# Beliefs and Behaviors: Internet Users' Understanding of Behavioral Advertising

Aleecia M. McDonald Lorrie Faith Cranor

August 16, 2010*

**Abstract**

**AUTHORS' PRE-PRESS VERSION**
**Please cite to the published paper from TPRC**

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

*This paper is a substantially extended version of [30].

# 1 Introduction

Real-time mass media was born with national radio networks in the 1920s. As mass media gave rise to mass advertising, advertisers' campaigns became national. However, typically only a subset of people are interested in any given product or service advertised. As the old advertisers' lament has it, "We know we're wasting half our ad dollars, we just don't know which half" [14]. Online advertising can be targeted to users most likely to be interested in a particular product or service. Customers may benefit from ads targeted to their personal interests, reducing irrelevant ads and the time it takes to find products.

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

The Internet is a form of mass media with targeted advertisements dependent on massive data collection on a tremendous scale. The Yahoo! ad server reaches over half a billion unique people each month, with 9.7% of the market [7]. Google's DoubleClick and AdSense ad servers have a combined total of 56% of the market and reach at least 1.5 billion unique users each month [7]. Google web beacons are on 88% of nearly 400,000 sampled websites and 92 of the top 100 most popular sites [20]. Google is reported to track approximately 90% of global Internet users [12]. The collection, storage, and use of the data that drives advertising has tremendous potential for privacy harm, as illustrated in the release of AOL search terms [10] and social networking information exposed by Gmail users trying Google Buzz [32]. There are four public policy domains that can benefit from understanding user perceptions of Internet advertising:

1. Legislation. State and Federal legislatures are considering new regulations around Internet privacy, including proposals from Representatives Boucher, Stearns, Rush, and Senator Kerry. Understanding what constituents know can help define legislative priorities: in areas where people are already able to protect their privacy interests, there is reduced justification for new laws.

2. Industry self-regulation. The Federal Trade Commission (FTC) and industry groups continue their efforts to improve corporate privacy practices without the burdens of regulation. Self-regulation presumes Internet users can make decisions to enact their privacy preferences, which makes understanding preferences, knowledge, and behavior a valuable contribution to evaluating self-regulation.

3. Consumer expectations. The FTC and privacy professionals within companies increasingly look to issues of surprise to decide which practices are acceptable [9]. With users' subjective responses to privacy loss being used as guidance, rather than formal approaches like privacy rights frameworks, it is crucial to know how users react to current online practices.

4. Education. The first "principle" in the *Self-Regulatory Program for Online Behavioral Advertising* is education [1]. Establishing a baseline of user knowledge before campaigns begin will help establish their successes and any short comings.

In this paper we review related work in section 2 and describe our methods in section 3. We present our findings regarding using cookie management as a self-help mechanism, participants' views of tailored advertising, and their willingness to pay for privacy in sections 4, 5, and 6 respectively. We conclude in section 7.

# 2 Background and Related Work

Targeted advertising has received a lot of scrutiny in the past few years. There are questions about consumer's online privacy, how easily seemingly anonymous information can be re-identified [33], and the legality of some behavioral advertising business practices. The advertising industry favors continuing an "industry self-regulation" approach. The Federal Trade Commission has held workshops and released guidelines for self-regulation [19, 18], and there are legislative proposals at the Federal [11] and State [6] level, including proposals from Representatives Boucher, Stearns, Rush, and Senator Kerry.

In 2008, TRUSTe commissioned a report on behavioral advertising, finding 57% of respondents are "not comfortable" with browsing history-based behavioral advertising, "even when that information cannot be tied to their names or any other personal information" [41]. In 2009, TRUSTe found that even if it "cannot be tied to my name or other personal information," only 28% of Internet users would feel comfortable with advertisers using web browsing history, and 35% believe their privacy has been invaded in the past year due to information on the Internet [42]. Anton, et. al., performed some of the earliest work on behavioral advertising in 2002, with a follow up study in 2009 [5]. They found the types of privacy concerns remained stable, but the level of concern has increased around information used for behavioral advertising. Gomez et al. estimated that Google Analytics tracks at least 329,330 unique domains, and found confusion in privacy policies containing "conflicting statements that third-party sharing is not allowed but third-party tracking and affiliate sharing are" [20]. Turow et. al. conducted a nationally representative phone survey in 2009. They found 66% of adults do not want tailored advertising, which increased to as high as 86% when participants were informed of three common techniques used in advertising [45]. In 2003, Turow found that when offered a choice between paying for their favorite website with cash or with their personal information, over half of respondents said they would rather stop using the site all together [44]. Several experiments investigated under which conditions people will pay more to purchase from websites offering better

privacy protections when privacy information is presented in search results and in other salient ways [43, 17].

Much of the current self-regulation approach to online privacy is grounded in the Fair Information Principle of notice. Notice, by its nature, requires communication. As Morgan et al. wrote, "An effective communication must focus on the things that people need to know but do not already. This seemingly simple norm is violated remarkably often in risk communication" [31]. We investigated people's mental models — beliefs about how a system works, interacts, or behaves. Incorrect mental models may form a view of the world that undermines decision making. For example, if people hold the mental model that any company with a privacy policy is bound by law not to release data, the existence of a link to a privacy policy would seem sufficient in and of itself with reduced reason to read the policy. Research shows that people do, in fact, believe the words "privacy policy" mean they are protected by law [24].

Economics literature suggests that the most someone is willing to pay (WTP) to buy something should be equal to the minimum they are willing to accept (WTA) in payment for it: there should be a point of indifference between the good and cash. A difference between WTP and WTA may be indicative of an *endowment effect*, a phrase coined by Richard Thaler to describe when people place more value on an object that they own. The canonical example is that if two groups are asked to put a value on a coffee mug, people answering without owning the mug will generally suggest a lower price than people who first receive the mug as their own property. The endowment effect does not always occur with abstract items. For example, giving people a token that they can redeem for a mug does not have the same effect as giving them the actual mug [16]. Prior work shows a gap between WTP and WTA for revealing private data (for example, number of sexual partners) in an offline experiment [21]. Acquisti et. al. found substantial differences between WTP and WTA with gift cards and inexpensive tangible goods, including "subjects who started from positions of greater privacy protection were five times more likely than other subjects to forego money to preserve that protection" [3]. We examine the Acquisti hypothesis in an online context. If there is also a gap between WTP and WTA online, then the way privacy choices are framed may affect the decisions people make about online privacy.

