data anonymized	Stores data encrypted	Mention how tracking works	Mention third-party information sources	Link to educational material	Suggests browser's privacy settings
Large Members										
AddThis	Yes	Yes	Yes	No	Yes	Yes	No	No	No	No
Adobe Advertising	No	No	Yes	No	No	No	Yes	Yes	Yes	Yes
Adobe Analytics	Yes	No	Yes	No	No	No	Yes	Yes	Yes	Yes
AppNexus	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Atlas	No	Yes	Yes	No	Yes	No	No	Yes	No	Yes
Audience Science	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
BlueKai	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Chango	No	No	Yes	No	Yes	No	Yes	Yes	No	Yes
Criteo	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes
eXelate	No	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes
Facebook Exchange	Yes	No	Yes	No	No	No	Yes	Yes	No	Yes
Google AdSense	Yes	No	Yes	No	No	No	Yes	No	No	Yes
Lotame	Yes	No	Yes	No	Yes	No	Yes	Yes	No	Yes
Neustar	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes
Quantcast	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No
Rubicon	No	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes
ShareThis	No	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes
ValueClick Mediaplex	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	No
Xaxis	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Large Non-members										
Disqus	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes
Gemius	Yes	No	Yes	Yes	Yes	No	Yes	No	No	Yes
Histats	No	Yes	No	Yes	Yes	No	Yes	No	Yes	No
Nielsen	No	No	No	No	Yes	No	Yes	Yes	No	Yes
OpenX	No	No	No	No	No	No	No	Yes	No	No
Optimizely	Yes	Yes	No	No	No	No	Yes	No	No	Yes
Right Media	No	No	No	No	Yes	No	Yes	Yes	No	Yes
Statcounter	No	No	No	No	No	No	Yes	Yes	No	No
Twitter	Yes	Yes	No	No	Yes	No	Yes	Yes	No	Yes
Tynt	No	Yes	No	No	Yes	Yes	No	No	No	No
VoiceFive	Yes	Yes	No	No	No	No	Yes	Yes	No	Yes
whos.amung.us	No	Yes	No	No	Yes	No	Yes	No	No	Yes
WordPress	No	No	No	No	No	No	Yes	No	No	Yes
Yandex	No	No	No	No	No	No	Yes	No	No	Yes
Random Members										
Acxiom	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
AOL	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes
APT from Yahoo!	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No
Bazaarvoice	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Media Innovation Group	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes
Pulsepoint Audience	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	No
Rocket Fuel	No	No	Yes	No	No	No	No	Yes	No	No
Sizmek	Yes	No	Yes	No	Yes	No	Yes	No	No	No
Specific Media	No	Yes	Yes	No	Yes	No	Yes	Yes	No	No
Vibrant Media	No	No	Yes	No	Yes	No	Yes	Yes	No	No
Random Non-members										
Ad Magnet	No	No	No	No	Yes	No	Yes	Yes	No	Yes
AdGear	No	No	No	No	No	No	Yes	No	No	No
Advanse	No	No	No	No	No	No	No	No	No	Yes
Apple	Yes	Yes	Yes	No	No	Yes	Yes	No	No	Yes
AT&T AdWorks	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
CBS Interactive	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes
ChineseAN	No	No	No	No	Yes	No	No	No	No	No
Digg	No	No	No	No	Yes	No	Yes	Yes	No	No
Dow Jones	No	Yes	No	No	No	No	Yes	Yes	No	Yes
Essence	No	No	No	No	No	No	No	No	No	No
ForeSee Results	Yes	Yes	No	No	No	No	No	Yes	No	Yes
Gay Ad Network	Yes	No	No	No	No	No	No	No	No	Yes
Httpool	No	No	No	No	Yes	No	Yes	No	No	No
MdotM	No	Yes	No	No	No	No	Yes	Yes	No	Yes
News Distribution Network	Yes	Yes	No	No	No	No	No	Yes	No	Yes
Open Amplify	No	No	No	No	No	No	No	No	No	No
Red Loop Media	No	No	No	No	No	No	No	No	No	No
RGM Alliance	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	No
SET Media	No	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes
Smowtion	No	Yes	No	No	No	No	Yes	Yes	No	Yes
Sojern	No	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes
Star Media	Yes	No	No	No	No	No	Yes	No	No	Yes
SymphonyAM	No	Yes	No	No	Yes	No	Yes	Yes	No	No
Tapjoy	No	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes
Traffiq	No	Yes	No	No	Yes	No	Yes	No	No	Yes
Twelvefold Media	No	No	No	No	No	No	No	No	No	Yes

Unite	No	Yes	No	No	Yes	No	No	Yes	Yes	No
Usability Sciences	Yes	Yes	No	No	Yes	No	Yes	No	No	No
UserReport	No	No	Yes	No	No	No	Yes	No	No	No
Verizon	No	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes
VisibleBrands	No	No	No	No	No	No	Yes	No	No	Yes
WildTangent Games	Yes	Yes	No	No	No	No	Yes	Yes	No	Yes

Table B.17: Other stated practices by companies that have an English-language privacy policy for tracked users. We coded the practice as "Yes," when the practice was explicitly mentioned, a "No" code means the practice was not mentioned. Three companies notably mention that they mask IP addresses. A large fraction of companies (38.7%) don't mention whether or not they receive information from third-parties, and those who do mention it, don't explicitly indicate who those third parties are. *Affiliation with any self-regulatory organization, not only DAA or NAI

Appendix C

Perceptions of Online Behavioral Advertising

C.1 Interview script

1. What is the first thing that comes to your mind when you hear “Internet advertising”?
2. How do you feel about Internet Advertising?
    - Do you like Internet advertising?
    - Is Internet advertising useful?
    - Is Internet advertising distracting?
3. In general, do you find the advertisements you see on the Internet to be relevant to your interests?
4. Do you think that the ads you see when browsing the Internet are tailored to your personal interests? If yes:
    - Is it useful for you to see ads that are tailored to your interests?
    - How do you think online companies decide which ads are more suitable for you?

    If not:

    - Would it be useful for you to see ads more tailored to your interests?
    - How do you think online companies could know which ads are more relevant for you?
5. Have you heard of the term “targeted advertising”? If yes:
    - What does targeted advertising mean to you?
    - How do you think it works?
6. Have you heard of the term “behavioral advertising”? If yes:
    - What does behavioral advertising mean to you?
    - How do you think it works?
7. When surfing the Internet, have you ever seen either of these icons? **[A page with the two OBA icons and taglines is shown here]** If yes:
    - Where did you see it?
    - Have you clicked on it?

    If yes:

- What happened when you clicked on it?

If not:

- What do you think would happen if you clicked on it?
- What do you think is the purpose of this icon?

If not:

- What do you think is the purpose of this icon?

8. These icons usually appear on Internet ads. Here are two examples: **[A page with the ads containing these icons and taglines is shown here]**

- Do you remember having seen any ads with this icon?

If yes:

- Have you ever clicked on the icon?

If yes:

- What happened when you clicked on it?

If not:

- What do you think would happen if you clicked on it?
- What do you think is the purpose of this icon?

If not:

- What do you think is the purpose of this icon?

[VIDEO was shown here]

9. What does behavioral advertising mean to you?

10. In your understanding, what is a third-party cookie?

11. What information do you think online advertising companies can collect about you?

12. Do you think online advertising companies can have access to:

- Your name?
- Your Address?
- Your Telephone?
- Your Email address?
- The city where you live?

13. How do you think behavioral advertising can benefit Internet users?

14. How do you think behavioral advertising can benefit online advertising companies?

15. Is there any other party that could benefit from behavioral advertising?

- Who?
- How?

16. Are there any negative aspects of behavioral advertising?

17. Overall, how do you feel about online behavioral advertising? Why?

Scenario	YES/NO	Reason
You are planning your next vacation using the internet.		
You are shopping for a car and a car loan.		
A friend of you has a STD and asks you to help him to find some treatment alternatives.		
You are job-hunting online.		
You are ordering all of your food and household goods for the week online.		
You are reading the news.		

18. Are there situations in which you would be more willing to let companies collect information about your web browsing in order to send you ads tailored to your interests?

19. For each of the following scenarios, please indicate if you would like online advertising companies to collect information about your web browsing in order to deliver tailored ads. Please explain the reasoning behind your decisions

    If the participant expressed any concern about OBA:

20. Would your attitude towards behavioral advertising change if

    - The advertising companies notified you what information is being collected and how that information is used?
    - The company that is collecting information allows you to decide when to allow or block the data collection?

21. For each of the following companies, please tell me a) if youÕre familiar with the company and b) if you would permit that company to collect information about your web browsing to show tailored ads.

Company name	I am familiar with this company (YES/NO)	I would like to let it collect data. Why?
Google		
Yahoo!		
24/7 Real Media		
AOL Advertising		
BlueKai		
Casale Media		
Microsoft Advertising		

22. Are there any circumstances in which you would NOT like online companies to collect data about your browsing in order to show tailored ads?

23. Are you aware of any ways that can help you stop receiving targeted ads? Y/N, Which? **If affirmative answer to previous question but the participant did not mention “software tools”:**

24. Are you aware of any software designed to help users manage the targeted ads that they receive?

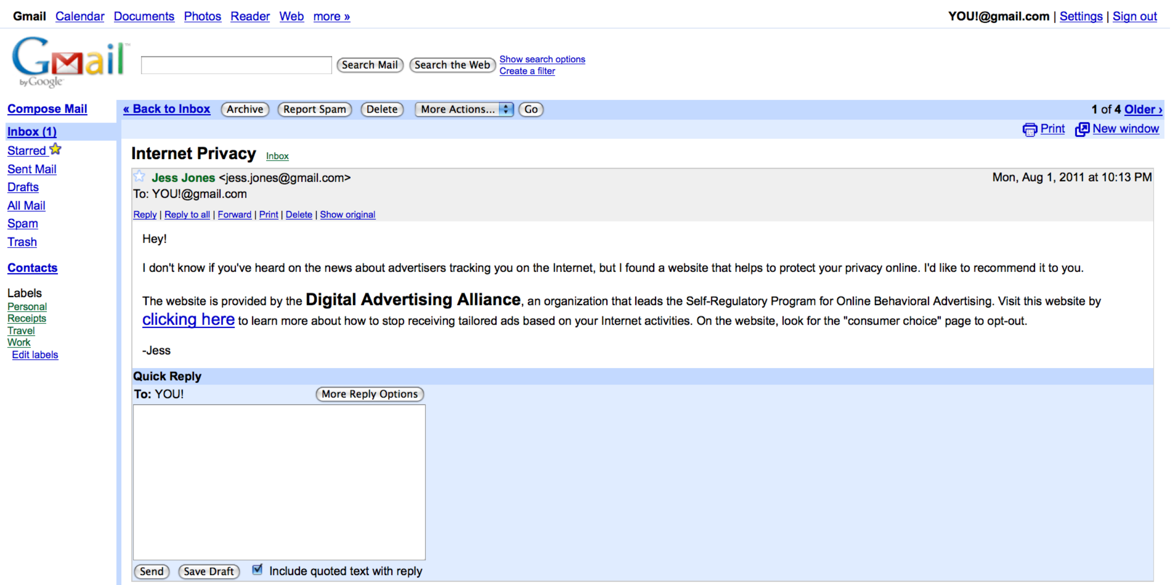
25. Are you aware of any laws dealing with online behavioral advertising?

26. Do you have any additional comments?

Appendix D

A Usability Evaluation of Tools to Limit Online Behavioral Advertising

D.1 Introducing participants to tested tools



Internet Privacy

Jess Jones <jess.jones@gmail.com>
To: YOU!@gmail.com
Mon, Aug 1, 2011 at 10:13 PM

Hey!

I don't know if you've heard on the news about advertisers tracking you on the Internet, but I found a website that helps to protect your privacy online. I'd like to recommend it to you.

The website is provided by the **Digital Advertising Alliance**, an organization that leads the Self-Regulatory Program for Online Behavioral Advertising. Visit this website by clicking here to learn more about how to stop receiving tailored ads based on your Internet activities. On the website, look for the "consumer choice" page to opt-out.

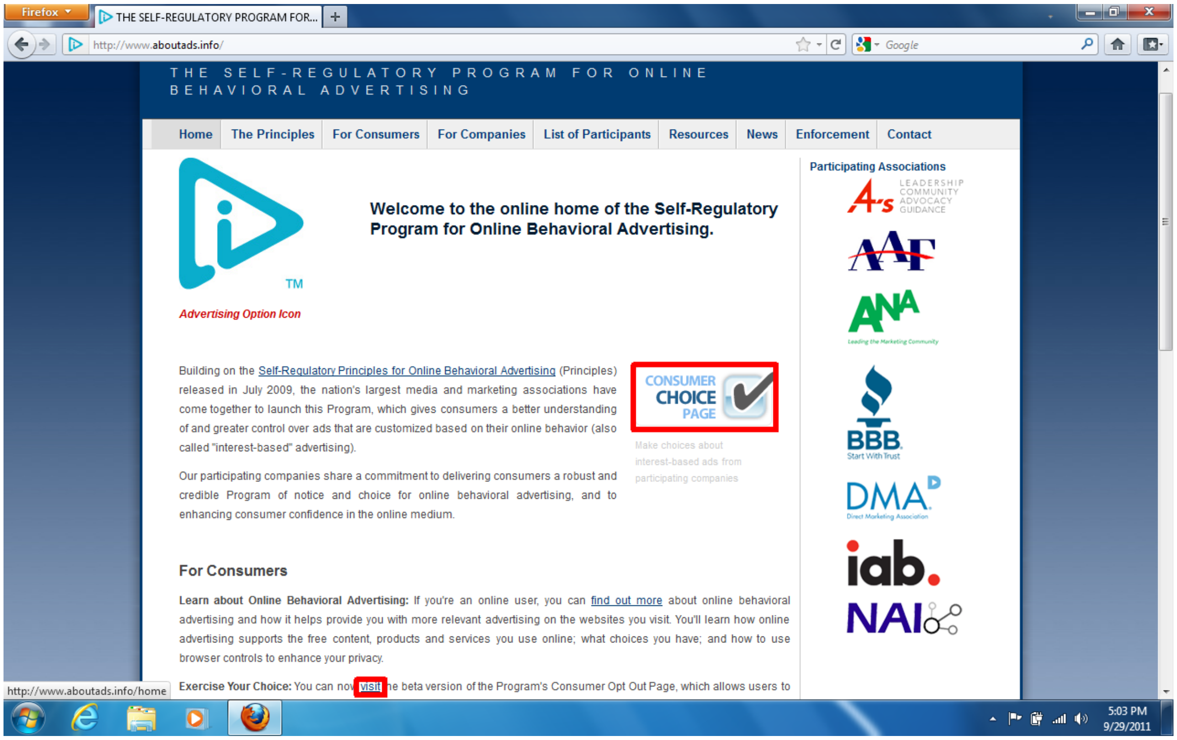
-Jess

Figure D.1: This screenshot shows the email that was used to introduce participants to the DAA website. Similar emails were used to introduce other participants to their assigned tools. When participants clicked on *clicking here* links, they were taken to a support webpage from the tool provider.

Tool	Tool's support webpage
Blocking	
TACO	http://abine.com/preview/taco.php
Ghostery	http://www.ghostery.com/
IE-TPL	http://windows.microsoft.com/en-US/windows7/How-to-use-Tracking-Protection-and-ActiveX-Filtering
AdBlock Plus	http://adblockplus.org/en/
Opt-out	
DAA	http://www.aboutads.info/
Evidon	http://www.evidon.com/
PrivacyMark	http://www.privacychoice.org/privacymark
Built-in	
IE-Settings	http://windows.microsoft.com/en-US/windows7/Change-Internet-Explorer-9-Privacy-settings
Firefox	http://support.mozilla.com/en-US/kb/Options%20window%20-%20Privacy%20panel

Table D.1: This table shows the URL of the support webpage for each of the tested tools. Participants were directed to these URLs to learn about their assigned tool.

D.2 Screenshots of opt-out tools


Firefox
THE SELF-REGULATORY PROGRAM FOR...
http://www.aboutads.info/
Google
Home
The Principles
For Consumers
For Companies
List of Participants
Resources
News
Enforcement
Contact
Participating Associations
4A's LEADERSHIP COMMUNITY ADVOCACY GUIDANCE
AAF
ANA Leading the Marketing Community
BBB Start With Trust
DMA Direct Marketing Association
iab.
NAI
TM
http://www.aboutads.info/home
5:03 PM
9/29/2011

THE SELF-REGULATORY PROGRAM FOR ONLINE BEHAVIORAL ADVERTISING

Welcome to the online home of the Self-Regulatory Program for Online Behavioral Advertising.

Advertising Option Icon

Building on the Self-Regulatory Principles for Online Behavioral Advertising (Principles) released in July 2009, the nation's largest media and marketing associations have come together to launch this Program, which gives consumers a better understanding of and greater control over ads that are customized based on their online behavior (also called "interest-based" advertising).

Our participating companies share a commitment to delivering consumers a robust and credible Program of notice and choice for online behavioral advertising, and to enhancing consumer confidence in the online medium.

CONSUMER CHOICE PAGE

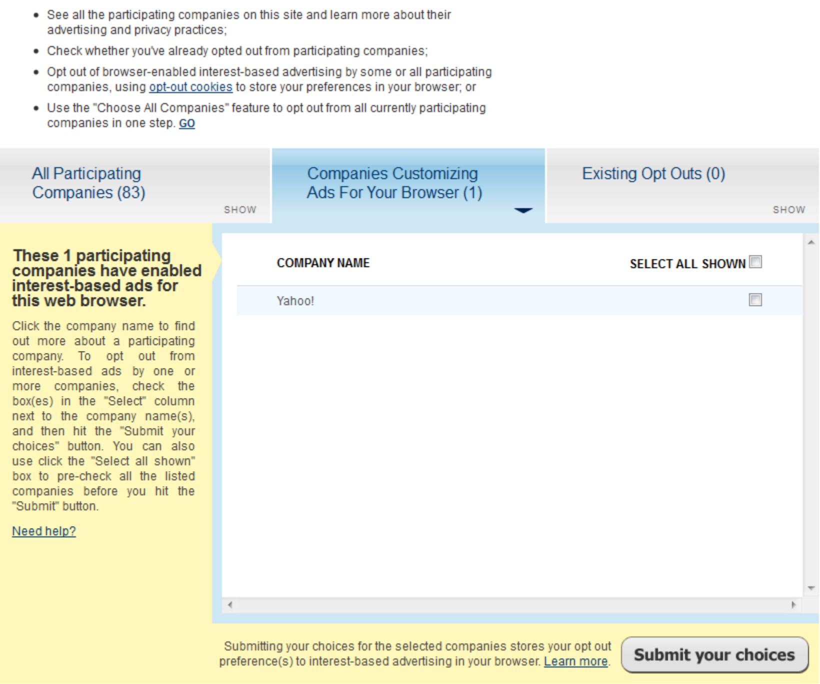
Make choices about interest-based ads from participating companies

For Consumers

Learn about Online Behavioral Advertising: If you're an online user, you can find out more about online behavioral advertising and how it helps provide you with more relevant advertising on the websites you visit. You'll learn how online advertising supports the free content, products and services you use online; what choices you have; and how to use browser controls to enhance your privacy.

Exercise Your Choice: You can now visit the beta version of the Program's Consumer Opt Out Page, which allows users to

Figure D.2: The DAA home page, with red rectangles indicating links to the opt-out page. Most users didn't realize the checkmark icon or "visit" links would lead them to the opt-out page. Instead, two of the five participants testing the DAA's opt-out instead clicked a "click here" link lower on the page, even though the full text of the sentence containing the link was "If you would like to register to use the icon, please *click here*." Those two users were very confused when they landed on a page where advertising companies can register to join the DAA, with one user wondering why opting out costs $5,000.



- See all the participating companies on this site and learn more about their advertising and privacy practices;
- Check whether you've already opted out from participating companies;
- Opt out of browser-enabled interest-based advertising by some or all participating companies, using opt-out cookies to store your preferences in your browser; or
- Use the "Choose All Companies" feature to opt out from all currently participating companies in one step. GO

All Participating Companies (83)
SHOW
Companies Customizing Ads For Your Browser (1)
Existing Opt Outs (0)
SHOW

These 1 participating companies have enabled interest-based ads for this web browser.