# 3 Research Methods

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

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

In the second study, we recruited 314 participants from the Mechancal Turk[1] website at the end of April, 2010. We paid participants $2 for what we advertised as a 20-30 minute study. Median completion time was 24 minutes, which is skewed slightly high by participants who likely started the survey, put it aside, and came back to it later. We saw a drop-out rate of 37%. We deliberately started the study with short-answer questions to encourage people not to take the survey unless they were willing to invest some time, and used the reasonableness of responses to short-answer questions to screen participants. We removed two outliers from our dataset; they had unusually short response times and response patterns that suggested they had not read the questions. We coded free-form responses to tabulate categories of responses, or "unclear" when we were unsure what participants had in mind.

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

## 3.1 Demographics

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

For the online study, we slightly over-represented women and our population was notably skewed younger than the adult American Internet population, as seen in Table 1. To estimate the demographics for US Adult Internet Users, we combined Pew data [34] with Census data [46]. Because Pew and the Census data record race differently, we cannot estimate the portion of Internet users by race. Instead we contrast to national race statistics from the Census. We under-sampled black and hispanic populations. Our respondents were 74% were white (contrast to 81% nationally), 9% American Indian or Alaskan Native (v. 2%), 6% Asian (v. 5%), 4% Black or African American (v. 14%),

[1]Mechanical Turk is crowd-source web portal run by Amazon. See `www.mturk.com` for details. Mechanical Turk users tend to be better educated, less likely to be working, and more likely to be female than our target population of adult US Internet users. However, the Mechanical Turk population may be "more appropriate" for Internet research, as mturk studies can be closer to representative of Internet users than a random sampling of the full US population [36]. There is a growing literature on how best to use Mechanical Turk in research; see [26], [27], [25], [13]

and 2% Latina/Latino or Hispanic (v. 16%).[2] We contrast to Quantcast's estimates for operating systems [35] and Axon's for web browsers [8]. Our sample is skewed toward Firefox users at the expense of Internet Explorer, which suggests a more technically sophisticated sample.

Table 1: Demographics for online study

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

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

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

## 3.2 Transferability

Early in the online study, before we asked questions that might affect participants' views, we asked the same three questions Turow et al. asked in their study designed to be representative of the US population [45]. As our sample is not a statistically representative sample of United States Internet users, we contrasted to the Turow work

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

to understand the transferability of results to other contexts [23]. We found similar results for two of their three questions, as shown in Table 2.

Table 2: Percentage of respondents who want tailored content

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

Our respondents' differ demographically from the Turow population in an important way: our sample is skewed younger. Where the representative Turow sample is comprised of 35% of people aged 18 – 34, our sample is 69% in that age range. However, despite age-linked differences in responses, our younger sample does not explain why we saw a substantially higher percentage interested in tailored discounts. We had approximately 20% more interest in tailored discounts in all of our age categories as compared to the Turow work, as seen in Figure 1. One possible explanation: we recruited participants willing to spend 20 minutes to answer our survey for $2 on the Mechanical Turk website. Our participants may be unusually sensitive to financial incentives. For tailored ads and news, our findings mirrored the Turow paper: most respondents are not interested in tailored advertisements or news.

# 4 Perceptions About Cookies

A variety of technologies help facilitate online behavioral tracking for targeted advertising. Third-party cookies are used by advertising companies to set cookies associated with ads embedded in first-party sites; when browsers load advertisers' ads, they also get advertisers' third-party cookies. Beacons are associated with invisible or hidden page elements, and again may be first- or third-party. Flash cookies were designed to store information like volume levels for flash content, but are now used in tracking [37]. Browser fingerprinting is a technique that uses information from web browsers' user agents (for example, the specific version number and operating system) plus potentially other information available from javascript (for example, the specific order fonts load on the system.) By using small bits of seemingly unidentifiable information in concert, approximately 80% of browsers are uniquely identifiable[15]. We planned to survey participants about what they understood about each of these tracking mechanisms.

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

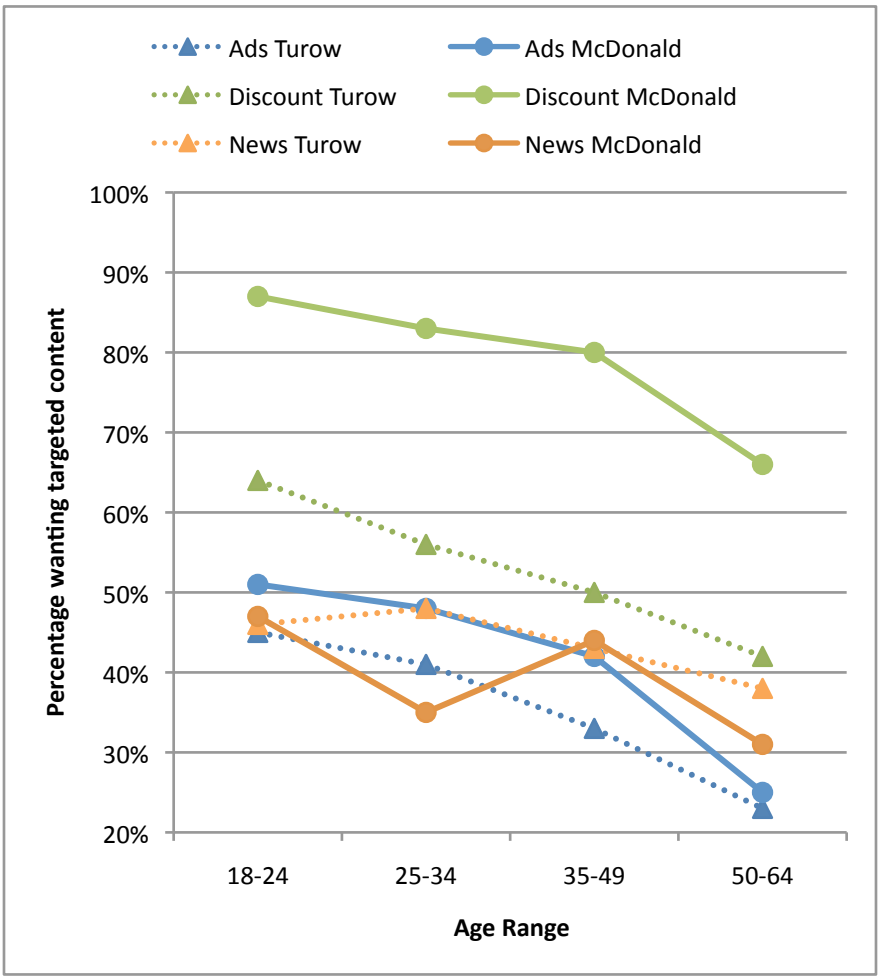



Figure 1: Percentage of respondents interested in targeted ads, discounts, and news by age groups in the Turow study, as contrasted with our results

mechanism in advertising, but also because they are the only technology users are even passingly familiar with. Cookies have been around, discussed, and studied across decades [2, 22]. If users understand behavioral advertising well enough to attempt to enact their privacy preferences, they are most likely to be able to do so via cookie management. We asked questions to study participants' knowledge of cookies, how they manage cookies, and see if they understand the industry self-regulation approach of setting cookies to opt out of viewing behavioral advertising.

All participants in the interviews had heard of cookies before but we observed widespread confusion. When asked, "What is a cookie?" nearly a third of participants replied immediately that they were not sure. Slightly more than a third of participants

gave an answer that was at least partially correct without also saying something factually incorrect. Only one person articulated that a cookie can contain a unique identifier.

## 4.1 Misperceptions of First Party Cookies

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

Only three participants said that cookies are related to personalized advertisement. They expressed three very different perspectives. One participant said she has no choices about cookies, because if you "say no then you don't get to go to the site. That's not much of an option." She could not think of any way cookies help her. For ways cookies do not help, she said sites use cookies to personalize, and that "could mean more personalized advertising. It makes me feel like they expect me to be gullible." A second said cookies are things "that programs use to gather information about sites [visited], functionality, and demographics for an ad." He said that "if asked for information [people] would say no," and believes he has "no choices" about cookies. He said that cookies are good when "a set pattern of behaviors, sites, topics, or hobbies" can give "information on products and services that are more interesting," but "some [cookies] are used negatively to exploit a person's history," and "cookies open pools of information one might prefer to stay private." Drawing an analogy to shopping offline, he said "you may be shopping in a public place but there is a privacy issue" with companies "knowing where you spend money and time." Even with a computer collecting and storing the data, there still must be a "person manipulating and interpreting that." A third participant said advertisers use cookies to "find out as much as [advertisers] can without asking for names," to gain an "idea of what sort of person" you are. He mentioned ISPs trying to "find ways to catalog this wealth of information," to pair ads to an audience. He described this practice as a "smart thing" and "reasonable." He then volunteered that he believes ISPs are constrained by law not to share information. When asked what the law entails, he answered he was not sure and perhaps constraints were not from law but that there would be a "public uproar" and a "bad image" for any company sharing even anonymous customer data. He made the analogy to phone service where recording conversations can be illegal, and said there are "certain cultural norms and expectations" to privacy. Notice the analogies to off-line settings as participants form their views of how privacy works online. Legal protection of privacy in telephone conversations and postal mail are often assumed to carry over to Internet communications as well.

## 4.2 Knowledge of Cookies

Based on the responses we heard in interviews, we asked questions in the online study to understand participants' knowledge of cookies with options of True, False, or Unsure. See Table 3 for details.

Table 3: Responses to factual questions about cookies — correct answers in bold

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

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

are confused, including half of participants who believe if they do not accept cookies their location can not be identified, and half believe cookies contain information from when they purchased their computer, including their name and home address, with more than another third unsure if this is true or not. Neither of these beliefs is true, and shows a lack of understanding of how cookies work and what they are. Similarly, 44% incorrectly believe cookies enable the forward and backward arrows in their browser, a third incorrectly believe all websites can read all cookies, and a quarter confuse cookies and history, incorrectly thinking cookies change visited hyperlinks to a different color. There is nearly an even split in thirds between participants who believe cookies are a type of spyware, are not a type of spyware, or are unsure — which is not unreasonable, as there is disagreement within the technical community as to whether some cookies, all cookies, or no cookies should be seen as spyware. 29% believe cookies are responsible for spam, which is not the case, with another 33% unsure. The greatest confusion is around legal protections. 30% incorrectly believe cookies may not contain credit card information by law, with 59% uncertain. 38% incorrectly believe the PATRIOT ACT allows law enforcement to read cookies if they exchange email with "someone on the terrorist watch list," with 56% unsure. This very last question is the only one we wrote ourselves: all of the rest come directly from perceptions from our lab study participants. Several participants expressed concern that the government could read cookies, but used vague language; we tested a specific example.

## 4.3 Managing Cookies

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

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

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

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

## 4.4 Unclear on Clearing Cookies

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

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

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

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

Of 278 reasons given to delete cookies:

- 33% were based on the idea that "many cookies slow down my computer." This seems unlikely in practice.[3]
- 30% had to do with privacy and security. About a fifth of the privacy and security reasons mentioned deleting history; history is commonly confused with cookies. The remaining four-fifths of privacy and security reasons generally reflected some understanding of how cookies work, for example, "I wouldn't want someone being able to get on my computer and remain logged into my accounts. Also, I don't want a website tracking me through them."
- 28% had to do with freeing up hard drive space, reducing clutter, or a notion of hygiene and cleanliness. Answers included "[I] like having a clean slate on the computer all the time," "[to] clear up clutter," and "to make space on my

[3]For DSL users, a webpage with a 3000 byte cookie takes approximately 80 milliseconds longer to load [40] so users are not wrong to associate cookies with delay. However, just deleting all cookies without blocking them does not improve time to load the page: websites would simply download new cookies to replace the deleted cookies. Participants may be confusing cookies with cached images.

computer." Few modern computers will run into space problems due to cookies.[4]

- 8% mention viruses, spam, or malware. Some tracking cookies are classified as spyware by Norton Anti-virus and other anti-malware programs.