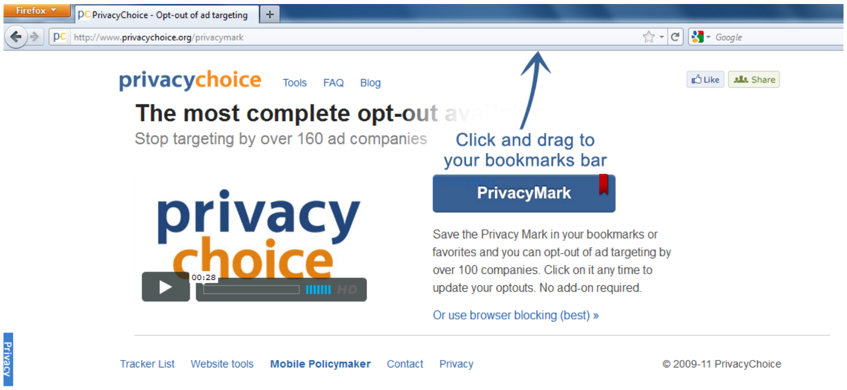
Click the company name to find out more about a participating company. To opt out from interest-based ads by one or more companies, check the box(es) in the "Select" column next to the company name(s), and then hit the "Submit your choices" button. You can also use click the "Select all shown" box to pre-check all the listed companies before you hit the "Submit" button.

Need help?

COMPANY NAME	SELECT ALL SHOWN □
Yahoo!	□

Submitting your choices for the selected companies stores your opt out preference(s) to interest-based advertising in your browser. Learn more.

Figure D.3: The DAA opt-out page, whose layout confused users. The page has three tabs: "All Participating Companies," "Companies Customizing Ads For Your Browser," and "Existing Opt Outs." The default tab is "Companies Customizing Ads For Your Browser," which appears even when a user clears her cookies. To actually opt out of all available companies, a user must first click the "All Participating Companies" tab before choosing "Select All Shown."



The most complete opt-out av

Stop targeting by over 160 ad companies

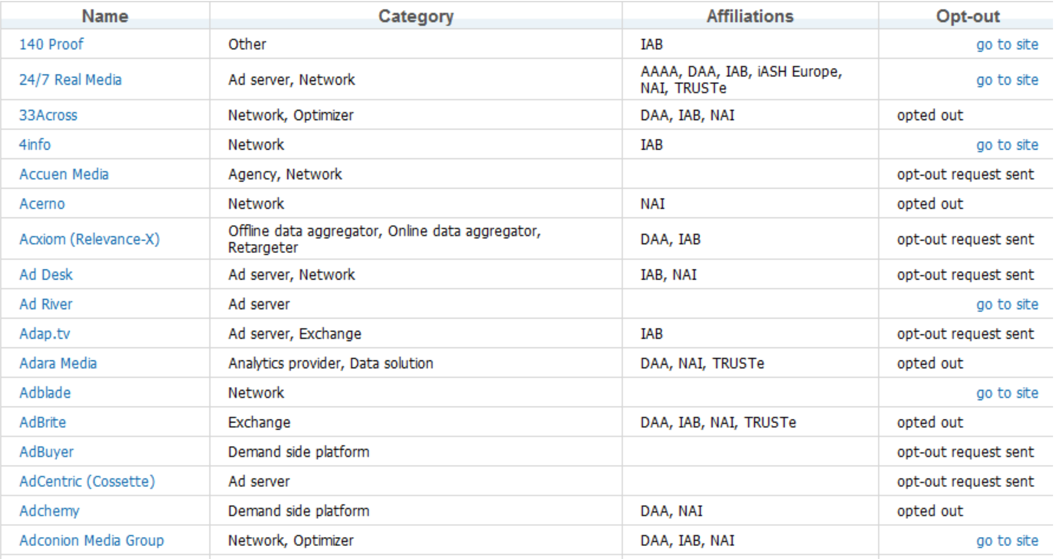
Click and drag to your bookmarks bar

Save the Privacy Mark in your bookmarks or favorites and you can opt-out of ad targeting by over 100 companies. Click on it any time to update your optouts. No add-on required.

Or use browser blocking (best) »

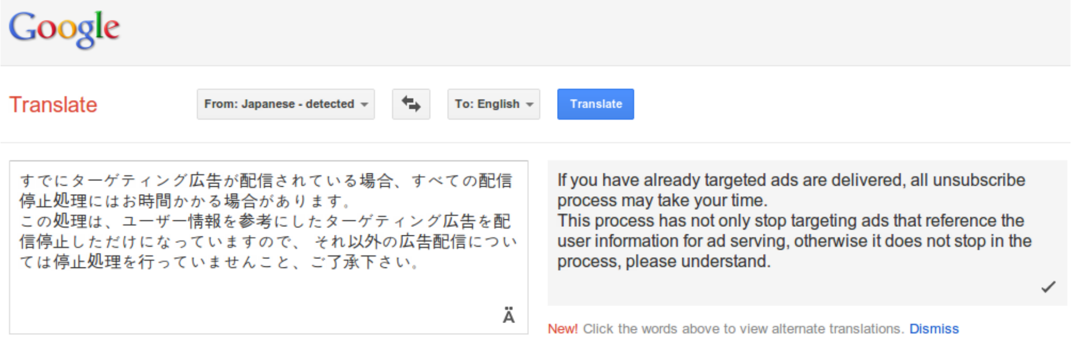
© 2009-11 PrivacyChoice

Figure D.4: PrivacyMark's installation website. Users had difficulty using PrivacyMark to opt out since it asks the user to drag the PrivacyMark icon to the browser's Bookmarks Toolbar. This toolbar is not enabled by default in newer versions of Firefox, which led to confusion for users.



Name	Category	Affiliations	Opt-out
140 Proof	Other	IAB	go to site
24/7 Real Media	Ad server, Network	AAAA, DAA, IAB, iASH Europe, NAI, TRUSTe	go to site
33Across	Network, Optimizer	DAA, IAB, NAI	opted out
4info	Network	IAB	go to site
Accuen Media	Agency, Network		opt-out request sent
Acerno	Network	NAI	opted out
Acxiom (Relevance-X)	Offline data aggregator, Online data aggregator, Retargeter	DAA, IAB	opt-out request sent
Ad Desk	Ad server, Network	IAB, NAI	opt-out request sent
Ad River	Ad server		go to site
Adap.tv	Ad server, Exchange	IAB	opt-out request sent
Adara Media	Analytics provider, Data solution	DAA, NAI, TRUSTe	opted out
Adblade	Network		go to site
AdBrite	Exchange	DAA, IAB, NAI, TRUSTe	opted out
AdBuyer	Demand side platform		opt-out request sent
AdCentric (Cossette)	Ad server		opt-out request sent
Adchemy	Demand side platform	DAA, NAI	opted out
Adconion Media Group	Network, Optimizer	DAA, IAB, NAI	go to site

Figure D.5: Evidon's opt-out page. Although participants were more successful opting out of companies on Evidon's page than on the DAA's page, they were confused and annoyed by the site's terminology. After choosing "Select All" and opting out, users receive one of three different messages for each company: "opted out," "opt-out request sent," or "go to site." Participants were particularly unhappy with the ambiguity of "opt-out request sent" and the extra effort required to "go to site" to opt out.



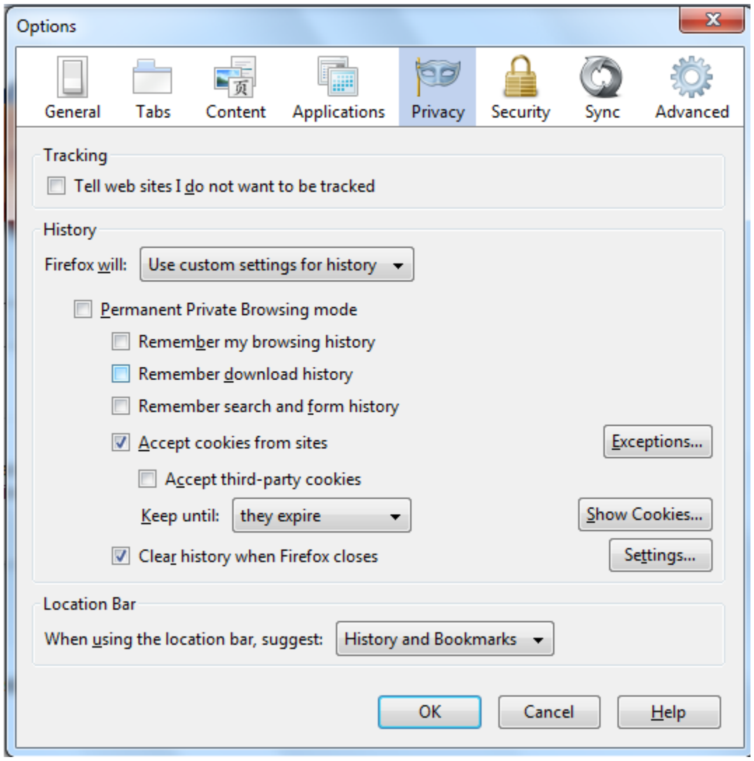
すでにターゲティング広告が配信されている場合、すべての配信停止処理にはお時間かかる場合があります。
この処理は、ユーザー情報を参考にしたターゲティング広告を配信停止しただけになっていますので、 それ以外の広告配信については停止処理を行っていませんこと、ご了承下さい。

If you have already targeted ads are delivered, all unsubscribe process may take your time.
This process has not only stop targeting ads that reference the user information for ad serving, otherwise it does not stop in the process, please understand.

New! Click the words above to view alternate translations. Dismiss

Figure D.6: Two determined participants chose to "go to site" for the companies from which they were unable to opt out automatically. The second of these participants opted out for 47 minutes. Since a handful of opt-out pages were offered only in languages other than English, he used Google Translate to learn how to opt out and confirm that his opt-out had been recorded. This figure shows part of the translation he generated while opting out of Freak Out, one of the Japanese networks.

D.3 Screenshots of tools built into browsers



Tracking

□ Tell web sites I do not want to be tracked

History

Firefox will: Use custom settings for history

□ Permanent Private Browsing mode

□ Remember my browsing history

□ Remember download history

□ Remember search and form history

☑ Accept cookies from sites

□ Accept third-party cookies

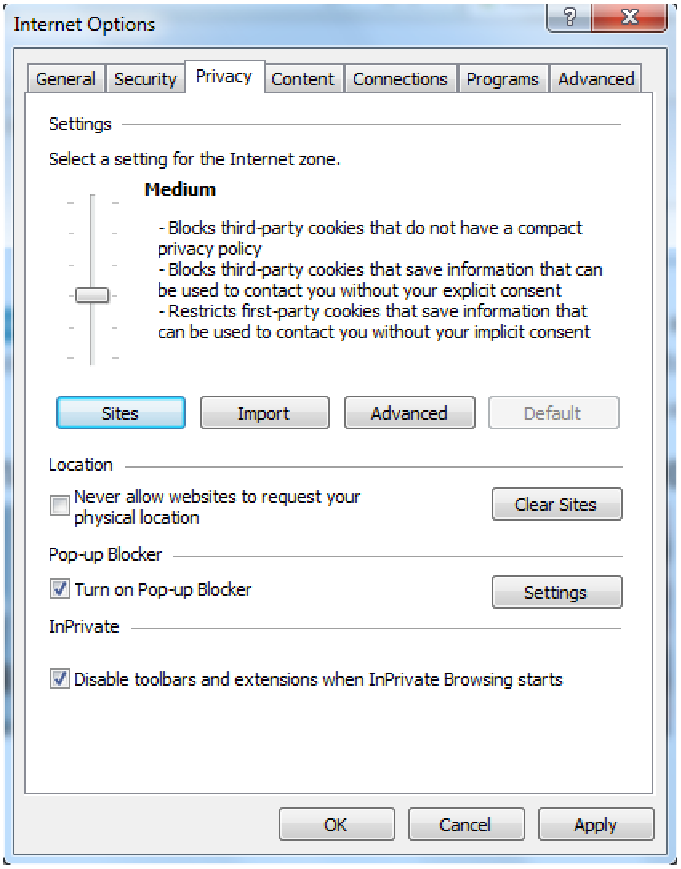
Keep until: they expire

☑ Clear history when Firefox closes

Location Bar

When using the location bar, suggest: History and Bookmarks

Figure D.7: Firefox 5's built-in privacy features. Using the privacy options built into the Mozilla Firefox 5 browser, participants were generally successful in blocking third-party cookies, which are often cookies from advertisers, while still accepting first-party cookies. Although Firefox doesn't show any of the checkboxes seen in this figure until the users chooses "Firefox Will: Use custom settings for history," all participants were able to find these options following the instructions on Mozilla's website that they read before configuring Firefox.



Settings

Select a setting for the Internet zone.

Medium

- Blocks third-party cookies that do not have a compact privacy policy
- Blocks third-party cookies that save information that can be used to contact you without your explicit consent
- Restricts first-party cookies that save information that can be used to contact you without your implicit consent

Location

□ Never allow websites to request your physical location

Pop-up Blocker

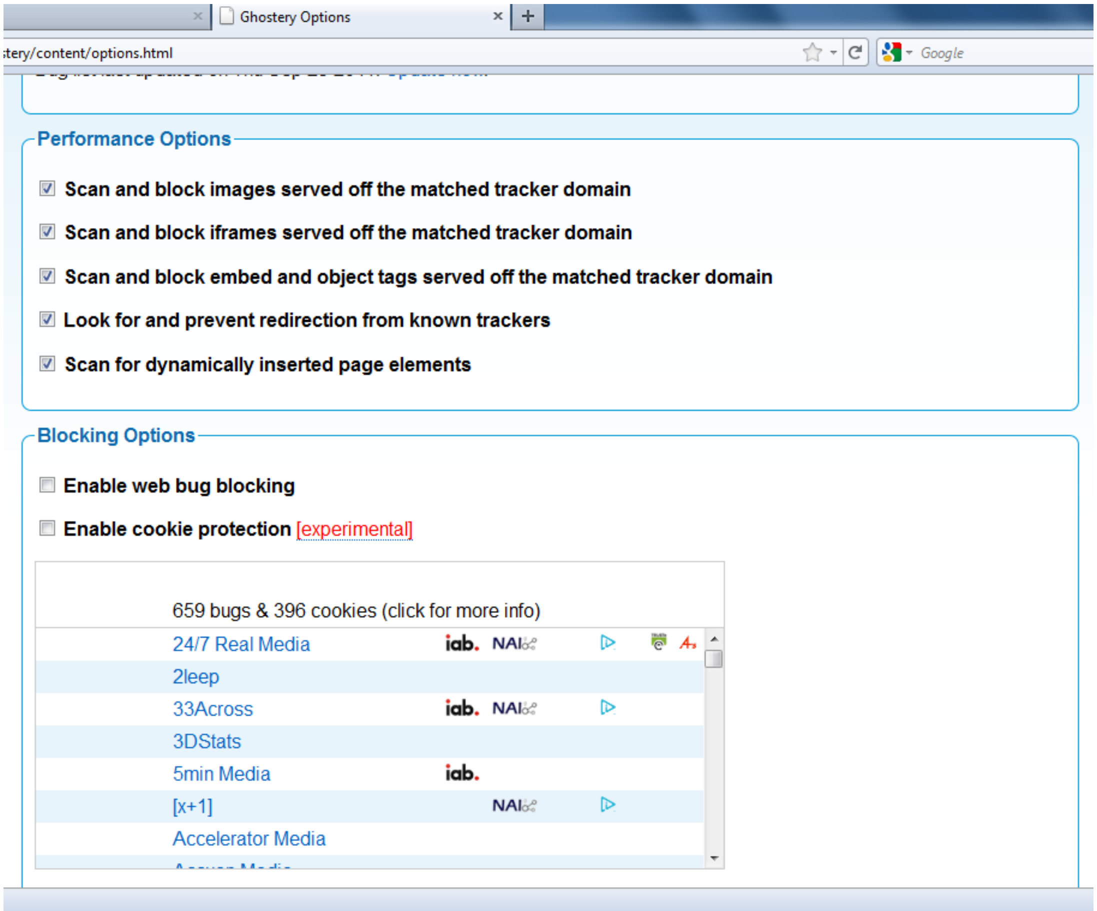
☑ Turn on Pop-up Blocker

InPrivate

☑ Disable toolbars and extensions when InPrivate Browsing starts

Figure D.8: Internet Explorer 9's built-in privacy features. In contrast to Firefox, participants testing Internet Explorer 9 were unable to block third-party cookies while enabling first-party cookies. The option to perform this blocking is part of the "advanced" menu, which no users chose to view. Users were confused by the slider for choosing privacy settings, neither understanding its references to compact privacy policies nor the options it presented.

D.4 Screenshots of blocking tools



Performance Options

☑ Scan and block images served off the matched tracker domain

☑ Scan and block iframes served off the matched tracker domain

☑ Scan and block embed and object tags served off the matched tracker domain

☑ Look for and prevent redirection from known trackers

☑ Scan for dynamically inserted page elements

Blocking Options

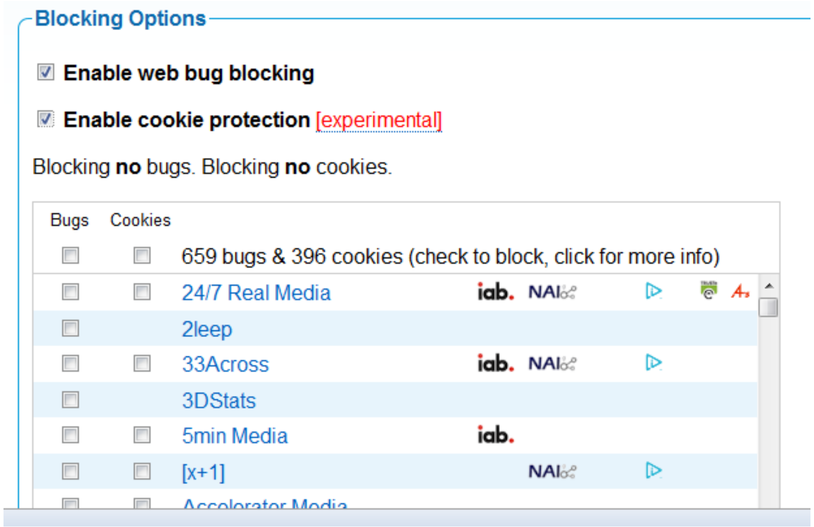
□ Enable web bug blocking

□ Enable cookie protection [experimental]

659 bugs & 396 cookies (click for more info)

24/7 Real Media	iab.	NAI
2leep		
33Across	iab.	NAI
3DStats		
5min Media	iab.	
[x+1]		NAI
Accelerator Media		

Figure D.9: Ghostery's configuration interface. Users found the configuration of Ghostery relatively confusing. Although it's intended as privacy software, Ghostery doesn't block any trackers by default. On this configuration screen, users must select both "Enable web bug blocking" and "Enable cookie protection" for full protection. Some participants were apprehensive about using cookie protection since it is labeled "experimental" in red, a color that often indicates a problem.



Blocking Options

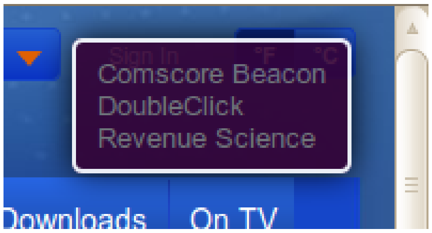
☑ **Enable web bug blocking**

☑ **Enable cookie protection** [experimental]

Blocking **no** bugs. Blocking **no** cookies.

Bugs	Cookies			
□	□	659 bugs & 396 cookies (check to block, click for more info)		
□	□	24/7 Real Media	iab.	NAI
□		2leep		
□	□	33Across	iab.	NAI
□		3DStats		
□	□	5min Media	iab.	
□	□	[x+1]		NAI
□	□	Accelerator Media		

Figure D.10: Ghostery's configuration interface, once cookie protection has been enabled. Once a user chooses to "enable web bug blocking" or "enable cookie protection," she must further select from a list of companies that appears for this blocking to take effect. While it comes first on the list, the button to select all options is unlabeled. Furthermore, participants didn't understand the difference between blocking web bugs and enabling cookie protection.



Comscore Beacon
DoubleClick
Revenue Science

Figure D.11: The alert Ghostery presents on each site a user visits. As users visit websites, Ghostery presents an ephemeral pop-up alert indicating which companies have trackers on that page. Participants noticed and correctly understood that those companies were attempting to track their browsing on the page.

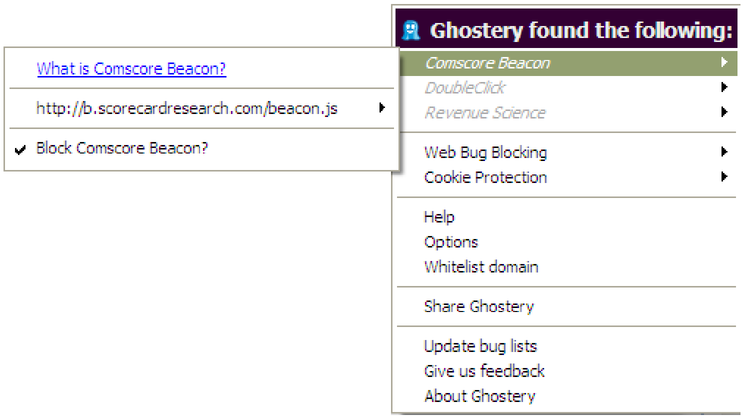
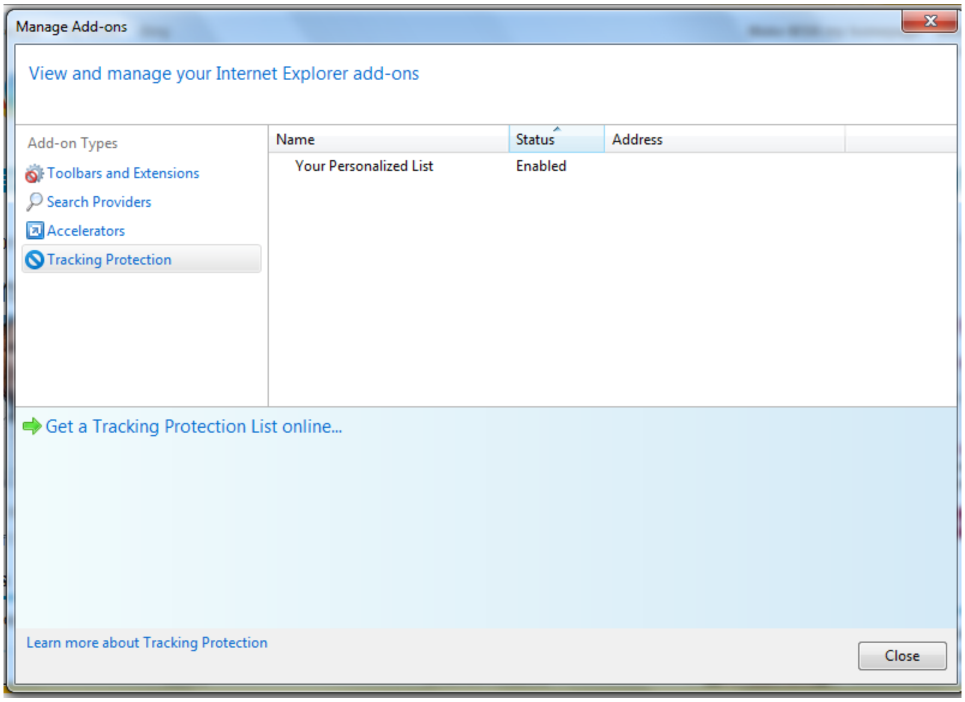
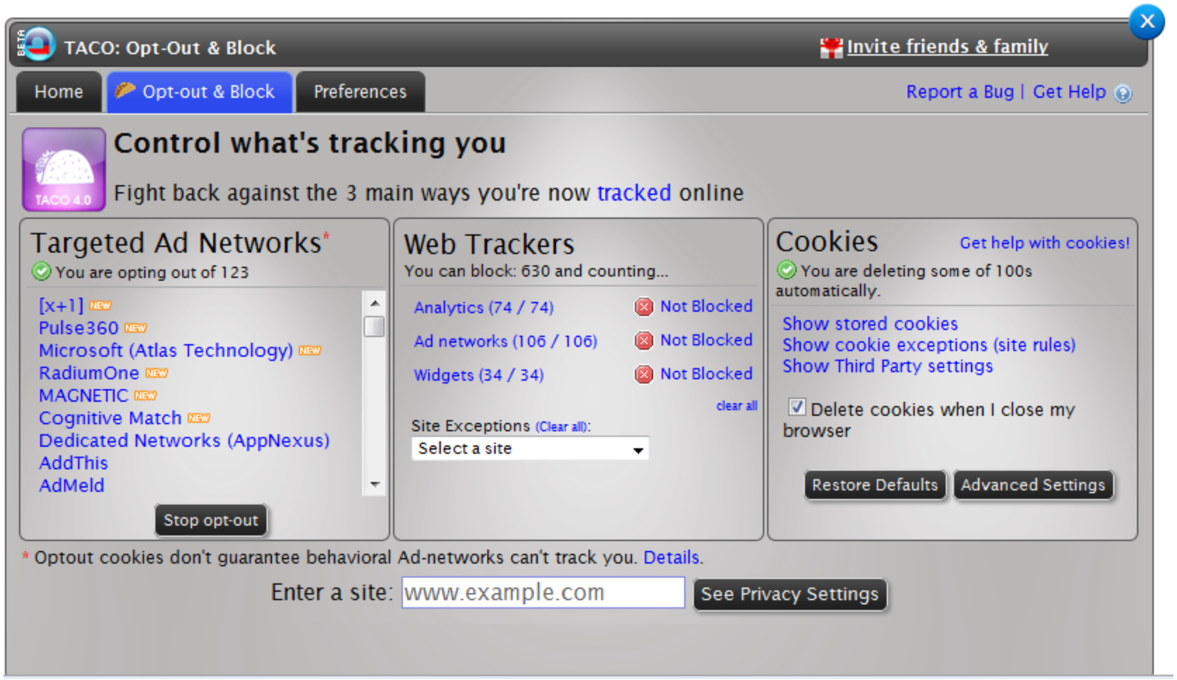


Figure D.12: The Ghostery options that appear when a user clicks on its icon in the toolbar. A user is able to block or unblock particular trackers.


Manage Add-ons
View and manage your Internet Explorer add-ons
Add-on Types
Toolbars and Extensions
Search Providers
Accelerators
Tracking Protection
Get a Tracking Protection List online...
Learn more about Tracking Protection
Close

Name	Status	Address
Your Personalized List	Enabled	

Figure D.13: Internet Explorer's Tracking Protection List configuration screen, after enabling "Your Personalized List." Users must click "Get a Tracking Protection List Online" to block tracking; participants in our study did not realize this.



TACO: Opt-Out & Block

Invite friends & family

Home | Opt-out & Block | Preferences

Report a Bug | Get Help

Control what's tracking you

Fight back against the 3 main ways you're now tracked online

Targeted Ad Networks*

✓ You are opting out of 123

- [x+1] NEW
- Pulse360 NEW
- Microsoft (Atlas Technology) NEW
- RadiumOne NEW
- MAGNETIC NEW
- Cognitive Match NEW
- Dedicated Networks (AppNexus)
- AddThis
- AdMeld

Stop opt-out

Web Trackers

You can block: 630 and counting...

Analytics (74 / 74)	☒ Not Blocked
Ad networks (106 / 106)	☒ Not Blocked
Widgets (34 / 34)	☒ Not Blocked

clear all

Site Exceptions (Clear all):

Select a site

Cookies

Get help with cookies!

✓ You are deleting some of 100s automatically.

Show stored cookies

Show cookie exceptions (site rules)

Show Third Party settings

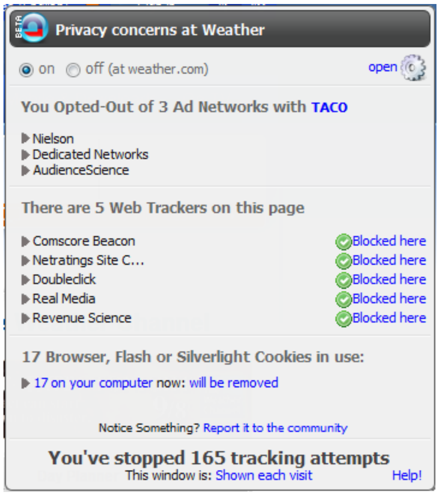
☑ Delete cookies when I close my browser

Restore Defaults | Advanced Settings

* Optout cookies don't guarantee behavioral Ad-networks can't track you. Details.

Enter a site: www.example.com | See Privacy Settings

Figure D.14: The interface for configuring TACO's blocking and opt-out features. Simply accessing this screen, which users found confusing, requires four steps. Once here, the user is presented with three categories of tracking: "Targeted Ad Networks," "Web Trackers," and "Cookies." The distinction between these categories was opaque to users. To enable blocking, a user must click on the three "Not Blocked" pieces of text that don't appear to be clickable. Even after choosing all three available categories, the user is informed, "You are blocking some of 630." No participants were ever told they were blocking all 630.



Privacy concerns at Weather

● on ○ off (at weather.com) open

You Opted-Out of 3 Ad Networks with TACO

- Nielson
- Dedicated Networks
- AudienceScience

There are 5 Web Trackers on this page

Comscore Beacon	✓ Blocked here
Netratings Site C...	✓ Blocked here
Doubleclick	✓ Blocked here
Real Media	✓ Blocked here
Revenue Science	✓ Blocked here

17 Browser, Flash or Silverlight Cookies in use:

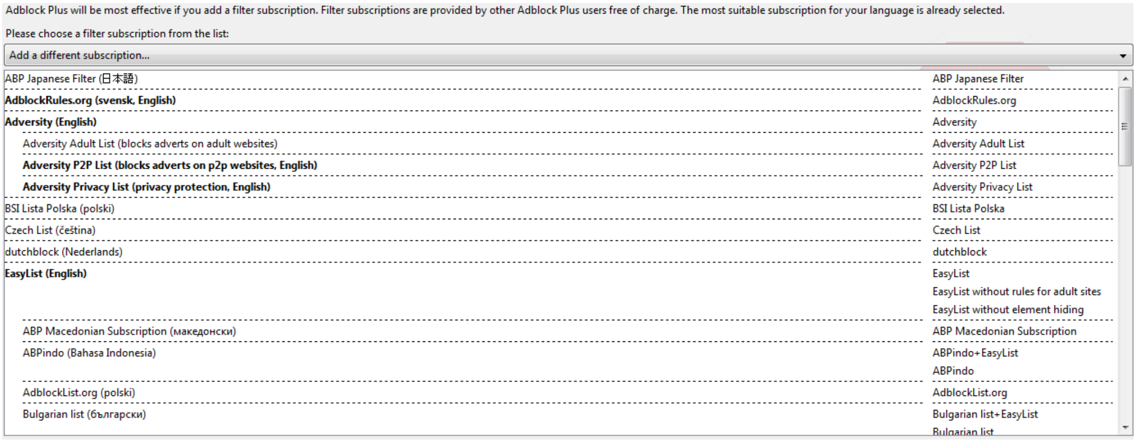
- 17 on your computer now: will be removed

Notice Something? Report it to the community

You've stopped 165 tracking attempts

This window is: Shown each visit Help!

Figure D.15: The alert TACO presents on each site a user visits. The distinction between "ad networks" and "web trackers" was confusing to users, as was the cumulative nature of "tracking attempts" stopped.

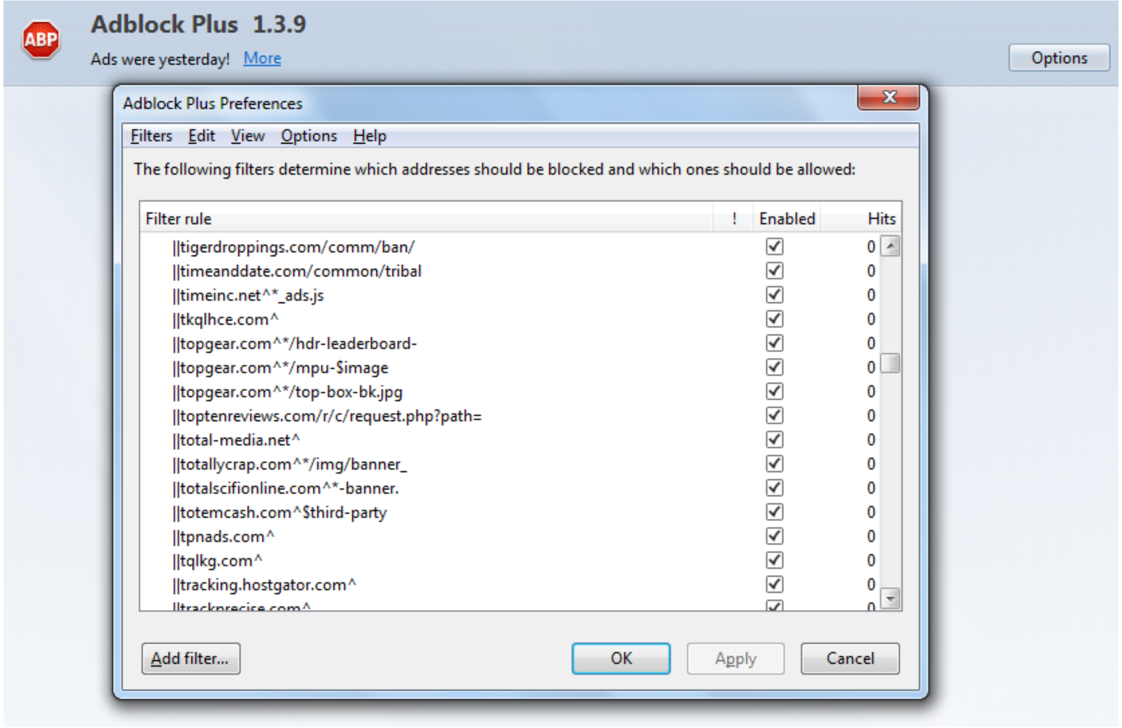


Adblock Plus will be most effective if you add a filter subscription. Filter subscriptions are provided by other Adblock Plus users free of charge. The most suitable subscription for your language is already selected.

Please choose a filter subscription from the list:

ABP Japanese Filter (日本語)	ABP Japanese Filter
AdblockRules.org (svensk, English)	AdblockRules.org
Adversity (English)	Adversity
Adversity Adult List (blocks adverts on adult websites)	Adversity Adult List
Adversity P2P List (blocks adverts on p2p websites, English)	Adversity P2P List
Adversity Privacy List (privacy protection, English)	Adversity Privacy List
BSI Lista Polska (polski)	BSI Lista Polska
Czech List (čeština)	Czech List
dutchblock (Nederlands)	dutchblock
EasyList (English)	EasyList EasyList without rules for adult sites EasyList without element hiding
ABP Macedonian Subscription (македонски)	ABP Macedonian Subscription
ABPindo (Bahasa Indonesia)	ABPindo+EasyList ABPindo
AdblockList.org (polski)	AdblockList.org
Bulgarian list (български)	Bulgarian list+EasyList Bulgarian list

Figure D.16: The main configuration screen for AdBlock Plus. The instructions at the top ask the user to subscribe to a filtering list. In contrast to Internet Explorer TPLs, all participants subscribed to a filter list when testing AdBlock Plus since the interface prompts the user to do so. However, subjects didn't know which filtering subscription to select or how to comparatively evaluate these subscriptions.



The following filters determine which addresses should be blocked and which ones should be allowed:

Filter rule	!	Enabled	Hits
||tigerdroppings.com/comm/ban/		☑	0
||timeanddate.com/common/tribal		☑	0
||timeinc.net^*_ads.js		☑	0
||tkqlhce.com^		☑	0
||topgear.com^*/hdr-leaderboard-		☑	0
||topgear.com^*/mpu-$image		☑	0
||topgear.com^*/top-box-bk.jpg		☑	0
||toptenreviews.com/r/c/request.php?path=		☑	0
||total-media.net^		☑	0
||totallycrap.com^*/img/banner_		☑	0
||totalscifionline.com^*-banner.		☑	0
||totemcash.com^$third-party		☑	0
||tpnads.com^		☑	0
||tqlkg.com^		☑	0
||tracking.hostgator.com^		☑	0
||trackprecise.com^		☑	0

Figure D.17: The options screen for AdBlock Plus, showing filter rules. Resolving problems was difficult for AdBlock Plus users since they didn't know which filters from a particular list had affected a particular website. If a user is trying to unblock filters that are causing problems on a website, she will be presented with an "options" screen containing all filter rules. Only experts can interpret these rules.

D.5 Participants' opinions about tools

We summarize here what each participants told us about the tool he or she tested during the exit interview. In some cases participants' perceptions are not accurate or their comments reflect what they read about a tool more than what they personally experienced.

Tool	Features liked	Features disliked	Desired features	Benefits perceived
Ghostery				
GH-1	Configuration on a per-tracker basis	None	Tutorial about trackers and tool usage	Controls information disclosure
GH-2	Ability to block advertising	Slow to configure	None	Not seeing ads
GH-3	Observing what trackers are on the web page, ability to block trackers on a web page	None	More contextual awareness	Controls information disclosure
GH-4	Seeing fewer ads, awareness of who is collecting data	None	Ability to remember choices	Fewer pop-up ads, awareness of trackers
GH-5	Awareness of source of ads, easy to configure, configuration on a per-tracker basis	Tool blocked a flash video	None	Control ads, controls information disclosure
TACO				
TACO-1	Awareness of who is collecting data	None	None	Allows users to specify who can track them, provides better awareness
TACO-2	Not seeing ads, fill out forms, removing cookies	None	None	Prevents all ads, removes cookies
TACO-3	Awareness of who is collecting data, awareness of blocked ads	None	None	Prevents tracking
TACO-4	Awareness of trackers	Difficult to use, creates false sense of security	Option to block all ads	Fewer ads
TACO-5	Awareness of trackers, ability to block trackers	None	Ability to block only certain trackers	Allows users to specify who can track them
AdBlock Plus				
ABP-1	Not seeing ads	Difficult to use	Better interface, easier access to preferences	Fewer distractions
ABP-2	Easy to configure	None	More information about what is blocked	Protects privacy, fewer distractions
ABP-3	Easy to use	Unintuitive	Better notice that installation is successful	Fewer ads, prevents tracking
ABP-4	Easy to use	Contents of filter lists unknown	None	Less annoying ads, prevents tracking
ABP-5	Easy to install	Difficult to configure	Ability to allow desired ads, ability to preview blocked ads	Fewer ads, improved security
IE TPL				
TPL-1	Ability to customize what is blocked	None	Better instructions, a help button	Provides more appropriate content in searches

Continued on next page ...

Continued from previous page

Tool	Features liked	Features disliked	Features wanted	Benefits perceived
TPL-2	Ability to customize what is blocked	Difficult to install and use	Better instructions	More privacy, blocks third-party cookies
TPL-3	Ability to customize what is blocked	Couldn't figure out how to personalize tracking list	Ability to know what is blocked, ability to unblock some trackers	Stops targetted ads
TPL-4	None	No feedback that tool is working	Feedback that tool is working	More privacy, controls information disclosure
TPL-5	Fewer ads	None	Notice that user is being tracked, mechanism for knowing which trackers to trust	Fewer ads

Table D.2: Participants' opinions about blocking tools, paraphrased from exit interviews.

Tool	Features liked	Features disliked	Desired features	Benefits perceived
		DAA Consumer Choice		
DAA-1	Easy to configure/use	None	None	Controls information disclosure
DAA-2	Speed of configuration	Not knowing why companies participate, not knowing if opt-out will be honored	Would prefer blocking tool	Prevents some companies from targeting ads
DAA-3	Easy to configure/use	None	None	Allows users to specify who can track them
DAA-4	Easy to configure/use	Not knowing if opt-out will be honored	More companies to choose from, easier website navigation	Allows users to specify who can track them
DAA-5	Listing of companies offering opt-out	None	Indication of what opting out means	Less obtrusive ads
		Evidon Global Opt-Out		
EV-1	Listing of companies offering opt-out	"Select all" feature does not work	Notification of successful opt-out	Fewer ads, fewer third-party cookies
EV-2	The "select all" feature	"Select all" feature does not work, not knowing if opt-out will be honored	Make opt-out feature more prominent	Better awareness of which companies perform tracking
EV-3	Configuration on a per-tracker basis	None	Knowing the websites on which tracking is performed	Allows users to specify who can track them
EV-4	Configuration on a per-tracker basis	Not knowing if opt-out will be honored	More information about what the affiliations such as NAI and DAA are, assurance that the opt-outs are honored	Allows users to specify who can track them
EV-5	Easy to configure/use	Time-consuming to configure	Better organized list of trackers	More privacy
		Privacy Mark		
PM-1	Easy to configure/use	None	Assurance that the opt-outs are honored	Blocking search-based and contextual ads

Continued on next page . . .