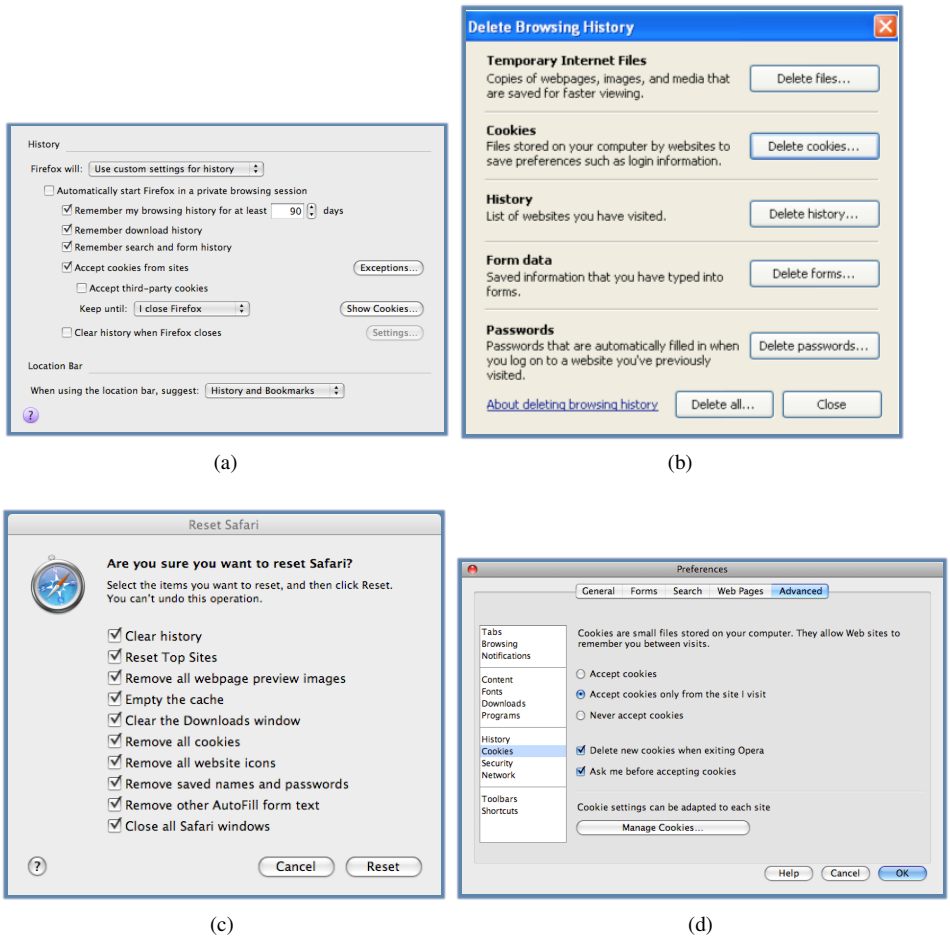
User confusion is high. Some do not know how to delete cookies and might wish to do so, which limits self-help mechanisms in privacy decision making. Some participants reported what seems to be over-clearing of cookies: they delete cookies to avoid issues that cookies do not cause. Cookie deletion creates uncertainty in measuring the number of people — and unique people — who have seen a given online ad, or have visited a given website. Disagreement over ad impressions has slowed the growth of web ads. Counting impressions often depends upon cookie data. Over- and under-counting ad impressions causes economic harms to members of the advertising community, with hundreds of thousands of dollars disputed in large ad campaigns [38]. When users delete their cookies for reasons that do not match their actual preferences, it causes harm without the gains users expect.

## 4.5 Cookies and Browser History

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

Browser user interfaces in which clearing cookies, clearing history, and clearing cache data settings are intermingled may contribute to user confusion. One component of this confusion is temporal: participants reported they delete cookies and clear history at the same time, which leads them to misattribute properties of browser history to cookies. The reason participants clear cookies and history together likely stems from the way they are swirled together in the user interfaces of web browsers. For example, Firefox presents choices about cookies, history, and bookmarks on the same tab, as shown in Figure 2a. There is no visual hint that these three topics are distinct. To the contrary, cookies are in the middle of options for history, which serves to convey history and cookies are related. Moreover, Firefox does not expose any cookie options

[4]RFC 2109 suggests browsers implement a maximum size of 4k per cookie and a maximum number of cookies per domain to avoid denial of service attacks from malicious servers filling hard drives [28], and hard drives today are typically measured in gigabytes.




(a) (b)

(c) (d)

Figure 2: Four browsers' interfaces for deleting cookies: Firefox, Internet Explorer, Safari, and Opera.

unless users know to change a setting from Remember history to Use custom settings for history. Anyone looking through preference tabs for cookies will not find them in the default configuration. In Internet Explorer, users must select the Tools menu and then choose Delete Browsing History in order to get to the cookie dialog, shown in Figure 2b. The easiest way to delete cookies in Safari is to select Reset Safari from the Safari menu, which then presents options to delete cookies and history together as shown in Figure 2c. The exception is Opera, shown in Figure 2d. Cookies are not mixed in with history. The Opera dialog attempts to define cookies and avoids jargon.

In the online study, we asked "Sometimes you hear about web browser history. Are cookies and history the same?" 35% of participants incorrectly answered yes. Those who answered no generally had a good working understanding of the difference

between cookies and history, with responses like "History is a list of your previous browsing, and cookies are files that registered each site visited." Of those who correctly answered no, 79% were able to give at least partially correct answers explaining how cookies and history differ, 12% gave clearly incorrect answers, and 8% gave answers that were so unclear we were not able to tell if they understood the difference or not.

## 4.6 Lack of Understanding of Cookies and Data Flows

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

We asked: "Please refer to the images below to answer questions at the bottom of this page. Imagine you are using a standard web browser to visit The Times website, which has ads as depicted in the diagrams. *There are no other non-visible components to the webpage.*" and gave a choice of four different figures, shown here as Figures 3a, 3b, 3c, and 3d, along with a brief text description of each image. We followed up by asking "Which, if any, of the diagrams above could not happen?"

- 22% selected Figure 3a, described as: "The Times' web server sets and reads cookies for all elements on the webpage, including cookies associated with specific ads." While advertising could work like this, with each host storing and displaying all ads from just the host's server, modern websites are usually more complicated. 9% answered, incorrectly, that this configuration could never happen.
- 20% selected Figure 3b, described as: "Multiple web servers set and read cookies from The Times' web page." This graphic introduces the concept of multiple actors with multiple servers, but incorrectly depicts them all being able to read and write cookies from the same section of the website. Servers cannot set and read cross-domain cookies, so this configuration is unlikely, especially in practice. 18% answered that this configuration could never happen, which is a reasonable answer.
- 18% selected Figure 3c, described as: "Only the Times' server can set and read cookies on the Times web page." This graphic does emphasize the lack of cross-domain cookies, but also shows ads that do not set cookies. While this is possible, it is highly unlikely on modern sites. 15% answered that this configuration could never happen.
- 40% selected Figure 3d, described as: "Different servers set and read cookies from different parts of the Times' web page." This is the best choice. It shows common relationships between hosts and advertisers. 10% answered, incorrectly, that this configuration could never happen.

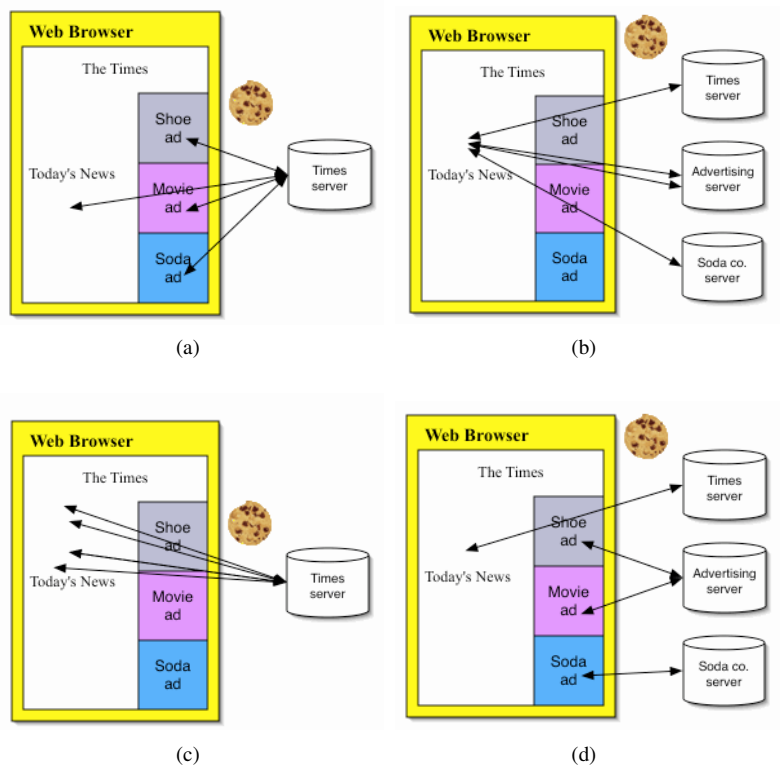



Figure 3: Four possible mental models of how advertising cookies work.

- 48% answered that all four figures are possible.

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

## 4.7 Consumers Do Not Understand NAI Opt-Out Cookies

None of our interview participants had heard of cookie-based methods to opt-out of tracking cookies, including TACO[5] and NAI opt-out cookies.[6] At the end of the pro-

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

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

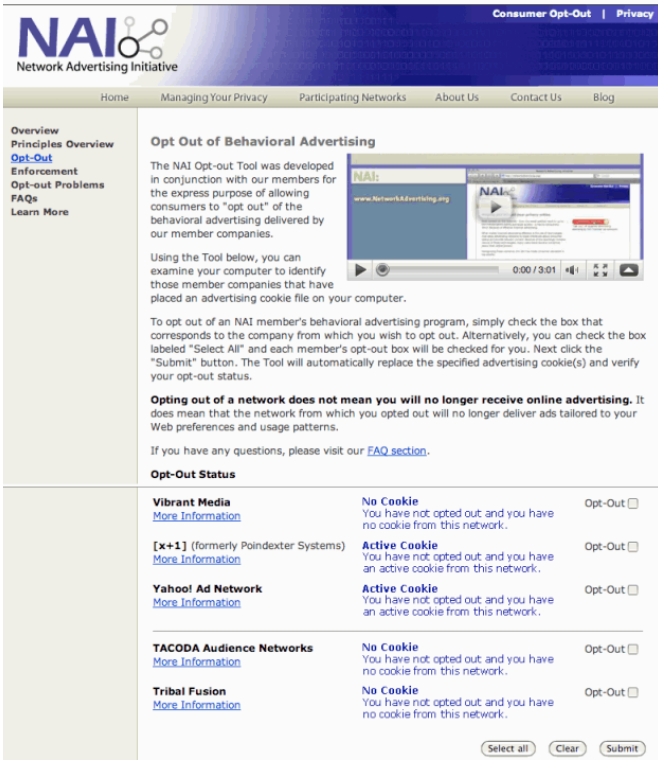



Figure 4: Screenshot of the NAI Opt Out page

tocol, we showed four participants a text description of NAI opt-out cookies from the NAI opt-out website (see Figure 4.)[7]

All four participants understood they would continue to see at least some online advertisements. However, there was substantial confusion about what the NAI opt-out does. The text does not disclose that companies may choose to continue all data collection and profiling, and that in some cases the only thing that changes is the type of ads displayed [4]. One participant understood this but the other three did not.

The first participant believed the NAI opt-out "sets your computer or ethernet so information doesn't get sent." She still expected to see ads, but now the ads would be "random." She said it might "sound old fashioned" but in a choice between "convenience and privacy, I'm going to pick privacy." She was afraid that opt-out meant "all these people get your information" and therefore "this could be a phishing expedition." A second participant began his comments by saying "Where do I click? I want this!" He believed the NAI opt-out to be an "opt-out tool so users opt out of being tracked." He thought "the ads are still there, they just get no data." A third participant thought it would "reduce the amount of online advertising you receive." He understood data

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

collection was also involved, but not how, just "some sort of control over what companies use that information." He would choose to opt-out of companies where "the information they would seek would be too personal to share with a group." Our final participant understood the NAI text. At first he said if you use Gmail, the opt-out cookie means "stop reading my email and tailoring ads." He later clarified "What you search is Google property, it's theirs. They're going to profile you but not show you that they are."

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

- 34% answered "A website that lets you tell companies not to collect data about you." There are some companies for which this is the case. However, some NAI members like Yahoo! continue to collect data exactly as before; they just do not tailor ads to reflect that data.
- 25% answered "A website that lets you tell companies you do not want to see ads from them, but you will still see as many ads overall." This is incorrect because companies continue to serve ads, just not targeted ads. The ad source is unchanged.
- 18% answered "A website that lets you see fewer online ads." This is both wrong and prominently disclaimed in the NAI text.
- 11% answered "A website that allows companies to profile you, but not show you ads based on those profiles." **Correct answer.**
- 6% answered "A scam website to collect your private information."
- 5% answered "A scam website to find out which websites you have visited."