Continued from previous page

Tool	Features liked	Features disliked	Features wanted	Benefits perceived
PM-2	Not seeing ads	Unable to configure	Assurance that the opt-outs are honored, ability to configure preferences	Controls information disclosure
PM-3	None	Not knowing if opt-out will be honored	Assurance that the opt-outs are honored, ability to configure preferences	Controls information disclosure, fewer ads
PM-4	None	Lack of information about trackers, creates false sense of security	Assurance that the opt-outs are honored	None
PM-5	Configuration on a per-tracker basis	Time-consuming to configure	Assurance that the opt-outs are honored, ability to configure preferences	Controls information disclosure

Table D.3: Participants' opinions about opt-out tools, paraphrased from exit interviews.

Tool	Features liked	Features disliked	Desired features	Benefits perceived
		IE Privacy Settings		
IE-1	None	Difficult to undo blocking	None	None
IE-2	Block tracking cookies	Difficult to configure	Notification of who is tracking and what collected information is used for	Allows users to specify who can track them
IE-3	Ability to configure third-party cookies	Lack of information about cookies	Assurance that the tool is working	Identity theft prevention
IE-4	Blocking pop-ups	None	None	Blocks third-party cookies, hides physical location
IE-5	Easy to configure	None	None	Blocks cookies
		Firefox Privacy Settings		
FF-1	Ability to stop specific websites from tracking, ability to see who is tracking	None	None	Fewer ads
FF-2	Ability to stop specific websites from tracking	None	Indicate which cookies are being used for tracking	Feeling of security, allows users to specify who can track them
FF-3	Not seeing ads	None	None	Controls information disclosure
FF-4	Block third-party cookies, clear browsing history, browse in private mode	None	None	More privacy
FF-5	Blocks websites	Difficult to remember what is blocked, perceived as ineffective	Simplify configuration	None

Continued on next page ...

Continued from previous page

Tool	**Features liked**	**Features disliked**	**Features wanted**	**Benefits perceived**

Table D.4: Participants' opinions about built-in browser tools, paraphrased from exit interviews.

D.6 Participants' understanding of tool capabilities

During the testing session, we asked participants multiple-choice questions that tested their understanding of the tools' capabilities. We asked some questions twice, once before and once after the browsing scenarios; we asked others only before or only after the browsing scenarios. Participants could respond with the answers true, false, or unsure. The tables in this section show the questions that the participants answered and the percentage of correct answers per tool. Overall, participants showed a lack of understanding about the tools' capabilities.

Question	Ghostery	TACO	ABP	IE-TPL	Privacy Mark	DAA	Evidon	Firefox	IE
I will not see advertising on webpages I visit	False (40%)	False (60%)	True (40%)	False (60%)	False (N/A)	False (80%)	False (100%)	False (80%)	False (80%)
I will be more secure from computer viruses	False (60%)	False (80%)	False (80%)	False (80%)	False (N/A)	False (20%)	False (40%)	False (20%)	False (0%)
While using this tool, if I delete the cookies that my browser has stored, I will protect my privacy even more	True (20%)	True (100%)	True (80%)	True (60%)	False (N/A)	False (0%)	False (20%)	True (60%)	True (80%)

Table D.5: This table shows the questions that we asked only before the browsing scenarios, after completing the changing configuration task. The table contains the correct answer to each question for each tool, and the percentage of participants who answered correctly. PrivacyMark participants did not perform the changing configuration task and were not required to answer these questions. Firefox and IE settings participants exhibited a particular low understanding for the second question. DAA and Evidon participants exhibited a very low understanding for the third question. In particular, DAA and Evidon participants did not understand that deleting cookies would render the testing tool ineffective

Question	Ghostery	TACO	ABP	IE-TPL	Privacy Mark	DAA	Evidon	Firefox	IE
I can block particular advertising companies from delivering any ads to me	True (80%,80%)	True (60%,80%)	True (80%,100%)	True (60%,60%)	False (*,80%)	False (40%,60%)	False (60%,40%)	False (60%,60%)	False (20%,60%)
I will see fewer ads that are tailored to my interests	True (100%,80%)	True (60%,80%)	True (80%,80%)	True (60%,60%)	True (*,100%)	True (80%,80%)	True (100%,100%)	True (80%,100%)	True (60%,80%)
I can see which online advertising companies are delivering ads to me	True (100%,100%)	True (80%,100%)	True (40%,100%)	False (20%,40%)	False (*,60%)	False (40%,40%)	False (80%,20%)	False (60%,40%)	False (60%,60%)
I can block particular advertising companies from delivering ads that are tailored specifically to me	True (80%,100%)	True (100%,80%)	True (80%,100%)	True (40%,60%)	True (*,20%)	True (100%,100%)	True (80%,80%)	True (100%,20%)	True (40%,40%)
While using this tool, my computer won't download any cookies while browsing the Internet	False (40%,60%)	False (60%,80%)	False (60%,60%)	False (40%,80%)	False (*,60%)	False (40%,60%)	False (60%,60%)	False (40%,100%)	False (40%,60%)

Table D.6: This table shows the questions that we asked both before and after the browsing scenarios. The table contains the correct answer to each question for each tool, and the percentage of participants who answered correctly before and after the browsing scenarios, respectively. PrivacyMark participants only answered these questions after the browsing scenarios. Blocking tools exhibited a clear improvement in understanding after having used the tool. DAA and Evidon participants did not show much improvement. Firefox and IE settings participants improved understanding for some questions but reduced it for some others, showing problems understanding the tools' capabilities.

Question	Ghostery	TACO	ABP	IE-TPL	Privacy Mark	DAA	Evidon	Firefox	IE
I can block all advertising companies from delivering ads that are tailored specifically to me	False (40%)	False (20%)	True (80%)	False (60%)	False (60%)	False (80%)	False (40%)	False (40%)	False (40%)
When I visit a website, I will never see any advertising based on other websites I've visited	False (40%)	False (60%)	True (60%)	False (60%)	False (80%)	False (60%)	False (20%)	False (40%)	False (40%)
I can decide when to allow websites that I visit to create a profile of me based on my activities on their own websites.	False (0%)	False (40%)	False (20%)	False (0%)	False (60%)	False (40%)	False (40%)	True (100%)	True (60%)
If I am visiting Amazon.com, I will not see advertisements based on other products I've viewed on Amazon.com	False (80%)	False (60%)	False (0%)	False (60%)	False (40%)	False (60%)	False (80%)	False (20%)	False (40%)
If an advertising company delivers ads to both Walmart.com and CNN.com, I could use this tool to prevent that advertising company from creating a profile of me based on the products I view on Walmart.com and the stories I read on CNN.com	True (100%)	True (80%)	True (60%)	True (80%)	False (40%)	False (20%)	False (0%)	True (60%)	True (60%)
It will be more difficult technologically for advertising networks to track which sites I visit	True (60%)	True (100%)	True (60%)	True (80%)	False (0%)	False (0%)	False (20%)	True (40%)	True (40%)

Table D.7: This table shows the questions that were asked only after the browsing scenarios. The table contains the correct answer to each question for each tool, and the percentage of participants who answered correctly. Evidon, DAA and PrivacyMark participants' low understanding for the last two questions in this table suggests that participants incorrectly believe that these tools prevent tracking.

Appendix E

What Do Online Behavioral Advertising Privacy Disclosures Communicate to Users?

E.1 Participants' Understanding of Privacy Disclosures

Q1: As best as you can tell, what is the purpose of placing this symbol and phrase [icon+tagline shown] on the top right corner of the above ad?	Why did I get this ad?	Learn about your ad choices	Interest based ads	AdChoices	Configure ad preferences	Sponsor ads	Blank
To tell you that the ad is targeted to you	82%	67%	62%	55%	45%	22%	22%
To allow you to choose which types of products appear in ads that you see	45%	78%	33%	50%	72%	18%	24%
To tell you that this ads are from a legitimate company	39%	33%	33%	46%	22%	57%	21%
To give you information about placing advertisements on this website	33%	39%	31%	46%	29%	38%	23%
To attract your attention to the ad	45%	42%	55%	41%	34%	49%	42%
To advertise the company that is delivering this ad	37%	41%	41%	65%	35%	57%	34%
To give you more information about the advertised product	35%	35%	35%	30%	28%	40%	51%
To get your reactions to the ad	35%	33%	22%	19%	28%	13%	24%
To get you to click on the ad	54%	55%	58%	37%	41%	48%	47%

Table E.1: Participants' responses to the question, "As best as you can tell, what is the purpose of placing this symbol and phrase [icon+tagline shown] on the top right corner of the above ad?' The percentage of participants who answered "Probably" or "Definitely" is shown.

Q2: To what extent, if any, does this combination of the symbol and phrase [icon+tagline shown], placed on the top right corner of the above ad suggest the following?	Why did I get this ad?	Learn about your ad choices	Interest based ads	AdChoices	Configure ad preferences	Sponsor ads	Blank
This ad has been tailored based on websites you have visited on the past	80%	66%	68%	58%	58%	26%	34%
The ads you see on the news website are based on your visits to other websites	77%	62%	66%	56%	47%	28%	32%
This website shows ads that are chosen to match your needs	78%	70%	66%	65%	67%	26%	31%
These ads have been chosen to be relevant to you	83%	73%	72%	68%	62%	27%	35%
You can stop tailored advertising	18%	31%	15%	18%	41%	6%	13%
You can click on that symbol/phrase	91%	85%	58%	75%	83%	67%	71%
You can turn off advertisements on this website	12%	20%	6%	12%	41%	7%	13%
This ad is from one of the website's premier partners	40%	34%	41%	44%	33%	66%	33%
You can choose to learn about the advertised product	50%	54%	58%	58%	43%	66%	55%
You can choose which ads you want to see on this website	37%	71%	26%	42%	72%	13%	19%

Table E.2: Participants' responses to the question, "To what extent, if any, does this combination of the symbol and phrase [icon+tagline shown], placed on the top right corner of the above ad suggest the following?" The percentage of participants who answered "Probably" or "Definitely" is shown.

Q3: What do you think would happen if you click on that symbol or that phrase?	Why did I get this ad?	Learn about your ad choices	Interest based ads	AdChoices	Configure ad preferences	Sponsor ads	Blank
It will take you to a page where you can tell the advertising company that you do not want to receive tailored ads	28%	34%	17%	27%	50%	16%	20%
It will take you to the advertised company site	45%	52%	64%	60%	39%	74%	71%
It will take you to a page where you can buy advertisements on this website	18%	32%	29%	45%	15%	45%	27%
It will take you to a page where you can tell the advertising company whether you are or not interested in the advertised product/service	46%	58%	36%	47%	60%	28%	33%
It will take you to a page where you can tell the advertising company what products/services you are interested in	51%	71%	50%	59%	73%	31%	40%
More ads will pop-up	46%	51%	57%	56%	42%	63%	57%
You will let the advertising company know that you are interested in those products	43%	53%	59%	51%	50%	53%	52%

Table E.3: Participants' responses to the question, "What do you think would happen if you click on that symbol or that phrase?" The percentage of participants who answered "Probably" or "Definitely" is shown.

Appendix F

Factors That Affect Users' Willingness to Share Information with Online Advertisers

F.1 Study One: Factor Analysis

Factor	Statement	Disclose	α if item removed	Factor Loading
Browsing Information ($\alpha = 0.92$)	How long I spent on each page of the *WebMD* website	25.3%	0.89	0.80
	My responses to health-related surveys	30.1%	0.91	0.77
	The medications I am taking (inferred from my interactions with the site)	18.7%	0.92	0.62
	The pages I've visited on the *WebMD* website	42.8%	0.89	0.98
	Which search terms I've entered on the *WebMD* website	41.2%	0.89	0.99
Computer Information ($\alpha = 0.93$)	The name and version of the web browser (e.g., Internet Explorer 9, Firefox 18.0.1, Safari 6.0.2, etc.) that I use to visit the *WebMD* website	42.9%	0.93	0.97
	The type of operating system (e.g., Windows, Mac, etc.) of my computer	44.8%	0.93	0.94
Demographic Information ($\alpha = 0.94$)	My age	39.1%	0.94	0.67
	My gender	46.3%	0.94	0.67
	My highest level of education	27.4%	0.93	0.89
	My hobbies	34.9%	0.94	0.76
	My income bracket	9.6%	0.94	0.65
	My marital status	25.2%	0.93	0.94
	My religion	17.3%	0.94	0.97
	My political preferences	16.7%	0.94	0.98
	My sexual orientation	21.1%	0.94	0.99
Location Information ($\alpha = 0.91$)	The country from which I'm visiting the *WebMD* website	52.7%	0.90	0.60
	The state from which I'm visiting the *WebMD* website	42.9%	0.90	0.79

Continued on next page . . .

Continued from previous page

Factor	Statement	Disclose	α if item removed	Factor Loading
	The town or city from which I'm visiting the *WebMD* website	24.4%	0.82	1.08
	The ZIP code from which I'm visiting the *WebMD* website	22.6%	0.85	1.06
Personally Identifiable ($\alpha = 0.81$)	My email address	14.0%	NA	*
	My name	13.6%	NA	*
Did not conform to a factor	My address	1.9%	NA	NA
	My credit card number	0.6%	NA	NA
	My credit score bracket	4.5%	NA	NA
	My phone number	2.6%	NA	NA
	My Social Security number	0.6%	NA	NA
	My weight and height	25.5%	NA	NA
	The exact address from which I'm visiting the *WebMD* website	3.9%	NA	NA
	The IP address of my computer (i.e., a computer identifier assigned by your Internet service provider)	15.3%	NA	NA

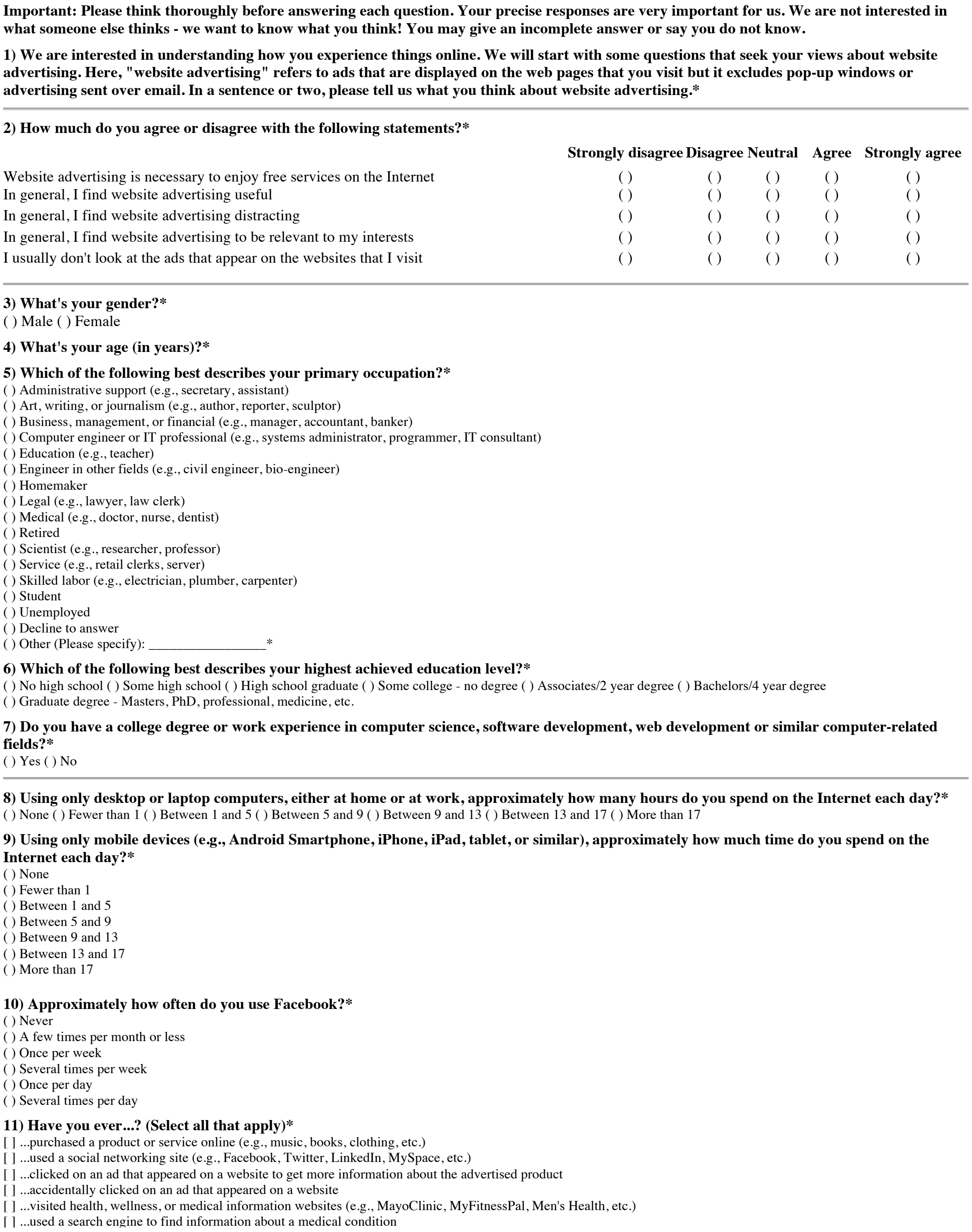
Table F.1: Participants' willingness to disclose different types of information for OBA purposes (N=2,912). Using exploratory factor analysis, we grouped 22 of the 30 types of information into five factors. The *disclose* column lists the percentage of participants who agreed or strongly agreed that they would be willing to disclose that type of information. The α *if item removed* column displays Cronbach's Alpha, the correlation of items in the group, if that item were to be removed from the group. The *loading* column displays the factor loading from exploratory factor analysis, or NA for types of information that had below 0.60 factor loading for all five factors. The two types of data with a factor loading of "*" did not meet the criteria for inclusion with a factor, while no types of information loaded sufficiently onto the fifth factor. Since these two types of information were correlated with each other ($\alpha = 0.81$), we considered them to be the fifth factor.

F.2 Study One: Multivariate Multiple Regression Model

Independent Variable	Control Category	β Estimate	SE	t-value	P>|t|
Dependent Variable: Browsing information					
Scope: Health site + Facebook	Only health site	-0.986	0.088	-11.202	<0.001
Scope: All sites	Only health site	-0.297	0.088	-3.363	<0.001
Retention: Indefinite	One day	-0.465	0.082	-5.639	<0.001
Facebook usage	Not Facebook user	0.154	0.043	3.560	<0.001
Privacy concern	Unconcerned	-0.291	0.030	-9.667	<0.001
Like targeted ads	Don't like	0.684	0.040	16.759	<0.001
Interaction: Facebook and Retention	NA	0.312	0.097	3.209	0.001
Dependent Variable: Computer information					
Facebook usage	Not Facebook user	0.220	0.051	4.302	<0.001
Privacy concern	Unconcerned	-0.254	0.035	-7.152	<0.001
Like targeted ads	Don't like	0.590	0.048	12.242	<0.001
Dependent Variable: Demographic information					
Retention: Indefinite	One day	-0.172	0.008	-2.248	0.025
Age	NA	-0.004	0.002	-2.228	0.026
IT experience	None	-0.073	0.031	-2.310	0.021
Facebook usage	Not Facebook user	0.210	0.040	5.206	<0.001
Privacy concern	Unconcerned	-0.326	0.028	-11.669	<0.001
Like targeted ads	Don't like	0.622	0.038	16.390	<0.001
Interaction: Facebook and Retention	NA	0.181	0.090	2.002	0.045
Dependent Variable: Location information					
Scope: Health site + Facebook	Only health site	-0.328	0.093	-3.531	<0.001
Retention: Indefinite	One day	-0.283	0.087	-3.249	0.001
Age	NA	0.008	0.002	3.924	<0.001
Facebook usage	Not Facebook user	0.187	0.046	4.092	<0.001
Privacy concern	Unconcerned	-0.340	0.032	-10.727	<0.001
Like targeted ads	Don't like	0.623	0.043	14.476	<0.001
Dependent Variable: Personally identifiable information					
Scope: Health site + Facebook	Only health site	0.329	0.083	3.959	<0.001
Facebook usage	Not Facebook user	0.190	0.041	4.638	<0.001
Privacy concern	Unconcerned	-0.262	0.028	-9.244	<0.001
Like targeted ads	Don't like	0.432	0.039	11.227	<0.001
Interaction: Facebook and Retention	NA	0.186	0.092	2.024	0.043

Table F.2: This table shows the multivariate multiple regression model underlying our analysis of participants' willingness to disclose information. In addition to the retention, scope, access, and site familiarity treatments, we included the following co-variates: age, gender, frequency of Facebook usage (Q10 in Appendix F.3), whether or not the participant held a degree or job in IT or a related field (Q7), privacy concerns (Q46), and whether the participant likes targeted ads (Q38). Only terms significant at α<0.05 are shown.