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

# 5 Tailored Content and Privacy Concerns

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

## 5.1 Mixed Identification of Internet Advertising

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

We asked how advertising on Google works. All participants understood that advertisers pay Google to run ads. Participants were less clear on the mechanics of payment. Some thought Google charges for all ads displayed, and some thought Google only charges for ads when people click on them. No one described beliefs that were technically impossible; everything described has occurred at one time. All told, this is a fairly sophisticated understanding of Google's contextual advertising during search tasks.

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

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

### 5.1.1 Inability to Distinguish Widgets

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

Widgets are another part of a web page, for example an embedded clock or the weather. Widgets are often designed to be customizable to match the look and feel of the web site they are dropped into. Industry guidelines assume people can distinguish third party widgets from first party content and assume that people understand that data flows differently to third party advertisers. Therefore they treat third party widget providers as first party data collectors, subject to fewer guidelines [1]:

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

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

## 5.2 Mixed Understanding of Current Practices

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

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

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

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

Table 4: Perceived likelihood of practices occurring

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

As shown in Table 4, participants seem to have a high degree of understanding that behavioral advertising happens, with only 13% of respondents casting doubt that current practices occur. Yet only 40% believe advertising based on email content is happening today, and 29% believe this common practice will never occur.

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

For both scenarios we asked, "How would you feel about this practice?" (Participants were able to select more than one answer.) As shown in Table 5, the most popular answer is that 46% of participants find behavioral advertising "creepy," but a small group of 18% welcome targeted advertisements. Responses on how people feel about advertising based on email are markedly more negative, with 62% saying email should be private and that they find ads based on email creepy. Only 4% of respondents saw email-based advertising as a benefit, and only 9% supported the trade off of data and advertising for free services. This matches what we heard in interviews: people understand ads support free content, but do not believe data are part of the deal.

Table 5: Attitudes toward current practices

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

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

## 5.3 Reasons to Accept or Reject Tailored Advertising

Based on discussions in the laboratory study, we compiled a list of reasons participants gave for being for or against behavioral advertising. We presented online participants with a seven point Likert scale from Strongly Agree (7) to Strongly Disagree (1), summarized in Table 6.

Privacy concerns are top priorities. Nearly two-thirds of our participants agreed or strongly agreed that "someone keeping track of my activities online is invasive," with only 4% disagreeing or strongly disagreeing. This phrase comes directly from a participant we interviewed in the lab study, and reflects the way she thought about behavioral advertising. It is phrased in a way that would likely garner maximum response by mentioning an unnamed, but presumably human, "someone" and using the possessive "my." We suggest the way to understand this result is that if behavioral advertising is framed this way in the press, most Americans will respond poorly to it.

Again expressing privacy concerns, 40% agreed or strongly agreed they would be more careful online if they knew advertisers were collecting data. The wording of this question limits data use to advertisers, which may reduce concern. It also explorers the notion of a chilling effect. Respondents at least believe they would self-censor if they knew advertisers were collecting data. While self-reported data is not always indicative of actual behavior, it appears people are considering leaving FaceBook in response to publicity about data flows to advertisers [39]. Advertiser's practices have the potential

Table 6: Mean Likert scores to accept or reject behavioral advertising (Strongly Agree = 7, Strongly Disagree = 1.)

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

to reduce Internet adoption and use, and may already be doing so.

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

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

agree or strongly agree,) with the most strictly apathetic option of not caring if ads are targeted or random in the middle (31% agree or strongly agree.)[8] This suggests that of those who ignore ads, they are likely to prefer data not be collected about them, since they do not see any benefit. However, just because someone claims to ignore ads does not mean that is always the case. Advertisers may still gain benefit from targeting these users. But an argument that targeted ads are a benefit will likely fall flat with the people who are not interested in any ads, let alone better ads. Interestingly, when we put that question to participants directly, we saw an even split. 21% agree or strongly agree that they want the benefits of relevant advertising while 21% disagree or strongly disagree, with a neutral Likert mean of 4.1. What emerges is neither a strong clamoring for nor a backlash against behavioral advertising, but rather several distinct groups with quite different preferences.

## 5.4 Privacy and Security Among Top Priorities for Buying Online

98% of our participants indicated they make purchases online. More than half said they never make purchases based on Internet ads or email advertising, as summarized in Table 7. This is self-reported data; people may make buying decisions based on ads without being aware they are doing so. Banner ads serve a billboard-like function for those who eventually buy online, even months later [29].

Table 7: Respondents who buy online

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

We asked participants how sellers could entice participants to purchase more products online, and listed 13 possible approaches with responses on a four point Likert scale of "Matters a lot," "Matters," "Matters a little," and "Does not matter." We created our 13 categories based on responses to a pilot test with an open-ended question. See Table 8 for results.

The most popular item was free shipping.[9] The next three most popular were clustered around privacy and security: not sharing data with advertisers, a policy against spam, and fraud protection. In contrast, the remaining privacy and security item on data retention scored near the very bottom. This may be a function of the specific description, or due to lack of understanding of how data retention limits reduce privacy and security risks, but suggests data retention is not currently a major concern for users.

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

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

Table 8: How sellers can entice more online purchases (Matters a lot = 4, Does not matter = 1)

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

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

# 6 Payment for Privacy

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

## 6.1 Gap Between Willingness to Pay and Willingness to Accept

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

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

## 6.2 Reasons to Pay or Refuse to Pay for Privacy

We followed up by asking questions to better understand why people would decide to pay or accept $1, based on reasons we heard from our lab study participants. We asked "Some websites may offer you a choice of paying for content or receiving content for free in exchange for letting them send you targeted advertising. How strongly do you agree or disagree with the following statements?" with a seven point Likert scale from Strongly Agree (7) to Strongly Disagree (1). See Table 9 for details.

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

Table 9: Reasons to pay for privacy or accept a discount

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

decision making, with only 11% willing to pay to avoid ads because they "hate" them. Most participants are accustomed to advertising. Mass media advertising has been part of life since before they were born. It is the data collection that is new, and, to many, a troubling aspect of online advertising.