F.3 Study One: Survey Questions



Important: Please think thoroughly before answering each question. Your precise responses are very important for us. 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) We are interested in understanding how you experience things online. We will start with some questions that seek your views about website advertising. Here, "website advertising" refers to ads that are displayed on the web pages that you visit but it excludes pop-up windows or advertising sent over email. In a sentence or two, please tell us what you think about website advertising.*

2) How much do you agree or disagree with the following statements?*

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Website advertising is necessary to enjoy free services on the Internet	( )	( )	( )	( )	( )
In general, I find website advertising useful	( )	( )	( )	( )	( )
In general, I find website advertising distracting	( )	( )	( )	( )	( )
In general, I find website advertising to be relevant to my interests	( )	( )	( )	( )	( )
I usually don't look at the ads that appear on the websites that I visit	( )	( )	( )	( )	( )

3) What's your gender?*
( ) Male ( ) Female

4) What's your age (in years)?*

5) Which of the following best describes your primary occupation?*
( ) Administrative support (e.g., secretary, assistant)
( ) Art, writing, or journalism (e.g., author, reporter, sculptor)
( ) Business, management, or financial (e.g., manager, accountant, banker)
( ) Computer engineer or IT professional (e.g., systems administrator, programmer, IT consultant)
( ) Education (e.g., teacher)
( ) Engineer in other fields (e.g., civil engineer, bio-engineer)
( ) Homemaker
( ) Legal (e.g., lawyer, law clerk)
( ) Medical (e.g., doctor, nurse, dentist)
( ) Retired
( ) Scientist (e.g., researcher, professor)
( ) Service (e.g., retail clerks, server)
( ) Skilled labor (e.g., electrician, plumber, carpenter)
( ) Student
( ) Unemployed
( ) Decline to answer
( ) Other (Please specify): _________________*

6) Which of the following best describes your highest achieved education level?*
( ) No high school ( ) Some high school ( ) High school graduate ( ) Some college - no degree ( ) Associates/2 year degree ( ) Bachelors/4 year degree
( ) Graduate degree - Masters, PhD, professional, medicine, etc.

7) Do you have a college degree or work experience in computer science, software development, web development or similar computer-related fields?*
( ) Yes ( ) No

8) Using only desktop or laptop computers, either at home or at work, approximately how many hours do you spend on the Internet each day?*
( ) None ( ) Fewer than 1 ( ) Between 1 and 5 ( ) Between 5 and 9 ( ) Between 9 and 13 ( ) Between 13 and 17 ( ) More than 17

9) Using only mobile devices (e.g., Android Smartphone, iPhone, iPad, tablet, or similar), approximately how much time do you spend on the Internet each day?*
( ) None
( ) Fewer than 1
( ) Between 1 and 5
( ) Between 5 and 9
( ) Between 9 and 13
( ) Between 13 and 17
( ) More than 17

10) Approximately how often do you use Facebook?*
( ) Never
( ) A few times per month or less
( ) Once per week
( ) Several times per week
( ) Once per day
( ) Several times per day

11) Have you ever...? (Select all that apply)*
[ ] ...purchased a product or service online (e.g., music, books, clothing, etc.)
[ ] ...used a social networking site (e.g., Facebook, Twitter, LinkedIn, MySpace, etc.)
[ ] ...clicked on an ad that appeared on a website to get more information about the advertised product
[ ] ...accidentally clicked on an ad that appeared on a website
[ ] ...visited health, wellness, or medical information websites (e.g., MayoClinic, MyFitnessPal, Men's Health, etc.)
[ ] ...used a search engine to find information about a medical condition

[ ] None of the above

Visiting a healthcare website
[WebMD/WebDR] is a healthcare information website. It provides information about the symptoms, treatment, and prevention of a range of health conditions.

Clicking on the link below will open a new tab or window in your browser displaying a version of the [WebMD/WebDR] website homepage with links disabled. Please look through this page at your own pace and make sure to scroll down and look at the entire page. Then, answer the following questions. Feel free to review the opened tab as many times as you want to answer these questions.
Click here to visit the [WebMD/WebDR] homepage

12) Please select from the list below at least three of the health conditions that appear on the left-hand side of the [WebMD/WebDR] homepage.*
[ ] Acne
[ ] Allergies
[ ] Alzheimer
[ ] Asthma
[ ] Bipolar disorder
[ ] Cancer
[ ] Carpal tunnel
[ ] Conjunctivitis
[ ] Depression
[ ] Glaucoma
[ ] Herpes
[ ] Hyperactivity
[ ] Hypertension
[ ] Osteoporosis

13) Indicate how much you agree or disagree with the following statements.*

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
I have a positive impression of the [WebMD/WebDR] website	( )	( )	( )	( )	( )
I believe [WebMD/WebDR] is a trustworthy website	( )	( )	( )	( )	( )
I believe the [WebMD/WebDR]website protects my privacy	( )	( )	( )	( )	( )
I am familiar with the [WebMD/WebDR]website	( )	( )	( )	( )	( )
I believe [WebMD/WebDR]is a well-known website	( )	( )	( )	( )	( )
I believe the [WebMD/WebDR]website has a good reputation	( )	( )	( )	( )	( )
I believe the [WebMD/WebDR]website provides useful information	( )	( )	( )	( )	( )

14) Had you ever visited the [WebMD/WebDR] website before (other than in this study)?*
( ) Yes ( ) No ( ) I don't remember

15) How often have you visited the [WebMD/WebDR] website in the last 12 months?*
( ) None ( ) Only once ( ) A few times ( ) A few times per month ( ) A few times per week ( ) A few times per day

16) Do you have a user account on the [WebMD/WebDR] website?*
( ) Yes ( ) No ( ) I don't remember

17) Have you visited other health or medical-information websites in the past?*
( ) Yes ( ) No ( ) I don't remember

Please read this information carefully. Then answer the questions below.
Many websites, including [WebMD/WebDR], are able to offer free services to their visitors by contracting with online advertising companies. The advertising companies pay websites for every ad they show, allowing the websites to provide free services for users like you.
Imagine that you are experiencing a flaky scalp condition and decide to visit the [WebMD/WebDR] website. [WebMD/WebDR] has contracted with ***[XYZ Advertising Company/Facebook*], which collects information about your interactions with the [WebMD/WebDR] website in order t **predict your preferences** and to show you ads that are most likely to be of interest to you. These ads are known as **targeted ads**. For example, if you search for "flaky scalp" or read an article about scalp problems on the [WebMD/WebDR] website, ***[****XYZ Advertising Company/Facebook]* could show you ads for dandruff shampoo or another related product.

In particular, *[**XYZ Advertising Company/Facebook***] will:

1. Collect your information **only from the** [WebMD/WebDR] website.
2. Use the collected information to show you **targeted ads only on the** [WebMD/WebDR] website.
3. Retain and use collected information for a [**maximum period of one day/indefinite period time]**.

[No text/ In addition: *[**XYZ Advertising Company/Facebook***] will provide you access to a webpage where you can **review, edit, and delete** the information that is being collected about you. For example, you can confirm that your information and preferences are accurate and remove information that you no longer feel comfortable sharing.]

18) Based on the information that you just read, which of the following are examples of the types of targeted ads that might occur as a result of your visit to [WebMD/WebDR]? (Choose any that apply)*
[ ] You see ads for bicycles on [WebMD/WebDR] since studies have found that many visitors to [WebMD/WebDR] are bicycle enthusiasts
[ ] You see ads for Acme cough syrup on Facebook because you read about cough remedies on [WebMD/WebDR]
[ ] You see ads for Acme cough syrup on [WebMD/WebDR] because a friend emailed you information about cough remedies
[ ] You see ads for Acme cough syrup on [WebMD/WebDR] because you read about cough remedies on [WebMD/WebDR]

[ ] You see ads for Acme cough syrup on www.WashingtonPost.com because you read about cough remedies on [WebMD/WebDR]

19) Based on the information that you just read, which of the following statements best explains how [XYZ Advertising Company/Facebook] may use the information that it collects about you?*

( ) To show me non-targeted ads on the websites that I visit
( ) To show me targeted ads only on the [WebMD/WebDR] website
( ) To show me targeted ads on the [WebMD/WebDR] website and other websites that I visit
( ) To show me targeted ads only on Facebook
( ) To show me targeted ads on Facebook and on the [WebMD/WebDR] website
( ) Other [Please explain]: _________________*

20) Based on the information that you just read, for how long may [XYZ Advertising Company/Facebook] use the information collected about you?*

( ) One day ( ) One week ( ) One year ( ) Indefinitely

Suppose that you use only your home computer to access the [WebMD/WebDR] website, and that nobody else uses this computer. Based only on the information that your read above, please answer the questions below indicating what information you would allow [XYZ Advertising Company/Facebook] to collect for the purpose of showing you targeted ads [on your Facebook page and the (WebMD/WebDR) website/only on the (WebMD/WebDR) website/on the (WebMD/WebDR) website and other websites you visit]

21) I would be willing to allow [XYZ Advertising Company/Facebook] to use and store the following information about my computer. This information will be retained [indefinitely/one day] [nothing/and you will be able to review, edit, and delete it]*

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
The type of operating system (e.g., Windows, Mac, etc.) of my computer	( )	( )	( )	( )	( )
The IP address of my computer (i.e., a computer identifier assigned by your Internet service provider)	( )	( )	( )	( )	( )
The name and version of the web browser (e.g., Internet Explorer 9, Firefox 18.0.1, Safari 6.0.2, etc.) that I use to visit the [WebMD/WebDR] website	( )	( )	( )	( )	( )

22) I would be willing to allow [XYZ Advertising Company/Facebook] to use and store the following demographic and preference information. This information will be retained [indefinitely/one day[nothing/and you will be able to review, edit, and delete it]*

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
My age	( )	( )	( )	( )	( )
My gender	( )	( )	( )	( )	( )
My highest level of education	( )	( )	( )	( )	( )
My income bracket	( )	( )	( )	( )	( )
My religion	( )	( )	( )	( )	( )
My political preferences	( )	( )	( )	( )	( )
My sexual orientation	( )	( )	( )	( )	( )
My marital status	( )	( )	( )	( )	( )
My hobbies	( )	( )	( )	( )	( )
My credit score bracket	( )	( )	( )	( )	( )

23) I would be willing to allow [XYZ Advertising Company/Facebook] to use and store the following information related to my interactions with the [WebMD/WebDR] website. This information will be retained [indefinitely/one day[nothing/and you will be able to review, edit, and delete it]*

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
The pages I've visited on the [WebMD/WebDR] website	( )	( )	( )	( )	( )
Which search terms I've entered on the [WebMD/WebDR] website	( )	( )	( )	( )	( )
My weight and height	( )	( )	( )	( )	( )
My responses to health-related surveys	( )	( )	( )	( )	( )
The medications I am taking (inferred from my interactions with the site)	( )	( )	( )	( )	( )
How long I spent on each page of the [WebMD/WebDR] website	( )	( )	( )	( )	( )

24) I would be willing to allow [XYZ Advertising Company/Facebook] to use and store the following information related to my location. This information will be retained [indefinitely/one day[nothing/and you will be able to review, edit, and delete it]*

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
The country from which I'm visiting the [WebMD/WebDR] website	( )	( )	( )	( )	( )
The state from which I'm visiting the [WebMD/WebDR] website	( )	( )	( )	( )	( )
The town or city from which I'm visiting the [WebMD/WebDR] website	( )	( )	( )	( )	( )
The zip code from which I'm visiting the [WebMD/WebDR] website	( )	( )	( )	( )	( )
The exact address from which I'm visiting the [WebMD/WebDR] website	( )	( )	( )	( )	( )

25) I would be willing to allow [XYZ Advertising Company/Facebook] to collect the following information. This information will be retained [indefinitely/one day[nothing/and you will be able to review, edit, and delete it]*

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
My name	( )	( )	( )	( )	( )
My email address	( )	( )	( )	( )	( )
My phone number	( )	( )	( )	( )	( )
My address	( )	( )	( )	( )	( )
My social security number	( )	( )	( )	( )	( )
My credit card number	( )	( )	( )	( )	( )

26) How would your willingness to allow [XYZ Advertising Company/Facebook] to collect your information change if it retained your information...*

	I would be less willing	**I would be equally willing**	**I would be more willing**
...only for the duration of a single web browsing session	( )	( )	( )
...for one week	( )	( )	( )
...for one month	( )	( )	( )
... for six months	( )	( )	( )
...for one year	( )	( )	( )
...indefinitely	( )	( )	( )

27) How would your willingness to allow [XYZ Advertising Company/Facebook] to collect your information change if it retained your information...*

	I would be less willing	**I would be equally willing**	**I would be more willing**
...only for the duration of a single web browsing session	( )	( )	( )
...for one day	( )	( )	( )
...for one week	( )	( )	( )
...for one month	( )	( )	( )
... for six months	( )	( )	( )
...for one year	( )	( )	( )

28) How would your willingness to allow [XYZ Advertising Company/Facebook] to collect your information change if it provided you access to a webpage where you could review, edit, and delete the information that is being collected about you? For example, you could confirm that your information and preferences are accurate and remove information that you no longer feel comfortable sharing.*

( ) I would be less willing
( ) I would be equally willing
( ) I would be more willing

29) Imagine that you are a frequent user of the [WebMD/WebDR] website, and that [WebMD/WebDR] offers you the opportunity to pay a monthly fee in exchange for not showing you any ads on the [WebMD/WebDR] website. In this case the information that [XYZ Advertising Company/Facebook] collects from you will not be used to show you ads, but may still be used for other purposes. What monthly fee, if any, in dollars and cents might you be willing to pay?*

30) Please explain how you chose the amount in the previous question.*

31) Imagine that you are a frequent user of the [WebMD/WebDR] website, and that [WebMD/WebDR] offers you the opportunity to pay a monthly fee in exchange for not showing you targeted ads but only generic ads on the [WebMD/WebDR] website. In this case the information that [XYZ Advertising Company/Facebook] collects from you will not be used to show you targeted ads, but may still be used for other purposes. What monthly fee (if an) in dollars and cents might you be willing to pay?*

32) Please explain how you chose the amount in the previous question.*

33) Imagine that you are a frequent user of the [WebMD/WebDR] website, and that [WebMD/WebDR] offers you the opportunity to pay a monthly fee in exchange for stopping [XYZ Advertising Company/Facebook] from collecting any information about you or your online activities on the [WebMD/WebDR] website. What monthly fee (if any) in dollars and cents might you be willing to pay?*

34) Please explain how you chose the amount in the previous question.*

35) How much do you agree or disagree with the following statements. I am interested in receiving targeted ads on the websites that I visit based on my online activities on...*

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**	**I don't use them**
...health websites	( )	( )	( )	( )	( )	( )
...online banking websites	( )	( )	( )	( )	( )	( )
...travel websites	( )	( )	( )	( )	( )	( )
...employment websites	( )	( )	( )	( )	( )	( )
...arts and entertainment websites	( )	( )	( )	( )	( )	( )
...dating websites	( )	( )	( )	( )	( )	( )
...news websites	( )	( )	( )	( )	( )	( )
...photo sharing websites	( )	( )	( )	( )	( )	( )
...social networking websites	( )	( )	( )	( )	( )	( )

36) What do you consider the main benefit, if any, of receiving ads that are targeted based on your online activities?*

37) What do you consider the main downside, if any, of receiving ads that are targeted based on your online activities?*

38) Overall, how do you feel about receiving ads that are targeted based on your online activities?*

( ) Strongly dislike
( ) Dislike
( ) Neutral
( ) Like
( ) Strongly like

39) Explain what, if anything, would make you feel more comfortable with receiving targeted ads?

40) How would you feel about seeing ads on Facebook that are targeted based on your activities on other websites that you visit? Please explain.*

41) How much do you agree or disagree with the following statements:*

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
It would be useful to see ads on my Facebook page based on my interactions with the [WebMD/WebDR] website	( )	( )	( )	( )	( )
I would feel comfortable seeing ads on my Facebook page based on my interactions with the [WebMD/WebDR] website	( )	( )	( )	( )	( )
It would be useful to see ads on my Facebook page based on my activities on the [WebMD/WebDR] website and other websites I visit	( )	( )	( )	( )	( )
I would feel comfortable if Facebook shows me ads on my Facebook page based on my activities on the [WebMD/WebDR] website and other websites I visit	( )	( )	( )	( )	( )
It would be useful to see ads on the websites that I visit based on my activities on my Facebook page and other websites that I've visited in the past	( )	( )	( )	( )	( )
I would feel comfortable seeing ads on the websites that I visit based on my activities on Facebook and other websites that I've visited in the past	( )	( )	( )	( )	( )

42) Please state how much you agree or disagree with the following statements.
I would be more willing to allow collection of ANONYMOUS information (i.e., information that cannot be used to identify me or contact me) for the purpose of receiving targeted ads if my web browser...*

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
...allowed me to choose ahead of time what information advertising companies can learn about me	( )	( )	( )	( )	( )
...allowed me to control which advertising companies can collect and use that information	( )	( )	( )	( )	( )
...allowed me to visualize what the advertising companies already know about me	( )	( )	( )	( )	( )
...allowed me to create different "personas" (i.e., fake or real characterizations of me) to show to these advertising companies at different points in time	( )	( )	( )	( )	( )
...allowed me to control on which websites my information can be collected	( )	( )	( )	( )	( )
...showed me on which websites my information has been collected	( )	( )	( )	( )	( )

43) Please state how much you agree or disagree with the following statements.
I would be more willing to allow collection of PERSONAL information (i.e. information that can be used to identify me and contact me) for the purpose of receiving targeted ads if my web browser....*

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
...allowed me to choose ahead of time what information advertising companies can learn about me	( )	( )	( )	( )	( )
...allowed me to control which advertising companies can collect and use that information	( )	( )	( )	( )	( )
...allowed me to visualize what the advertising companies already know about me	( )	( )	( )	( )	( )
...allowed me to control on which websites my information can be collected	( )	( )	( )	( )	( )
...showed me on which websites my information has been collected	( )	( )	( )	( )	( )
...allowed me to create different "personas" (i.e., fake or real characterizations of me) to show to these advertising companies at different points in time	( )	( )	( )	( )	( )

44) Please tell us what functionality would you like to have in your web browser to control the information that online advertising companies collect about you for the purpose of showing you targeted ads.

This is the last page of the survey. Please answer these last questions as accurately as possible.

45) Please indicate whether you have ever done any of the following.*

	Yes	**No**
Refused to give information to a website because you felt it was too personal or unnecessary	( )	( )
Decided not to use a website or not to purchase something online because you were not sure how your personal information would be used	( )	( )
Read a website's privacy policy	( )	( )
Deleted cookies from your web browser	( )	( )
Turned on the "do not track" option in your web browser	( )	( )

46) How much do you agree or disagree with the following statements:*

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
When websites ask for personal information, I usually think twice about providing it	( )	( )	( )	( )	( )
Consumers have lost all control over how personal information is collected and used by companies	( )	( )	( )	( )	( )
I feel that as a result of my visiting websites, others know more about me than I am comfortable with	( )	( )	( )	( )	( )

47) Do you have any further comments?

F.4 Study Two: Regression Model

Independent Variable	Control Category	Coefficient	Std. Error	Z	P>|Z|
Dependent Variable: Articles Read, Videos Watched, and Pages Visited (Online Interactions)					
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	-0.16	0.2	-0.8	0.42
Scope 3: AllNews + Others	Scope 1: Only AllNews	-0.18	0.18	-0.97	0.33
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	-0.58	0.19	-3.06	0.002
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	-0.62	0.21	-2.87	0.004
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	-0.55	0.20	-2.80	0.005
Scope 7: AllNews + FB	Scope 1: Only AllNews	-0.85	0.19	-4.5	<0.001
Purpose: Positive	Purpose: Negative	0.32	0.17	1.84	0.07
Purpose: Ambiguous	Purpose: Negative	0.17	0.24	0.71	0.48
Privacy Concerned: Yes	No	-0.62	0.14	-4.43	<0.001
Targeted Ads Opinion: Positive	Negative	1.54	0.15	10.12	<0.001
Opinion on AllNews	Negative	0.91	0.10	8.85	<0.001
Has FB Account: Yes	No	0.33	0.13	2.61	0.009
Dependent Variable: Purchasing Interests					
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	0.07	0.20	0.36	0.72
Scope 3: AllNews + Others	-0.014	0.19	-0.08	0.94	
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	-0.06	0.19	-0.29	0.77
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	-0.33	0.22	-1.51	0.13
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	-0.44	0.20	-2.21	0.03
Scope 7: AllNews + FB	Scope 1: Only AllNews	-0.35	0.19	-1.83	0.07
Purpose: Positive	Purpose: Negative	0.22	0.18	1.22	0.22
Purpose: Ambiguous	Purpose: Negative	0.37	0.24	1.55	0.13
Privacy Concerned: Yes	No	-0.43	0.14	-3.06	0.002
Targeted Ads Opinion: Positive	Negative	1.87	0.17	11.28	<0.001
AllNews Opinion: Positive	Negative	0.71	0.10	6.79	<0.001
Dependent Variable: Gender					
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	0.21	0.20	1.05	0.29
Scope 3: AllNews + Others	-0.009	0.19	-0.05	0.96	
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	0.24	0.19	1.28	0.20
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	-0.05	0.21	-0.25	0.80
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	-0.34	0.20	-1.70	0.09
Scope 7: AllNews + FB	Scope 1: Only AllNews	0.33	0.18	1.78	0.08
Purpose: Positive	Purpose: Negative	0.32	0.17	1.82	0.07
Purpose: Ambiguous	Purpose: Negative	0.32	0.24	1.32	0.19
Privacy Concerned: Ye	No	-0.42	0.14	-3.11	0.002
Targeted Ads Opinion: Positive	Negative	1.1	0.14	7.93	<0.001
AllNews Opinion: Positive	Negative	0.72	0.10	6.99	<0.001
Has FB Account: Yes	No	0.57	0.13	4.33	<0.001
Age	NA	-0.01	0.004	-2.82	0.004
Gender: Male	Female	0.29	0.10	2.79	0.005
Dependent Variable: ZIP code					

Continued on next page . . .

Continued from previous page

Independent Variable	Control Category	Coefficient	Std. Error	Z	P>|Z|
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	0.06	0.22	0.29	0.77
Scope 3: AllNews + Others	-0.05	0.21	-0.24	0.81	
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	-0.25	0.21	-1.20	0.23
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	-0.40	0.24	-1.64	0.10
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	-0.64	0.23	-2.82	0.005
Scope 7: AllNews + FB	Scope 1: Only AllNews	-0.13	0.21	-0.63	0.53
Purpose: Positive	Purpose: Negative	0.46	0.20	2.35	0.02
Purpose: Ambiguous	Purpose: Negative	0.39	0.27	1.47	0.14
Privacy Concerned: Yes	No	-0.74	0.14	-5.25	<0.001
Targeted Ads Opinion: Positive	Negative	1.17	0.14	8.18	<0.001
AllNews Opinion: Positive	Negative	0.60	0.12	5.1	<0.001
Age	NA	0.03	0.004	5.52	<0.001
Gender: Male	Female	0.26	0.11	2.28	0.02
Dependent Variable: Sexual Orientation					
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	0.30	0.27	1.08	0.28
Scope 3: AllNews + Others	Scope 1: Only AllNews	0.25	0.25	1.01	0.31
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	0.37	0.25	1.46	0.14
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	0.56	0.28	2.01	0.04
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	-0.07	0.28	-0.25	0.80
Scope 7: AllNews + FB	Scope 1: Only AllNews	0.80	0.24	3.35	<0.001
Purpose: Positive	Purpose: Negative	0.21	0.22	0.96	0.34
Purpose: Ambiguous	Purpose: Negative	0.39	0.30	1.30	0.19
Privacy Concerned: Yes	No	-0.79	0.15	-5.29	<0.001
Targeted Ads Opinion: Positive	Negative	1.0	0.15	6.82	<0.001
Has FB Account: Yes	No	0.29	0.17	1.71	0.09
AllNews Opinion: Positive	Negative	0.58	0.13	4.36	<0.001
Dependent Variable: Email					
Scope 2: Only AllNews+Other Purposes	Scope 1: Only AllNews	0.37	0.34	1.11	0.27
Scope 3: AllNews + Others	-0.10	0.34	-0.30	0.77	
Scope 4: AllNews+Others+Offline	Scope 1: Only AllNews	0.56	0.30	1.84	0.07
Scope 5: AllNews+Others+Other Purposes (No sharing)	Scope 1: Only AllNews	-0.19	0.40	-0.47	0.64
Scope 6: AllNews+Others+Other Purposes	Scope 1: Only AllNews	0.22	0.34	0.65	0.51
Scope 7: AllNews + FB	Scope 1: Only AllNews	0.88	0.29	3.04	0.002
Purpose: Positive	Purpose: Negative	0.10	0.28	0.34	0.74
Purpose: Ambiguous	Purpose: Negative	-0.49	0.50	-0.98	0.33
Privacy Concerned: Yes	No	-0.58	0.19	-3.05	0.002
Targeted Ads Opinion: Positive	Negative	0.98	0.18	5.45	<0.001
Has FB Account: Yes	No	1.02	0.28	3.64	<0.001
AllNews Opinion: Positive	Negative	0.62	0.18	3.53	<0.001

Table F.3: The logistic regression models of participants' willingness to disclose information. In addition to the scenario treatment, we included the following co-variates: age, gender, whether or not a participant used Facebook, Internet literacy, privacy concerns, whether participants like targeted ads, opinion of the AllNews website, and whether or not a participant answered at least one of the scenario understanding questions correctly. Only variables significant at α<0.05 are shown. If one or more levels of a categorical variable was significant, we show all the levels of that categorical variable.

F.5 Study Two: Survey Questions

( ) Fox News	( ) Huffington Post
( ) Los Angeles Times	( ) MSN
( ) NBC News	( ) reddit
( ) Reuters	( ) The Guardian
( ) The New York Post	( ) The New York Times
( ) The Wall Street Journal	( ) The Washington Post
( ) USA Today	( ) Yahoo!
( ) I never visit news websites	( ) Other [Please specify]: ____*

13) How often have you visited **[News Site Name]** in the last 12 months?*
( ) Only once ( ) A few times ( ) A few times per month ( ) A few times per week ( ) A few times per day

14) Do you have a user account on the **[News Site Name]** website?*
( ) Yes
( ) No
( ) I don't remember

15) Indicate how much you agree or disagree with the following statements.*

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I believe **[News Site Name]** has a good reputation	( )	( )	( )	( )	( )
I have a positive impression of **[News Site Name]**	( )	( )	( )	( )	( )
I believe **[News Site Name]** provides useful information	( )	( )	( )	( )	( )
I believe **[News Site Name]** protects my privacy	( )	( )	( )	( )	( )

Visiting a news website

AllNews is a news website. On allnews.com you can read articles and watch videos on breaking news, events, opinions, and interviews. allnews.com allows you to search for articles and videos.

Clicking on the link below will open a new tab or window in your browser displaying a version of the AllNews website homepage with links disabled. Please look through this page at your own pace and make sure to scroll down and look at the entire page. Then, answer the following questions. Feel free to review the opened tab as many times as you want to answer these questions.
Click here to visit the AllNews homepage

16) Please select from the list below one article that appears on the left-hand side of AllNews homepage.*
[ ] Bad news for Obamacare success story
[ ] Zimmerman agrees to go weaponless
[ ] Washington politics holding back growth
[ ] Ex-NFL player dies in high-speed crash
[ ] Rare, good news about U.S. deficit

17) Indicate how much you agree or disagree with the following statements.*

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I believe the AllNews website has a good reputation	( )	( )	( )	( )	( )
I have a positive impression of the AllNews website	( )	( )	( )	( )	( )
I believe the AllNews website provides useful information	( )	( )	( )	( )	( )
I believe the AllNews website protects my privacy	( )	( )	( )	( )	( )

Logic: Hidden unless: Scope is equal to 6. **Note: We only show scope 6 as example, other scopes have different practices.**

Please read this information carefully. Then answer the questions below.

Many websites, including AllNews, are able to offer free services to their visitors by contracting with online advertising companies. The advertising companies pay websites for every ad they show, allowing the websites to provide free services to users.

Imagine that you provided some information about yourself (e.g., email address, gender, etc.) when you signed up for an account with the AllNews website. Further imagine that AllNews has contracted with Best Ads, an advertising company that is interested in learning about you to show you ads that are most likely to be of interest to you. These ads are known as targeted ads.

For example, if you watch a video about the 2014 winter Olympic games on the AllNews website and then visit a traveling website and look up hotels near the Olympic venue, next time you visit the AllNews or any other news, entertainment, travel, or retail website, Best Ads could show you a targeted ad for a discounted hotel near the Olympic venue.

The following table summarizes Best Ads' data collection and use practices.

Best Ads may collect information from	• The AllNews website • Other news, entertainment, travel, and retail websites you visit
Best Ads may use information it collects to show you targeted ads on	• The AllNews website • Other news, entertainment, travel, and retail websites you visit
Best Ads may use information it collects for	• Targeted ads • Other purposes
Best Ads may retain information for	• **[One week / 3 months / One Year]**

18) The information you just read states that **[Best Ads / Facebook]** may use the information it collects about you also for purposes other than targeted ads. What do you think these other purposes might be?*

19) Based only on the information that you just read, for how long may **[Best Ads / Facebook]** retain the information it collects about you?*
( ) One week
( ) One month
( ) Three months
( ) Six months
( ) One year
( ) Indefinitely

20) Based only on the information that you read in the description above, which of the following are examples of the types of targeted ads that might occur as a result of your visits to **[AllNews, AllNews and Other/ AllNews and Facebook]**? (Choose all that apply)*
[ ] You see ads for Olympics t-shirts on Facebook because you read about the Olympics on AllNews
[ ] You see ads for Olympics t-shirts on AllNews because you read about the Olympics on AllNews
[ ] You see ads for Olympics t-shirts on hoteldeals.com because you read about the Olympics on AllNews
[ ] You get a coupon at your local department store for half-price Olympics t-shirts because you read about the Olympics on AllNews
[ ] You see ads for hotels on AllNews because you visited a travel website

Logic: Hidden unless: Scope is equal to 6 and understanding question was incorrectly answered

Your responses to the previous question are not completely right. We are going to let you try again in the next page, but we need you to fully understand Best Ads' practices first. In particular, you missed at least one of the following three true statement(s). Please review them and make sure you understand them before continuing with the survey.

Statement	True/False?	Explanation
You see ads for Olympics t-shirts on AllNews because you read about the Olympics on AllNews	TRUE	Best Ads may show you targeted ads on the AllNews website based on what you do on the AllNews website
You see ads for Olympics t-shirts on hoteldeals.com because you read about the Olympics on AllNews	TRUE	Best Ads may show you targeted ads on AllNews and other entertainment, travel, and retail websites
You see ads for hotels on AllNews because you visited a travel website	TRUE	Best Ads may collect information from other websites you visit to show you targeted ads on AllNews

Logic: Hidden unless: Scope is equal to 6 and understanding question was incorrectly answered

Please review again **[Best Ads / Facebook]**'s practices. Then answer the question below.

Best Ads may collect information from	• The AllNews website • Other news, entertainment, travel, and retail websites you visit
Best Ads may use information it collects to show you targeted ads on	• The AllNews website • Other news, entertainment, travel, and retail websites you visit
Best Ads may use information it collects for	• Targeted ads • Other purposes
Best Ads may retain information for	• **[One week / 3 months / One Year]**

27) I would be comfortable if **[Best Ads / Facebook]** collected or otherwise inferred the following information about me:*

Strongly Disagree Neutral Agree Strongly

	disagree				agree
The type of operating system (e.g., Windows, Mac, etc.) of my computer	( )	( )	( )	( )	( )
The IP address of my computer (i.e., a computer identifier assigned by your Internet service provider)	( )	( )	( )	( )	( )
The articles I read, videos I watch, and pages I visit on the AllNews website	( )	( )	( )	( )	( )
My income bracket	( )	( )	( )	( )	( )
My gender	( )	( )	( )	( )	( )
The ZIP code from which I visit the [AllNews, AllNews and Other/ AllNews and Facebook] website	( )	( )	( )	( )	( )
My email address	( )	( )	( )	( )	( )
My credit score bracket	( )	( )	( )	( )	( )
My sexual orientation	( )	( )	( )	( )	( )
The products I may be interested in purchasing	( )	( )	( )	( )	( )

Logic: Hidden unless: Question (ID 27.1 – 27.10) contains any ("Agree","Strongly agree")

28) Please explain why you would be COMFORTABLE with **[Best Ads / Facebook]** knowing your **[Data Type]**?*

Logic: Hidden unless: Question (ID 27.1-27.10) contains any ("Strongly disagree","Disagree")

29) Please explain why you would NOT be COMFORTABLE with **[Best Ads / Facebook]** knowing your **[Data Type]**?*

Advertising companies create individual profiles based on the information they collect or infer from users' online activities. Some of these companies provide Internet users access to their profiles. The table below shows an example of what information such a user profile may include. The information has been taken from actual user profiles created by an advertising company. Please review this sample profile carefully and then answer the questions below.

Data Type	Value
Location	Region:[Participant's Region]
	City:[Participant's City]
	IP Address:[Participant's Computer IP Address]
Demographics-Individual	Gender:[Participant's Gender]
	Single
	[Participant's Age] years old
	Education: [Participant's Education]
	Type of Job:[Participant's Occupation]
Demographics-Household	Income: $50K - $75K
	Household size:1
	Number of Adults:1
	Children in Residence: No
	Home Type: Multifamily Dwelling
	Home Value: Less than $100K
	Length of Residence: Fewer than 3 years
	Discretionary spending: $30K-$40K
	Voter Indicator: Republican
	Automobile: Less than $20K
Interest	General Health>Bones, Joints, Muscles>Pain
	Religion code: Tiers 1 - 3
	Video Games: Computer, PlayStation 3
	Travel Destinations>North America>US>New York>NYC
	Miscellaneous>News>Business and Finance
	Automobile:Coupes
	Online Activities: Research
Activities	Past Purchase>Products>Clothing>Jeans
	Offline Purchases>P&G>Charmin Ultra Soft
	Student Loan Consolidation
	Volunteering: Tier 1 - 3
Attitudes	Buy American: Not likely
	Look at Me Now: Most likely
	Never Show Up Empty Handed: Most likely
	It's all in the Name: Most likely
Behavior	Green Living
	Eco Friendly Vehicle Owner
	Mass Market and Discount Shopper
	Gift buyer
	Prepaid wireless plan subscriber
	Premium channel viewer
Predictive	Credit card interest score: 16-17%
	Credit card appl. intent score: 10 -11%
	Auto insurance online buyer: High propensity
	Online Higher Education Enrollee: High propensity
	In-market: Cell phones and plans

48) Please select from the list below two items that appears in the sample profile.*

[ ] Married
[ ] Credit card interest score 16-17%
[ ] Income: $75K - $100K
[ ] In-market: Jewelry
[ ] Children in Residence: No

49) What do you see as benefits (if any) of users from having access to the profiles that advertising companies create about them?*

50) Think about the information that is shown in the sample profile. How much do you agree or disagree with the following statements.*

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I am comfortable with the information that such profiles may contain	( )	( )	( )	( )	( )
I am concerned about the information that such profiles may contain	( )	( )	( )	( )	( )
I am surprised about the information that such profiles may contain	( )	( )	( )	( )	( )

Logic: Hidden unless: Question (ID 50.3) contains any ("Agree","Strongly agree")

51) Please explain what exactly surprised you about the sample profile?*

Logic: Hidden unless: Question (ID 50.2) contains any ("Agree","Strongly agree")

52) Please explain what exactly is concerning to you about the sample profile?*

53) Indicate how much you agree or disagree with the following statements. In general, I am comfortable sharing the following information with advertising companies:*

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
My online activities	( )	( )	( )	( )	( )
My demographic information	( )	( )	( )	( )	( )
My interests	( )	( )	( )	( )	( )
My contact information	( )	( )	( )	( )	( )

54) Think about the ability to view and edit the information that advertising companies know about you. How much do you agree or disagree with the following statements.*

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
hould be given the opportunity to view and edit the profiles that advertising companies eate about me	( )	( )	( )	( )	( )
aving access to the profiles that advertising companies create about me is beneficial to me	( )	( )	( )	( )	( )
ing able to edit the profiles that advertising companies create about me allows those mpanies to serve me better	( )	( )	( )	( )	( )
ing able to edit the profiles that advertising companies create about me provides those mpanies with more accurate information about me	( )	( )	( )	( )	( )
ing able to edit the profiles that advertising companies create about me allows me to decide at advertising companies can know about me	( )	( )	( )	( )	( )
ing able to edit the profiles that advertising companies create about me is beneficial to me	( )	( )	( )	( )	( )

Logic: Hidden unless: Question (ID 54.1) contains any ("Agree","Strongly agree")

55) You indicated that you would like to be given the opportunity to view and edit the profiles that advertising companies create about you. Please explain why you think having access to your profile is important.*

56) In general, how do you feel about receiving ads that are targeted based on your online activities?*
( ) Strongly dislike ( ) Dislike ( ) Neutral ( ) Like ( ) Strongly like

This is the last page of the survey. Please answer these last questions as accurately as possible.

57) How much do you agree or disagree with the following statements:*

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
usually bothers me when online companies ask me for personal information	( )	( )	( )	( )	( )
m concerned that online companies are collecting too much personal information about me	( )	( )	( )	( )	( )
bothers me to give personal information to so many online companies	( )	( )	( )	( )	( )
hen online companies ask for personal information, I usually think twice before providing it	( )	( )	( )	( )	( )
eel that as a result of me visiting online companies, others know more about me than I am mfortable with	( )	( )	( )	( )	( )
nsumer online privacy is really a matter of consumer' right to exercise control and tonomy over decisions about how their information is collected, used, and shared	( )	( )	( )	( )	( )
nsumer control of personal information lies at the heart of consumer privacy	( )	( )	( )	( )	( )
elieve that online privacy is invaded when control is lost or unwillingly reduced as a result a marketing transaction	( )	( )	( )	( )	( )

58) Do you have any further comments?

Thank you for taking the survey. Below is your confirmation code. You must retain this code to be paid - it is recommended that you store your code in a safe place (either by writing it down, or by printing this page).

REMINDER: You must correctly copy and paste the confirmation code into Mechanical Turk to be paid!

YOUR CODE IS
{Code Inserted Here}

Bibliography

[1] The White House, "Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Digital Economy," February 2012.

[2] P. P. Swire and K. Ahmad, *U.S. Private-sector Privacy. Law and Practice for Information Privacy Professionals*. IAPP, 2012.

[3] Federal Trade Commission, "Online behavioral advertising moving the discussion forward to possible self-regulatory principles," http://www.ftc.gov/os/2007/12/P859900stmt.pdf, December 2007, retrieved April 2013.

[4] ——, "FTC staff report: Self-regulatory principles for online behavioral advertising," February 2009.

[5] ——, "Protecting consumer privacy in an era of rapid change," March 2012.

[6] U.S. Securities and Exchange Commission, "The Laws That Govern the Securities Industry." [Online]. Available: http://www.sec.gov/about/laws.shtml

[7] U.S. Food and Drug Administration. Guide to Nutrition Labeling and Education Act (NLEA) Requirements. [Online]. Available: http://www.fda.gov/iceci/inspections/inspectionguides/ucm074948.htm

[8] National Conference of State Legislatures, "Security Breach Notification Laws," September 2014. [Online]. Available: http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx

[9] A. Fung, M. Graham, and D. Weil, *Full disclosure: The perils and promise of transparency*. Cambridge University Press, 2007.

[10] F. Cate, "The failure of fair information practice principles," 2008.

[11] M. A. Edwards, "Empirical and behavioral critiques of mandatory disclosure: Socio-economics and the quest for truth in lending," *Cornell JL & Pub. Pol'y*, vol. 14, p. 199, 2004.

[12] A. F. Westin, "Privacy and freedom," *Washington and Lee Law Review*, vol. 25, no. 1, p. 166, 1968.

[13] F. H. Cate and R. Litan, “Constitutional issues in information privacy,” *Mich. Telecomm. & Tech. L. Rev.*, vol. 9, p. 35, 2002.

[14] H. F. Nissenbaum, *Privacy in context: Technology, policy, and the integrity of social life*. Stanford Law & Politics, 2010.