# 7 Conclusions and Discussion

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

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

Most users understand that cookies store data on their computers, enable tailored ads, and allow tracking across sites. They are unclear on important details like whether cookies may be combined with other data, what data is stored in cookies, if blocking cookies preserves geolocational privacy, and they are particularly unclear about laws

and law enforcement. Web browsers may contribute to users' confusion. Browsers may also be an avenue to help with user understanding and decision making in the future. Thus far, browser makers have been largely absent from behavioral advertising issues, and some do not directly profit from behavioral advertising. Microsoft has been involved with behavioral advertising for years, and their adoption of P3P in Internet Explorer changed the third-party cookie landscape. Google's Chrome browser may be another opportunity to welcome browser makers as major stakeholders with tremendous ability to help Internet users make privacy decisions. However, the *Wall Street Journal* reported that Microsoft re-designed Internet Explorer 8 specifically to enable third party tracking for business reasons [47]. It may be naive to expect browser makers to support user privacy at their own expense. The NAI is as a major player in behavioral advertising but their opt-out cookie page is very confusing, with only 11% understanding what it is for. With their leadership role in self-regulation, the NAI may not be supporting Internet users' ability to avail themselves of self-help options.

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

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

With lack of understanding of and a lack of interest in tailored content, unless industry moves rapidly towards an effective self-regulatory solution, regulation may be needed. One possible path for regulation is to require opt-in for all forms of advertising other than contextual. However, opt-in systems are not a panacea: they can be designed so users click them away without understanding them, and once users opt-in it may be difficult to reverse the choice. If industry elected to, they could use self-regulation mechanisms to improve decision making through education, improved technology and tools, and more privacy-protective policies far more quickly than regulators could act. These tasks will be challenging no matter which parties take the initiative.

# 8 Acknowledgments

Thanks to Faisal N. Jawdat and to Greg Norcie for coding open-ended questions. Thanks to participants at the Privacy Law Scholars Conference (PLSC 2010) and to CUPS laboratory members for invaluable feedback. This research was supported in

part by CyLab at Carnegie Mellon under grant DAAD19-02-1-0389 from the Army Research Office and by Microsoft Research.

# References

[1] AAAA, ANA, BBB, DMA, AND IAB. Self-regulatory program for online behavioral advertising, 2009. http://www.iab.net/media/file/ven-principles-07-01-09.pdf Accessed 23 January 2010.

[2] ACKERMAN, M. S., CRANOR, L. F., AND REAGLE, J. Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM Conference on Electronic Commerce* (November 1999). http://doi.acm.org/10.1145/336992.336995 Accessed 1 March 2010.

[3] ACQUISTI, A., JOHN, L., AND LOEWENSTEIN, G. What is privacy worth? Tech. rep., Heinz College, Carnegie Mellon University, 2009.

[4] ANDERSON, S. House subcommittees hold joint hearing on behavioral advertising. *Security, Privacy and the Law* (July 2009). http://www.securityprivacyandthelaw.com/2009/07/articles/recent-legislation-1/house-subcommittees-hold-joint-hearing-on-behavioral-advertising/ Original testimony available from http://www.youtube.com/watch?v=-Wklp2qdbmw. Accessed 9 November 2009.

[5] ANTON, A. I., EARP, J. B., AND YOUNG, J. D. How Internet users' privacy concerns have evolved since 2002. Tech. Rep. Computer Science Technical Report TR-2009-16, North Carolina State, 2009. http://theprivacyplace.org/blog/wp-content/uploads/2009/07/tr_2009_16.pdf Accessed 3 March 2010.

[6] ARIAS, M. L. Internet law – behavioral advertising in the United States. *Internet Business Law Services* (June 2009). http://www.ibls.com/internet_law_news_portal_view.aspx?s=latestnews&id=2237 Accessed 9 November 2009.

[7] ATTRIBUTOR. Google ad server share now at 57%. microhoo less than 15% market share., December 2008. http://www.attributor.com/blog/google-ad-server-share-now-at-57-microhoo-less-than-15-market-share/ Accessed 3 March 2010.

[8] AXON, S. Chrome gains, IE slumps in browser wars. The Social Media Guide, May 2010. http://mashable.com/2010/05/04/chrome-firefox-ie-stats/ Accessed 15 May 2010.

[9] BAMBERGER, K. A., AND MULLIGAN, D. K. Privacy on the books and on the ground. *Stanford Law Review 63* (2010). UC Berkeley Public Law Research Paper No. 1568385. http://ssrn.com/abstract=1568385 Accessed 14 May 2010.

[10] BARBARO, M., AND ZELLER JR, T. A face is exposed for AOL searcher no. 4417749. *New York Times* (August 2006).

[11] BOORTZ, A. R. New federal privacy bill in the works: Behavioral advertising "beneficial," but must be done "appropriately". *AdLaw By Request* (August 2009). http://www.adlawbyrequest.com/2009/08/articles/legislation/new-federal-privacy-bill-in-the-works-behavioral-advertising-beneficial-but-must-be-done-appropriately/ Accessed 9 November 2009.

[12] CLELAND, S. The blind eye to privacy law arbitrage by Google – broadly threatens respect for privacy. Testimony Before the House Energy & Commerce Subcommittee On Telecommunications and the Internet. Hearing on "What Your Broadband Provider Knows About Your Web Use: Deep Packet Inspection and Communications Laws and Policies", July 2008. netcompetition.org/Written_Testimony_House_Privacy.pdf Accessed 14 May 2010.

[13] DOWNS, J. S., HOLBROOK, M. B., SHENG, S., AND CRANOR, L. F. Are your participants gaming the system? screening mechanical turk workers. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems* (New York, NY, USA, 2010), ACM, pp. 2399–2402.

[14] DYER, D., DALZELL, F., AND OLEGARIO, R. *Rising Tide: Lessons from 165 Years of Brand Building at Procter & Gamble*. Harvard Business Press, 2004.

[15] ECKERSLEY, P. How unique is your browser? In *Proceedings of the Privacy Enhancing Technologies Symposium (PETS)* (2010).

[16] The endowment effect: It's mine, I tell you. *The Economist* (June 2008). http://www.economist.com/science/displaystory.cfm?story_id=11579107 Accessed 2 September 2008.

[17] EGELMAN, S., TSAI, J., CRANOR, L. F., AND ACQUISTI, A. Timing is everything? The effects of timing and placement of online privacy indicators. In *CHI 2009* (Boston, MA, USA, April 2009).

[18] FEDERAL TRADE COMMISSION. FTC staff revises online behavioral advertising principles, February 2009. http://www.ftc.gov/opa/2009/02/behavad.shtm Accessed 15 May 2009.

[19] FEDERAL TRADE COMMISSION. Self-regulatory principles for online behavioral advertising: Tracking, targeting, and technology. Staff Report, February 2009. http://www.ftc.gov/os/2009/02/P085400behavadreport.pdf Accessed 9 November 2009.

[20] GOMEZ, J., PINNICK, T., AND SOLTANI, A. Knowprivacy, June 2009. http://www.knowprivacy.org/report/KnowPrivacy_Final_Report.pdf Accessed 4 March 2010.

[21] GROSSKLAGS, J., AND ACQUISTI, A. When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. In *Workshop on the Economics of Information Security (WEIS)* (2007).

[22] HA, V., INKPEN, K., AL SHAAR, F., AND HDIEB, L. An examination of user perception and misconception of Internet cookies. In *CHI '06 Extended Abstracts of Human Factors in Computing Systems* (April 2006). http://doi.acm.org/10.1145/1125451.1125615 Accessed 1 March 2010.

[23] HARPER, S. R., AND KUH, G. D. Myths and misconceptions about using qualitative methods in assessment. In *Using qualitative methods in institutional assessment. New Directions for Institutional Research*, S. R. Harper and S. D. Museus, Eds., no. 136. Jossey-Bass, San Francisco, 2007, pp. 5–14.

[24] HOOFNAGLE, C. J., AND KING, J. What Californians understand about privacy online, 2008 September. http://ssrn.com/abstract=1262130 Accessed 7 Nov 2008.

[25] KELLEY, P. G. Conducting usable privacy & security studies with amazon's mechanical turk. In *Symposium on Usable Privacy and Security (SOUPS)* (Redmond, WA, July 2010).

[26] KITTUR, A., CHI, E. H., AND SUH, B. Crowdsourcing user studies with mechanical turk. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (2008), ACM, pp. 453–456. http://doi.acm.org/10.1145/1357054.1357127 Accessed 5 March 2010.

[27] KOSARA, R., AND ZIEMKIEWICZ, C. Do mechanical turks dream of square pie charts? In *BELIV '10: BEyondtime and errors: Novel evaLuation methods for Information Visualization* (2010), ACM, pp. 373–382.

[28] KRISTOL, D., AND MONTULLI, L. RFC2109 - HTTP state management mechanism. http://www.faqs.org/rfcs/rfc2109.html Accessed 10 May 2010.

[29] MANCHANDA, P., DUBE, J.-P., GOH, K. Y., AND CHINTAGUNTA, P. K. The effect of banner advertising on internet purchasing. *Journal of Marketing Research XLIII* (February 2006), 98–108.

[30] MCDONALD, A. M., AND CRANOR, L. F. Americans' attitbudes about internet behavioral advertising practices. In *Proceedings of the 9th Workshop on Privacy in the Electronic Society (WPES)* (October 4 2010).

[31] MORGAN, M. G., FISCHHOFF, B., BOSTROM, A., AND ATMAN, C. J. *Risk Communication: A Mental Models Approach*. Cambridge University Press, 2002.

[32] MULLINS, R. Privacy group argues buzz breaks wiretap laws. *VentureBeat* (February 2010). http://venturebeat.com/2010/02/17/privacy-group-argues-buzz-breaks-wiretap-laws/ Accessed 14 May 2010.

[33] OHM, P. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review (forthcoming)* (2010). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006 Accessed 10 November 2009.

[34] PEW INTERNET & AMERICAN LIFE PROJECT. Report: Internet, broadband, and cell phone statistics, 2009. http://www.pewinternet.org/Reports/2010/Internet-broadband-and-cell-phone-statistics.aspx?r=1 Accessed 26 July 2010.

[35] PRESCOTT, L. OS X share up 29% in past year, slowly chipping away at Microsoft. VentureBeat, February 2010. http://venturebeat.com/2010/02/26/os-x-share-up-29-in-past-year-slowly-chipping-away-at-microsoft/ Accessed 15 May 2010.

[36] ROSS, J., ZALDIVAR, A., IRANI, L., AND TOMLINSON, B. Who are the turkers? worker demographics in amazon mechanical turk. Technical report socialcode-2009-01, Univeristy of California, Irvine, 2009.

[37] SOLTANI, A., CANTY, S., MAYO, Q., THOMAS, L., AND HOOFNAGLE, C. J. Flash cookies and privacy, August 11 2009. http://ssrn.com/abstract=1446862 Accessed 15 Apr 2010.

[38] STORY, L. How many site hits? depends who's counting. *New York Times* (October 2007).

[39] SUTTE, J. D. Some quitting facebook as privacy concerns escalate. *CNN Tech* (May 2010). http://www.cnn.com/2010/TECH/05/13/facebook.delete.privacy/index.html?iref=allsearch Accessed 15 May 2010.

[40] THEURER, T. Performance research, part 3: When the cookie crumbles. *Yahoo! User Interface Blog* (March 2007). http://yuiblog.com/blog/2007/03/01/performance-research-part-3/ Accessed 10 May 2010.

[41] TRUSTE. 2008 study: Consumer attitudes about behavioral targeting, March 2008. http://danskprivacynet.files/wordpress.com/2009/02/truste2008_tns_bt_study_summary1.pdf Accessed 9 November 2009.

[42] TRUSTE AND TNS. 2009 study: Consumer attitudes about behavioral targeting, March 2009.

[43] TSAI, J., EGELMAN, S., CRANOR, L. F., AND ACQUISTI, A. The effect of online privacy information on purchasing behavior: An experimental study. In *The 6th Workshop on the Economics of Information Security (WEIS)* (2008). http://weis2007.econinfosec.org/papers/57.pdf Accessed 22 Feb 2009.

[44] TUROW, J. Americans & Online Privacy: The System is Broken. Annenberg Public Policy Center Report, 2003.

[45] TUROW, J., KING, J., HOOFNAGLE, C. J., BLEAKLEY, A., AND HENNESSY, M. Americans reject tailored advertising and three activities that enable it. Tech. rep., Annenberg School for Communications, University of Pennsylvania, September 2009. http://repository.upenn.edu/asc_papers/137/ Accessed 4 March 2010.

[46] U.S. CENSUS BUREAU, POPULATION DIVISION. Annual estimates of the resident population by sex and five-year age groups for the united states: April 1, 2000 to july 1, 2009 (nc-est2009-01), June 2010.

[47] WINGFIELD, N. Microsoft quashed effort to boost online privacy. *Wall Street Journal* (August 2 2010).