[15] A. D. Thierer, “The Pursuit of Privacy in a World Where Information Control is Failing,” *Harvard Journal of Law and Public Policy*, vol. 36, no. 2, 2013.

[16] L. F. Cranor, “Necessary but not sufficient: Standardized mechanisms for privacy notice and choice,” *Journal of Telecommunications and High Technology Law*, vol. 10, no. 2, 2012.

[17] S. Komanduri *et al.*, “AdChoices? Compliance with online behavioral advertising notice and choice requirements,” *ISJLP*, vol. 7, pp. 603–721, 2012.

[18] J. R. Reidenberg, “Privacy wrongs in search of remedies,” *Hastings LJ*, vol. 54, p. 877, 2002.

[19] S. Gordon, “Privacy: A Study of Attitudes and Behaviors in US, UK and EU Information Security Professionals,” *Symantec White Paper*, 2004.

[20] A. McDonald *et al.*, “A comparative study of online privacy policies and formats,” in *Privacy enhancing technologies*. Springer, 2009, pp. 37–55.

[21] C. J. Hoofnagle and J. King, “What Californians understand about privacy online,” *Available at SSRN 1262130*, 2008.

[22] H. A. Simon, “Models of man: Social and rational; mathematical essays on rational human behavior in a social setting,” 1957.

[23] L. Willis, “Decisionmaking and the limits of disclosure: The problem of predatory lending: Price,” *Maryland Law Review*, vol. 65, p. 707, 2006.

[24] S. K. Ripken, “The dangers and drawbacks of the disclosure antidote: toward a more substantive approach to securities regulation,” *Baylor Law Review*, vol. 58, no. 1, pp. 139–204, 2006.

[25] A. Acquisti and J. Grossklags, “Privacy and rationality in individual decision making,” *IEEE Security & Privacy*, vol. 2, pp. 24–30, 2005.

[26] P. G. Leon, L. F. Cranor, A. M. McDonald, and R. McGuire, “Token attempt: The misrepresentation of website privacy policies through the misuse of p3p compact policy tokens,” in *Proceedings of the 9th Annual ACM Workshop on Privacy in the Electronic Society*, ser. WPES ’10. New York, NY, USA: ACM, 2010, pp. 93–104. [Online]. Available: http://doi.acm.org/10.1145/1866919.1866932

[27] W3C, “The platform for privacy preferences 1.0,” April 2002, http://www.w3.org/TR/REC-P3P-20020416. [Online]. Available: http://www.w3.org/TR/2002/REC-P3P-20020416/

[28] Microsoft, "Privacy in Internet Explorer 6," visited April 26, 2010. http://msdn.microsoft.com/en-us/library/ms537343

[29] http://www.w3.org/2002/12/18-p3p-workshop report.html, "Summary report - W3C workshop on the future of P3P," W3C, Tech. Rep., November 2002. [Online]. Available: http://www.w3.org/2002/12/18-p3p-workshop-report.html

[30] L. F. Cranor, *Web privacy with P3P*. O'Reilly, 2002.

[31] L. F. Cranor, P. Guduru, and M. Arjula, "User interfaces for privacy agents." *ACM Transactions on Computer-Human Interaction*, pp. 135–178, June 2006.

[32] L. F. Cranor, S. Egelman, S. Sheng, A. M. McDonald, and A. Chowdhury, "P3P deployment on websites," *Electronic Commerce Research and Applications*, vol. 7, no. 3, pp. 274–293, 2008.

[33] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder, "A "nutrition label" for privacy," *SOUPS '09: Proceedings of the 5th Symposium on Usable Privacy and Security*, 2009.

[34] S. Byers, L. F. Cranor, and D. Kormann, "Automated analysis of P3P-enabled web sites," *ICEC '03: Proceedings of the 5th international conference on Electronic commerce*, pp. 326–338, 2003.

[35] W3C, "The platform for privacy preferences 1.1," November 2006, http://www.w3.org/TR/P3P11/. [Online]. Available: http://www.w3.org/TR/P3P11/

[36] I. Reay, S. Dick, and J. Miller, "A large-scale empirical study of P3P privacy policies: Stated actions vs. legal obligations," *ACM Transactions on the Web*, 2009.

[37] J. R. Reidenberg and L. F. Cranor, "Can user agents accurately represent privacy policies?" *The 38th Research Conference on Communication, Information and Internet Policy*, 2002.

[38] J. A. Harvey and K. M. Sanzaro, "P3P and IE 6: good privacy medicine or mere placebo?" *The Computer and Internet Lawyer*, 2002.

[39] TRUSTe, "Directory of TRUSTe certified web sites," visited May 3, 2010. http://www.truste.com/trusted_sites/index.html.

[40] Microsoft Support, "Session variables are lost if you use FRAMESET in Internet Explorer 6," April 2006, http://support.microsoft.com/kb/323752.

[41] G. Patel, "How to set third-party cookies with iframe," December 2008, http://viralpatel.net/blogs/2008/12/how-to-set-third-party-cookies-with-iframe.html. [Online]. Available: http://viralpatel.net/blogs/2008/12/how-to-set-third-party-cookies-with-iframe.html

[42] A. Young, "IE blocking iframe cookies," September 2008, http://adamyoung.net/IE-Blocking-iFrame-Cookies. [Online]. Available: http://adamyoung.net/IE-Blocking-iFrame-Cookies

[43] K. Patil, "Session lost in iframe," April 2010, http://kiranpatils.wordpress.com/2010/04/13/session-lost-in-iframe/. [Online]. Available: http://kiranpatils.wordpress.com/2010/04/13/session-lost-in-iframe/

[44] S. Hacker, "P3P in IE: Frustrating failure," June 2002, http://www.oreillynet.com/mac/blog/2002/06/ p3p_in_ie6_frustrating_failure.html. [Online]. Available: http://www.oreillynet.com/mac/blog/2002/06/

[45] Piskvor, "Cookie blocked/not saved in IFRAME in IE," February 2009, http://stackoverflow.com/questions/389456/cookie-blocked-not-saved-in-iframe-in-internet-explorer. [Online]. Available: http://stackoverflow.com/questions/389456/cookie-blocked-not-saved-in-iframe-in-internet-explorer

[46] L. F. Cranor, K. Idouchi, P. G. Leon, M. Sleeper, and B. Ur, "Are they actually any different? Comparing thousands of financial institutions' privacy practices," in *Workshop on the Economics of Information Security (WEIS)*, June 2013.

[47] O. Ireland and R. Howell, "The fear factor: Privacy, fear, and the changing hegemony of the American people and the right to privacy," *NCJ Int'l L. & Com. Reg.*, vol. 29, p. 671, 2003.

[48] R. Nader *et al.*, "Joint petition for rulemaking on privacy notices," http://www.ftc.gov/bcp/workshops/glb/comments/, July 2001.

[49] OCC, Federal Reserve System, FDIC, OTS, NCUA, FTC, CFTC, and SEC, "Final model privacy form under the Gramm-Leach-Bliley Act," *Federal Register*, vol. 74, pp. 62 890–62 994, December 1, 2009.

[50] Terms of Service; Didn't Read, http://tosdr.org/.

[51] K. Badenhausen, "America's best and worst banks 2012," Forbes, http://www.forbes.com/sites/kurtbadenhausen/2012/12/18/full-list-americas-best-and-worst-banks-2012/, December 2012.

[52] J.D. Power & Associates, "2012 U.S. credit card satisfaction study," Press release, http://www.jdpower.com/content/press-release/xdTqU1T/2012-u-s-credit-card-satisfaction-study.htm, August 2012.

[53] "Gramm-Leach-Bliley Act," Pub. L. No. 106-102, 113 Stat. 1338, 1999.

[54] L. J. White, "The Gramm-Leach-Bliley Act of 1999: A bridge too far—or not far enough," *Suffolk UL Rev.*, vol. 43, p. 937, 2009.

[55] B. Shull, "Banking, commerce and competition under the Gramm-Leach-Bliley act," *Antitrust Bull.*, vol. 47, p. 25, 2002.

[56] J. R. Macey, "The business of banking: Before and after Gramm-Leach-Bliley," *J. Corp. L.*, vol. 25, p. 691, 1999.

[57] FTC, "Privacy of consumer financial information; final rule," Federal Register, May 2000.

[58] "Fair and Accurate Credit Transactions Act," Pub. L. No. 108-159, 117 Stat. 1952, 2003.

[59] P. L. McCorkell and A. M. Smith, "Fair Credit Reporting Act update—2008," *Business Lawyer*, vol. 64, no. 2, pp. 579–591, 2009.

[60] E. J. Janger and P. M. Schwartz, "Gramm-Leach-Bliley Act, information privacy, and the limits of default rules, the," *Minn. L. Rev.*, vol. 86, 2001.

[61] J. C. Schiller, "Informational privacy v. the commercial speech doctrine: Can the Gramm-Leach-Bliley Act provide adequate privacy protection," *Commlaw Conspectus*, vol. 11, p. 349, 2003.

[62] E. H. Freeman, "Privacy notices under the Gramm-Leach-Bliley Act," *Information Systems Security*, vol. 12, no. 2, pp. 5–9, 2003.

[63] G. T. Nojeim, "Financial privacy," *NYL Sch. J. Hum. Rts.*, vol. 17, p. 81, 2000.

[64] P. P. Swire, "Efficient confidentiality for privacy, security, and confidential business information," *Brookings-Wharton Papers on Financial Services*, vol. 2003, no. 1, pp. 273–310, 2003.

[65] J. M. Lacker, "The economics of financial privacy: to opt out or opt in?" *Economic Quarterly-Federal Reserve Bank of Richmond*, vol. 88, no. 3, pp. 1–16, 2002.

[66] M. Furletti and S. Smith, "Financial privacy: perspectives from the payment cards industry," *Payment Cards Center Discussion Paper*, 2003.

[67] US Financial Regulators, "Interagency financial institution web site privacy survey report," Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, Office of Comptroller of the Currency, Office of Thrift Supervision, Tech. Rep., 1999.

[68] X. Sheng and L. F. Cranor, "An evaluation of the effect of us financial privacy legislation through the analysis of privacy policies," *ISJLP*, vol. 2, p. 943, 2005.

[69] A. I. Antón, J. B. Earp, Q. He, W. Stufflebeam, D. Bolchini, and C. Jensen, "Financial privacy policies and the need for standardization," *IEEE Security & Privacy*, vol. 2, no. 2, pp. 36–45, 2004.

[70] Kleimann Communication Group Inc., "Evolution of a prototype financial privacy notice," http://www.ftc.gov/privacy/privacyinitiatives/ftcfinalreport060228.pdf, February 2006.

[71] Macro International Inc., "Mall intercept study of consumer understanding of financial privacy notices: Methodological report," http://www.ftc.gov/reports/quantitative-research-macro-international-report, September 2008.

[72] A. Levy and M. Hastak, “Consumer comprehension of financial privacy notices,” Interagency Notice Project, http://ftc.gov/privacy/privacyinitiatives/Levy-Hastak-Report.pdf, December 2008.

[73] Kleimann Communication Group Inc., “A report on validation testing results,” http://www.ftc.gov/reports/financial-privacy-notice-report-validation-testing-results-kleimann-validation-report, February 2009.

[74] L. Garrison, M. Hastak, J. M. Hogarth, S. Kleimann, and A. S. Levy, “Designing evidence-based disclosures: A case study of financial privacy notices,” *Journal of Consumer Affairs*, vol. 46, no. 2, pp. 204–234, 2012.

[75] A. L. Negroni and J. P. Kromer, “Gramm-Leach-Bliley: Tip of the privacy iceberg,” *Banking Law Journal*, vol. 118, no. 10, pp. 958–969, 2001.

[76] R. J. McMahon, “Developments in the Gramm-Leach-Bliley Act during 2005–06: An overview of important changes in case law and pending legislation,” *I/S: A Journal of Law & Policy for the Information Society*, vol. 2, no. 3, pp. 737–759, 2006.

[77] *Proskauer on Privacy: A Guide to Privacy and Data Security Law in the Information Age*. Practicing Law Institute, 2013.

[78] OECD, “Guidelines on the protection of privacy and transborder flows of personal data,” September 1980. [Online]. Available: http://www.oecd.org/internet/interneteconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

[79] Federal Trade Commission, “Privacy online: A report to Congress,” June 1998.

[80] Z. Tang, Y. J. Hu, and M. D. Smith, “Gaining trust through online privacy protection: Self-regulation, mandatory standards, or caveat emptor,” *Journal of Management Information Systems*, vol. 24, no. 4, pp. 153–173, 2008.

[81] J. Y. Tsai, S. Egelman, L. F. Cranor, and A. Acquisti, “The effect of online privacy information on purchasing behavior: An experimental study,” *Info. Sys. Research*, vol. 22, no. 2, pp. 254–268, Jun. 2011.

[82] M. Graber, D. D’Alessandro, and J. Johnson-West, “Reading level of privacy policies on Internet health web sites,” *Journal of Family Practice*, 2002.

[83] C. Jensen and C. Potts, “Privacy policies as decision-making tools: an evaluation of online privacy notices,” in *Proc. CHI*, 2004.

[84] B. Ur, M. Sleeper, and L. F. Cranor, “{Privacy, Privacidad, Приватност} policies in social media: Providing translated privacy notice,” *ISJLP*, vol. 9, no. 2, 2013.

[85] A. M. McDonald and L. F. Cranor, “The cost of reading privacy policies,” *I/S: A Journal of Law and Policy for the Information Society*, vol. 4, no. 3, pp. 540–565, 2008.

[86] L. F. Cranor, "Necessary but not sufficient: Standardized mechanisms for privacy notice and choice," pp. 273–307, 2012.

[87] L. F. Cranor, S. Egelman, S. Sheng, A. M. McDonald, and A. Chowdhury, "P3P deployment on websites," *Electronic Commerce Research and Applications*, vol. 7, no. 3, pp. 274–293, 2008.

[88] P. G. Kelley, J. Bresee, L. F. Cranor, and R. W. Reeder, "A "nutrition label" for privacy," in *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS)*, 2009, pp. 4:1–4:12. [Online]. Available: http://doi.acm.org/10.1145/1572532.1572538

[89] FDIC, "Institution directory," http://www2.fdic.gov/IDASP/, Accessed July 26, 2014.

[90] Federal Reserve, "Federal reserve's financial institution types," http://www.ffiec.gov/nicpubweb/content/help/LinkAdvancedSearchAllinstitutions.htm, Accessed July 26, 2014.

[91] FDIC, "Trust examination manual," http://www.fdic.gov/regulations/examinations/trustmanual/section_10/section_x.html#A, Accessed June 1, 2013.

[92] P. P. Swire, "Surprising virtues of the new financial privacy law, the," *Minn. L. Rev.*, vol. 86, p. 1263, 2001.

[93] "Ten steps to develop a multilayered privacy notice," The Center for Information Policy Leadership, 2007.

[94] L. F. Cranor, C. Hoke, P. G. Leon, and A. Au, "Are they worth reading? an in-depth analysis of online advertising companies' privacy policies," *Telecommunications Policy Research Conference*, Forthcoming (September 2014).

[95] Network Advertising Initiative, "2013 NAI code of conduct," http://www.networkadvertising.org/2013_Principles.pdf, 2013.

[96] AAAA, ANA, BBB, DNA, and IAB, *Self-Regulatory Principles for Online Behavioral Advertising*, Digital Advertising Alliance, July 2009.

[97] F. H. Cate, "The failure of fair information practice principles," *Consumer protection in the age of the information economy*, 2006.

[98] S. Zimmeck and S. M. Bellovin, "Privee: An architecture for automatically analyzing web privacy policies," *USENIX Security*, 2014.

[99] N. Sadeh, A. Acquisti, T. D. Breaux, L. F. Cranor, A. M. McDonald, J. Reidenberg, N. Smith, F. Liu, N. C. Rusell, F. Schaub, and S. Wilson, "The usable privacy policy project: Combining crowdsourcing, machine learning and natural language processing to semi-automatically answers those privacy questions users care about," http://usableprivacy.org/, Carnegie Mellon University, Tech. Rep., 2013. [Online]. Available: http://usableprivacy.org/

[100] S. Greengard, "Advertising gets personal," *Communications of the ACM*, vol. 55, no. 8, pp. 18–20, 2012.

[101] J. Turow, *The daily you: How the new advertising industry is defining your identity and your worth*. Yale University Press, 2012.

[102] R. Calo, "Digital market manipulation," *George Washington Law Review, Forthcoming*, 2013.

[103] P. Dixon and R. Gellman, "The scoring of America: How secret consumer scores threaten your privacy and your future," World Privacy Forum, April 2014.

[104] "Little blue book: A buyers guide," *Bluekai*, 2013.

[105] "Experian list services catalog," *Experian Marketing Services*, 2012.

[106] S. Sengupta. (2013, April) Facebook refines ad targeting.

[107] J. Valentino-Devries, "They know what you're shopping for," December 2012.

[108] K. D. Harris, *Making Your Privacy Practices Public*, http://oag.ca.gov/sites/all/files/agweb/pdfs/cybersecurity/making_your_privacy_practices_public.pdf, California Department of Justice, May 2014.

[109] J. B. Earp *et al.*, "Examining internet privacy policies within the context of user privacy values," *Engineering Management, IEEE Transactions on*, 2005.

[110] M. A. Graber, D. M. D Alessandro, and J. Johnson-West, "Reading level of privacy policies on internet health web sites," *Journal of Family Practice*, vol. 51, no. 7, pp. 642–642, 2002.

[111] G. R. Milne, M. J. Culnan, and H. Greene, "A longitudinal assessment of online privacy notice readability," *Journal of Public Policy & Marketing*, vol. 25, no. 2, pp. 238–249, 2006.

[112] A. I. Antón *et al.*, "HIPAA's effect on web site privacy policies," *Security & Privacy, IEEE*, 2007.

[113] J. Turow *et al.*, "Americans reject tailored advertising and three activities that enable it," *SSRN eLibrary*, 2009. [Online]. Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1478214

[114] K. Purcell, J. Brenner, and L. Rainie, "Search engine use 2012," Pew Internet and American Life Project. http://pewinternet.org/Reports/2012/Search-Engine-Use-2012.aspx, Tech. Rep., March 2012.

[115] L. Awarwal *et al.*, "Do not embarrass: re-examining user concerns for online tracking and advertising," in *Proc. SOUPS '13*. ACM, 2013.

[116] Evidon, Inc., "Evidon global tracker report," 2013, http://www.evidon.com/research.

[117] National Advertising Review Council, "Self-Regulatory Program for Children's advertising," Council of Better Business Bureaus, Inc., Tech. Rep., 2009.

[118] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang, "Smart, useful, scary, creepy: Perceptions of online behavioral advertising," in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, ser. SOUPS '12. New York, NY, USA: ACM, 2012, pp. 4:1–4:15. [Online]. Available: http://doi.acm.org/10.1145/2335356.2335362

[119] H. Beales, "The value of behavioral targeting," The Netwok Advertising Initiative http://www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf, Tech. Rep., 2010.

[120] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen, "How much can behavioral targeting help online advertising?" *WWW 2009*. [Online]. Available: http://doi.acm.org/10.1145/1526709.1526745

[121] K. W. Lendenmann, "Consumer perspectives on online advertising," Preference Central http://www.slideshare.net/mfredactie/preference-central-surveyfullreport, Tech. Rep., 2010.

[122] A. Farahat and M. Bailey, "How effective is targeted advertising?" *WWW 2012*.

[123] NAI, "Learn about online behavioral advertising, privacy, cookies, and how this all works!" http://networkadvertising.org/managing/learn_more.asp, Retrieved November 2011.

[124] B. Krishnamurthy and C. Wills, "Privacy diffusion on the web: A longitudinal perspective," in *Proc. WWW*, 2009.

[125] R. Balebako, P. Leon, R. Shay, B. Ur, Y. Wang, and L. Cranor, "Measuring the effectiveness of privacy tools for limiting behavioral advertising," *Web 2.0 Security and Privacy Workshop*, May 2012.

[126] P. Eckersley, "How unique is your web browser?" Electronic Frontier Foudation, EFF Report, 2009. [Online]. Available: panopticlick.eff.org/browser-uniqueness.pdf

[127] J. R. Mayer and J. C. Mitchell, "Third-party web tracking: Policy and technology," in *IEEE Symposium on Security and Privacy*, 2012.

[128] NAI, "FAQs," http://www.networkadvertising.org/managing/faqs.asp. Last accessed June 2012.

[129] P. Kosmala, "Yes, Johnny can benefit from transparency and control," DAA Blog, http://www.aboutads.info/blog/yes-johnny-can-benefit-transparency-and-control, October 2011, Retrieved November 2011.

[130] Network Advertising Initiative, "NAI code of conduct," http://www.networkadvertising.org/sites/default/files/imce/principles.pdf, 2008.

[131] T. Vega, “Ad group unveils plan to improve web privacy,” New York Times, http://www.nytimes.com/2010/10/04/business/media/04privacy.html, October 2010, retrieved March 2011.

[132] S. Rodgers, “The interactive advertising model tested: The role of Internet motives in ad processing,” *Journal of Interactive Advertising*, vol. 2, no. 2, pp. 22–33, 2002.

[133] R. Rettie, H. Robinson, and B. Jenner, “Does Internet advertising alienate users?” *Occasional Paper Series*, no. 52, 2003.

[134] S. McCoy, A. Everard, P. Polak, and D. Galletta, “The effects of online advertising,” *Communications of the ACM*, vol. 50, no. 3, pp. 84–88, 2007.

[135] D. Campbell and R. Wright, “Shut-up I don’t care: Understanding the role of relevance and interactivity on customer attitudes toward repetitive online advertising,” *Journal of Electronic Commerce Research*, vol. 9, no. 1, pp. 62–76, 2008.

[136] A. M. McDonald and L. F. Cranor, “Beliefs and behaviors: Internet users’ understanding of behavioral advertising,” *TPRC 2010*, 2010.

[137] M. Hastak and M. J. Culnan, “Online behavioral advertising “icon” study,” http://futureofprivacy.org/final_report.pdf, January 2010, unpublished Report.

[138] TRUSTe Research, “Consumer research results: Privacy and online behavioral advertising,” July 2011, http://www.truste.com/adprivacy/TRUSTe-2011-Consumer-Behavioral-Advertising-Survey-Results.pdf.

[139] KPMG International, “The converged lifestlye,” http://www.kpmg.com/convergence, 2011.

[140] DAAicon, “Advertising option icon application,” http://www.aboutads.info/participants/icon/. Last accessed June 2012.

[141] NAI, “Advertising Option Icon application,” http://www.aboutads.info/participants/icon/, Retrieved March 2012.

[142] V. Toubiana and V. Verdot, “Show me your cookie and I will tell you who you are,” arXiv.org, http://arxiv.org/abs/1108.5864, Tech. Rep., August 2011.

[143] AVG, “What is tracking and AVG do not track?” http://www.avg.com/ww-en/do-not-track, 2012.

[144] H. Nissenbaum, “Privacy as contextual integrity,” *Washington Law Review*, vol. 79, no. 1, 2004.

[145] Yahoo! Press Release, “Yahoo! announces global implementation of Do Not Track (DNT),” http://www.reuters.com/article/2012/03/29/idUS104733+29-Mar-2012+BW20120329, March 2012.

[146] P. Leon, B. Ur, R. Shay, Y. Wang, R. Balebako, and L. Cranor, "Why johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12. New York, NY, USA: ACM, 2012, pp. 589–598. [Online]. Available: http://doi.acm.org/10.1145/2207676.2207759

[147] M. Ayenson, D. J. Wambach, A. Soltani, N. Good, and C. J. Hoofnagle, "Flash cookies and privacy II," *SSRN eLibrary*, 2011. [Online]. Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1898390

[148] A. Soltani, S. Canty, Q. Mayo, L. Thomas, and C. J. Hoofnagle, "Flash cookies and privacy," *SSRN eLibrary*, 2009. [Online]. Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1446862

[149] L. F. Cranor, P. Guduru, and M. Arjula, "User interfaces for privacy agents," *ACM Transactions on Computer-Human Interaction*, vol. 13, 2006. [Online]. Available: http://dl.acm.org/citation.cfm?doid=1165734.1165735

[150] V. Ha, K. Inkpen, F. Al Shaar, and L. Hdeib, "An examination of user perception and misconception of internet cookies," in *CHI extended abstracts*. ACM, 2006.

[151] S. Lederer, J. Hong, A. Dey, and J. Landay, "Personal privacy through understanding and action: five pitfalls for designers," *Personal and Ubiquitous Computing*, vol. 8, no. 6, pp. 440–454, 2004.

[152] B. Brunk, "A user-centric privacy space framework," in *Security and Usability*, L. F. Cranor and S. Garfinkel, Eds. O'Reilly, 2005, ch. 20, pp. 401–420.

[153] L. F. Cranor, "Privacy policies and privacy preferences," in *Security and Usability*, L. F. Cranor and S. Garfinkel, Eds. O'Reilly, 2005, ch. 22, pp. 447–472.

[154] J. R. Lewis, *Handbook of Human Factors and Ergonomics*. John Wiley & Sons, Inc., 2006, ch. 49 Usability Testing, pp. 1275–1316.

[155] J. S. Dumas, *The Human-Computer Interaction Handbook*. Lawrence Erlbaum Associates, 2003, ch. User-Based Evaluations, pp. 1093–1117.

[156] L. F. Cranor, "A first look at Internet Explorer 9 privacy features," http://www.techpolicy.com/Blog/March-2011/A-first-look-at-Internet-Explorer-9-privacy-featur.aspx. [Online]. Available: http://www.techpolicy.com/Blog/March-2011/A-first-look-at-Internet-Explorer-9-privacy-featur.aspx

[157] P. G. Leon, J. Cranshaw, L. F. Cranor, J. Graves, M. Hastak, B. Ur, and G. Xu, "What do online behavioral advertising privacy disclosures communicate to users?" in *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, ser. WPES '12. New York, NY, USA: ACM, 2012, pp. 19–30. [Online]. Available: http://doi.acm.org/10.1145/2381966.2381970

[158] S.-C. Huang and R. G. Bias, "Icons versus texts in accuracy and efficiency," in *Proc. iConference*, 2012.

[159] S. Wiedenbeck, "The use of icons and labels in an end user application program: an empirical study of learning and retention," *Behaviour & Information Technology*, vol. 18, no. 2, 1999.

[160] K. Haramundanis, "Why icons cannot stand alone," *SIGDOC Asterisk Journal of Computer Documentation*, vol. 20, no. 2, pp. 1–8, May 1996. [Online]. Available: http://doi.acm.org/10.1145/381815.381819

[161] M. L. A. Kunnath, R. A. Cornell, M. K. Kysilka, and L. Witta, "An experimental research study on the effect of pictorial icons on a user-learner's performance," *Computers in Human Behavior*, vol. 23, no. 3, pp. 1454–1480, May 2007. [Online]. Available: http://dx.doi.org/10.1016/j.chb.2005.05.005

[162] A. Wang, "Visual priming of pharmaceutical advertising disclosures: effects of a motivation factor," *Corporate Communications: An International Journal*, vol. 17, no. 1, 2012.

[163] J. C. Andrews, S. Burton, and J. Kees, "Is simpler always better? Consumer evaluations of front-of-package nutrition symbols," *Journal of Public Policy & Marketing*, vol. 30, no. 2, pp. 175 – 190, 2011.

[164] A. Y. Lee, "Effects of implicit memory on memory-based versus stimulus-based brand choice," *Journal of Marketing Research*, vol. 39, no. 4, pp. pp. 440–454, 2002. [Online]. Available: http://www.jstor.org/stable/1558556

[165] D. M. Boush, "How advertising slogans can prime evaluations of brand extensions," *Psychology and Marketing*, vol. 10, no. 1, pp. 67–78, 1993. [Online]. Available: http://dx.doi.org/10.1002/mar.4220100106

[166] M. Dahlén and S. Rosengren, "Brands affect slogans affect brands? Competitive interference, brand equity and the brand-slogan link," *Journal of Brand Management*, vol. 12, no. 3, pp. 151–164, 2005.

[167] S. S. Williams and C. P. Koepke, "Using multiple methods to identify theme lines for media marketing campaigns to promote Medicare information sources," *Journal of Consumer Marketing*, vol. 23, no. 6, 2006.

[168] P. G. Kelley, L. Cesca, J. Bresee, and L. F. Cranor, "Standardizing privacy notices: An online study of the nutrition label approach," in *Proc. CHI*, 2010.

[169] Microsoft, "Privacy in Internet Explorer 6," Jun. 2011.

[170] J. Ross, L. Irani, M. S. Silberman, A. Zaldivar, and B. Tomlinson, "Who are the crowdworkers? Shifting demographics in Mechanical Turk," in *Proc. CHI Extended Abstracts*, 2010.

[171] P. G. Ipeirotis, "Demographics of Mechanical Turk," New York University, Tech. Rep. CeDER-10-01, 2010.

[172] G. Paolacci, J. Chandler, and P. G. Ipeirotis, "Running experiments on Amazon Mechanical Turk," *Judgment and Decision Making*, vol. 5, no. 5, 2010.

[173] M. Buhrmester, T. Kwang, and S. D. Gosling, "Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?" *Perspectives on Psychological Science*, vol. 6, no. 1, pp. 3–5, 2011.

[174] P. G. Leon, B. Ur, Y. Wang, M. Sleeper, R. Balebako, R. Shay, L. Bauer, M. Christodorescu, and L. F. Cranor, "What matters to users?: Factors that affect users' willingness to share information with online advertisers," in *Proceedings of the Ninth Symposium on Usable Privacy and Security*, ser. SOUPS '13. New York, NY, USA: ACM, 2013, pp. 7:1–7:12. [Online]. Available: http://doi.acm.org/10.1145/2501604.2501611

[175] J. McEntegart, "Microsoft sticks to "Do Not Track" plans for IE in Windows 8," in *Tom's Hardware*, August 2012.

[176] M. Geuss, "Firefox will block third-party cookies in a future version," in *Ars Technica*, February 2013.

[177] N. Singer, "Mediator joins contentious effort to add a "Do Not Track" option to web browsing," in *New York Times*, November 2012.

[178] F. Roesner, T. Kohno, and D. Wetherall, "Detecting and defending against third-party tracking on the web," in *Proc. NSDI*, 2012.

[179] J. Burkell and A. Fortier, "Consumer health websites and behavioural tracking," in *Proc. of the 40th Annual Conference of the CAIS*, 2012.

[180] B. Krishnamurthy and C. E. Wills, "On the leakage of personally identifiable information via online social networks," in *Proc. WOSN*, 2009.

[181] A. Roosendaal, "We are all connected to Facebook...by Facebook!" in *European Data Protection: In Good Health?* Springer, 2012, pp. 3–19.

[182] P. Golle, "Revisiting the uniqueness of simple demographics in the us population," in *Proc. WPES*, 2006.

[183] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Proc. IEEE Symposium on Security and Privacy*, 2008.

[184] A. Lambrecht and C. Tucker, "When Does Retargeting Work? Information Specificity in Online Advertising," *Journal of Marketing Research*, vol. 5, pp. 561–576, 2013.

[185] "Social Advertising," *SSRN eLibrary*, 2012, http://ssrn.com/abstract=1975897.

[186] J. Gomez, T. Pinnick, and A. Soltani, "KnowPrivacy," http://www.knowprivacy.org/report/KnowPrivacy_Final_Report.pdf, Jun. 2009.

[187] N. F. Awad and M. Krishnan, "The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization," *Management Information Systems Quarterly*, vol. 30, no. 1, 2006.

[188] C. E. Wills and M. Zeljkovic, "A personalized approach to web privacy: awareness, attitudes and actions," *Information Management & Computer Security*, vol. 19, no. 1, pp. 53–73, 2011.

[189] A. McDonald and J. Peha, "Track gap: Policy implications of user expectations for the "Do Not Track" internet privacy feature," *Information Privacy Law eJournal*, vol. 5, 2012.

[190] C.-H. Cho and H. J. Cheon, "Why do people avoid advertising on the internet?" *Journal of Advertising*, 2004.

[191] A. Acquisti, L. K. John, and G. Loewenstein, "The impact of relative standards on the propensity to disclose," *Journal of Marketing Research*, vol. 49, no. 2, pp. 160–174, 2012.

[192] D. G. Taylor, D. F. Davis, and R. Jillapalli, "Privacy concern and online personalization: The moderating effects of information control and compensation," *Electronic Commerce Research*, vol. 9, no. 3, pp. 203–223, 2009.

[193] A. N. Joinson, U.-D. Reips, T. Buchanan, and C. B. P. Schofield, "Privacy, trust, and self-disclosure online," *Human–Computer Interaction*, vol. 25, no. 1, pp. 1–24, 2010.

[194] M. J. Metzger, "Effects of site, vendor, and consumer characteristics on web site trust and disclosure," *Communication Research*, vol. 33, no. 3, pp. 155–179, 2006.

[195] E. Van De Garde-Perik, P. Markopoulos, B. De Ruyter, B. Eggen, and W. Ijsselsteijn, "Investigating privacy attitudes and behavior in relation to personalization," *Social Science Computer Review*, vol. 26, no. 1, pp. 20–43, 2008.

[196] B. Krishnamurthy, K. Naryshkin, and C. Wills, "Privacy leakage vs. protection measures: the growing disconnect," in *Proc. W2SP*, 2011.

[197] H. Nissenbaum, "A contextual approach to privacy online," *Daedalus*, vol. 140, no. 4, pp. 32–48, 2011.

[198] N. K. Malhotra, S. S. Kim, and J. Agarwal, "Internet users' information privacy concerns (iuipc): the construct, the scale, and a causal model," *Information Systems Research*, vol. 15, no. 4, pp. 336–355, 2004.

[199] Y. Wang, P. G. Leon, A. Acquisti, L. F. Cranor, A. Forget, and N. Sadeh, "A field trial of privacy nudges for facebook," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 2367–2376. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557413

[200] Y. Wang, S. Komanduri, P. G. Leon, G. Norcie, A. Acquisti, and L. F. Cranor, ""I regretted the minute I pressed share": A qualitative study of regrets on facebook," in *SOUPS*, 2011.

[201] A. Acquisti, *Privacy in Electronic Commerce and the Economics of Immediate Gratification*, 2004.

[202] I. Adjerid, A. Acquisti, L. Brandimarte, and G. Loewenstein, "Sleights of privacy: Framing, disclosures, and the limits of transparency," in *Proceedings of the Ninth Symposium on Usable Privacy and Security*. ACM, 2013, p. 9.

[203] R. H. Thaler and C. R. Sunstein, "Libertarian paternalism," *Am. Econ. Rev.*, vol. 93, no. 2, pp. 175–179, 2003. [Online]. Available: http://www.aeaweb.org/articles.php?doi=10.1257/000282803321947001

[204] A. Acquisti, "Nudging privacy: The behavioral economics of personal information," *IEEE Security and Privacy*, vol. 7, no. 6, pp. 82–85, 2009.

[205] E. Goffman, *The Presentation of Self in Everyday Life*, 1st ed. Anchor, Jun. 1959.

[206] A. Marwick and d. boyd, "I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience," *New Media & Society*, vol. 13, no. 1, p. 114, 2011.

[207] A. Acquisti and R. Gross, "Imagined communities: Awareness, information sharing, and privacy on the Facebook," *6th Workshop on Privacy Enhancing Technologies*, pp. 36–58, 2006. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.73.2056

[208] Michelle Madejski, Maritza Johnson, and Steven M. Bellovin, "The failure of online social network privacy settings," Columbia University, Tech. Rep. CUCS-010-11, Feb. 2011.

[209] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in *WPES*, 2005, pp. 71–80. [Online]. Available: http://dx.doi.org/10.1145/1102199.1102214

[210] d. boyd and N. Ellison, "Social network sites: Definition, history, and scholarship," *J. of Computer-Mediated Commun.*, vol. 13, no. 1, 2007. [Online]. Available: http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html

[211] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *science*, vol. 185, no. 4157, pp. 1124–1131, 1974.

[212] H. A. Simon, "A behavioral model of rational choice," *The Quarterly J. of Econ.*, vol. 69, no. 1, pp. 99–118, Feb. 1955. [Online]. Available: http://qje.oxfordjournals.org/content/69/1/99

[213] G. A. Akerlof, "The market for "lemons": Quality uncertainty and the market mechanism," *The quarterly journal of economics*, pp. 488–500, 1970.

[214] M. S. Bernstein, E. Bakshy, M. Burke, and B. Karrer, "Quantifying the invisible audience in social networks," in *SIGCHI*. ACM, 2013, pp. 21–30. [Online]. Available: http://doi.acm.org/10.1145/2470654.2470658

[215] D. Laibson, "Golden eggs and hyperbolic discounting," *The Quarterly J. of Econ.*, vol. 112, no. 2, pp. 443–478, May 1997. [Online]. Available: http://qje.oxfordjournals.org/content/112/2/443

[216] S. Frederick, G. Loewenstein, and O. T., "Time discounting and time preference: A critical review," *J. of Econ. Lit.*, vol. 40, no. 2, pp. 351 – 401, 2002.

[217] D. Kahneman, "A perspective on judgment and choice: Mapping bounded rationality," *American psychologist*, pp. 697–720, 2003.

[218] R. H. Thaler and C. R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness*, 1st ed. Yale University Press, Apr. 2008.

[219] A. Forget, S. Chiasson, P. C. van Oorschot, and R. Biddle, "Improving text passwords through persuasion," in *SOUPS*. ACM, 2008, pp. 1–12.

[220] B. Ur, P. Kelley, S. Komanduri, J. Lee, M. Maass, M. Mazurek, T. Passaro, R. Shay, T. Vidas, L. Bauer, N. Christin, and L. Cranor, "How does your password measure up? the effect of strength meters on password creation," in *Proc. USENIX Security*, 2012.

[221] S. Wilson, J. Cranshaw, N. Sadeh, A. Acquisti, L. F. Cranor, J. Springfield, S. Y. Jeong, and A. Balasubramanian, "Privacy manipulation and acclimation in a location sharing application," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 2013, pp. 549–558.

[222] E. Choe, J. Jung, B. Lee, and K. Fisher, "Nudging people away from privacy-invasive mobile apps through visual framing," in *INTERACT*. Springer Berlin Heidelberg, 2013, pp. 74–91.

[223] F. Stutzman, R. Gross, and A. Acquisti, "Silent listeners: The evolution of privacy and disclosure on facebook," *Journal of Privacy and Confidentiality*, vol. 4, no. 2, p. 2, 2013.

[224] L. Fang and K. LeFevre, "Privacy wizards for social networking sites," in *WWW*. ACM, 2010, pp. 351–360.

[225] H. R. Lipford, J. Watson, M. Whitney, N. Carolina, H. Lipford, K. Froiland, and R. W. Reeder, "Visual vs . Compact : A Comparison of Privacy Policy Interfaces," *Interfaces*, pp. 1111–1114, 2010.

[226] A. Besmer and H. Lipford, "Tagged photos: concerns, perceptions, and protections," in *Proc. CHI Ext. Abs.*, 2009, pp. 4585–4590. [Online]. Available: http://portal.acm.org/citation.cfm?id=1520704

[227] K. E. Jenni and G. Loewenstein, "Explaining the identifiable victim effect," *Journal of Risk and Uncertainty*, vol. 14, no. 3, pp. 235–257, 1997. [Online]. Available: http://papers.ssrn.com/abstract=5187

[228] Y. Wang, P. G. Leon, K. Scott, X. Chen, A. Acquisti, L. F. Cranor, and N. Sadeh, "Privacy nudges for social media: An exploratory facebook study," *PSOSM*, 2013.

[229] F. H. Cate, P. Cullen, and V. Mayer-Schőnberg, "Data protection principles for the 21st century. revising the 1980 oecd guidelines," Tech. Rep., March 2014, http://www.oii.ox.ac.uk/publications/Data_Protection_Principles_for_the_21st_Century.pdf.

[230] R. Balebako, R. Shay, and L. F. Cranor, "Is your inseam a biometric? a case study of the role of usability studies in developing public policy," *Workshop on Usable Security Experiments (USEC)*, 2014.

[231] L. Brandimarte, A. Acquisti, and G. Loewenstein, "Misplaced confidences privacy and the control paradox," *Social Psychological and Personality Science*, vol. 4, no. 3, pp. 340–347, 2013.

[232] A. Cavoukian and D. Reed, "Big privacy: Bridging big data and the personal data ecosystem through privacy by design," 2013, http://www.privacybydesign.ca/content/uploads/2013/12/pbd-big_privacy.pdf.

[233] United States Government Accountability Office, "Information Resellers: Consumer privacy framework needs to reflect changes in technology and market place," GAO-13-663, September 2013.

[234] D. D. Hirsch, "The glass house effect: Big data, the new oil, and the power of analogy," *Maine Law Review*, vol. 66, p. 2, 2014.