

Behavioral Advertising: The Offer You Cannot Refuse

*Chris Jay Hoofnagle,¹ Ashkan Soltani,² Nathaniel Good,³
Dietrich J. Wambach,⁴ and Mika D. Ayenson^{5,6}*

INTRODUCTION

At UC Berkeley, we are informing political debates surrounding online privacy through empirical study of website behaviors. In 2009 and 2011, we surveyed top websites to determine how they were tracking consumers. We found that advertisers were using persistent tracking technologies that were relatively unknown to consumers. Two years later, we found that the number of tracking cookies expanded dramatically and that advertisers had developed new, previously unobserved tracking mechanisms that users cannot avoid even with the strongest privacy settings.

These empirical observations are valuable for the political debate surrounding online privacy because they inform the framing and assumptions surrounding the merits of privacy law.

Our work demonstrates that advertisers use new, relatively unknown technologies to track people, specifically because consumers have not heard of these techniques. Furthermore, these technologies obviate choice mechanisms that consumers exercise.

In the political debate, “paternalism” is a frequently invoked objection to privacy rules. Our work inverts the assumption that privacy interventions are paternalistic while market approaches promote freedom. We empirically demonstrate that advertisers are making it impossible to avoid online tracking. Advertisers are so invested in the idea of a personalized web that they do not think consumers are competent to decide to reject it. We argue that policymakers should fully appreciate the idea that consumer privacy inter-

¹ Lecturer in Residence, University of California, Berkeley, School of Law (Boalt Hall).

² MIMS, University of California, Berkeley, School of Information; independent researcher and consultant focused on privacy, security, and behavioral economics.

³ Ph.D., Chief Scientist and Principal of Good Research.

⁴ Dietrich J. Wambach is a senior at the University of Wyoming.

⁵ Mika D. Ayenson is a junior at the Worcester Polytechnic Institute.

⁶ This work was supported exclusively by Team for Research in Ubiquitous Secure Technology (TRUST), which receives support from the National Science Foundation (NSF award number CCF-0424422) and the following organizations: AFOSR (#FA9550-06-1-0244), BT, Cisco, ESCHER, HP, IBM, iCAST, Intel, Microsoft, ORNL, Pirelli, Qualcomm, Sun, Symantec, Telecom Italia, and United Technologies. We are grateful for the opportunities offered by the TRUST Research Experiences for Undergraduates program (REU), and to its former program leader, Dr. Kristen Gates, who is now the Director of Graduate Programs at the Buck Institute. The technical version of this report is available here: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1898390; supplemental information is available here: http://ashkansoltani.org/docs/respawn_redux.html.

ventions can enable choice, while the alternative, pure marketplace approaches can deny consumers opportunities to exercise autonomy.

THE ONLINE TRACKING DEBATE

The rise of telemarketing created tensions between marketers and consumers. Prior to the creation of the National Do Not Call Registry, telemarketers could call any phone number in the country, and the burden was upon the consumer to opt out from each caller. Doing so was not easy because telemarketers adopted a number of choice-invalidating techniques that prevented consumers from avoiding sales calls. On the political front, the telemarketing industry opposed the creation of a universal opt out, wishing to preserve a company-by-company opt out approach.⁷ They also wanted consumers to have to reenroll in the Registry regularly, perhaps every two years. These policies increased transaction costs for consumers and allowed every telemarketer in the world to ring consumers' phones at least once.

On the technical front, consumers who used devices to avoid telemarketing, such as the Telezapper, soon found the intervention to be ineffective, as telemarketers developed countermeasures.⁸ Telecommunications companies played both sides of the market, by marketing Caller ID service to consumers and at the same time, by marketing telephone equipment that did not send Caller ID to telemarketers. These steps rendered technical interventions to avoid telemarketing ineffective.

By 2003, rules required the transmission of Caller ID and required telemarketers to respect opt out choices on a universal level.⁹ Now, 209 million people have listed their phone numbers on the Registry,¹⁰ and those who enroll receive fewer sales calls.¹¹

⁷ See AM. TELESERVICES ASS'N, COMMENTS OF THE AMERICAN TELESERVICES ASSOCIATION ON THE REVIEW OF THE TELEMARKEETING SALES RULE 10 (2000), available at <http://www.ftc.gov/bcp/rulemaking/tsr/comments/ata.pdf> ("Additionally, the company specific 'Do-Not-Call' list is the best way to empower consumers to make the type of informed purchasing decisions that are necessary for a satisfactory sale. For consumers who do not want to receive calls, all they have to do is inform the caller at anytime during the call. However, for those consumers who want to receive calls or who only want to receive certain types of calls, the existing federal rule allows them the freedom to determine which calls they want to receive and prohibits those calls they don't.").

⁸ Scott Hovanyetz, *Call Center Mailer Touts TeleZapper Immunity*, DIRECT MARKETING NEWS (Feb. 18, 2003), <http://www.dmnews.com/call-center-mailer-touts-telezapper-immunity/article/80083/>.

⁹ See generally Telemarketing Sales Rule, 16 C.F.R. § 310 (2011).

¹⁰ FED. TRADE COMM'N, BIENNIAL REPORT TO CONGRESS: UNDER THE DO NOT CALL REGISTRY FEE EXTENSION ACT OF 2007, FY 2010 AND 2011 (2011), available at <http://www.ftc.gov/os/2011/12/111230dncreport.pdf>.

¹¹ Press Release, Harris Interactive, National Do-Not-Call Registry: Seven in Ten Are Registered and All of Them Will Renew Their Registration, Large Majority Who Have Registered Report Receiving Far Fewer Telemarketing Calls 1 (Oct. 31, 2007), available at <http://www.harrisinteractive.com/vault/Harris-Interactive-Poll-Research-Do-Not-Call-2007-10.pdf>.

The modern direct marketing debate concerns Internet tracking. This debate possesses some of the features of the telemarketing controversy. An innumerable array of companies specialize in monitoring individuals' use of the Internet. They do so for the purposes of testing the performance and functionality of websites, for measuring how popular sites are, and for tailoring advertising to individual users. This last purpose—tailoring advertising—has become politically controversial because in order to pitch relevant advertising to individuals, companies have strong incentives to monitor individuals' use of the Internet pervasively and to build profiles of users.¹² These profiles are a kind of file about the consumer; they could include information about past Internet use or demographic information, or classify the consumer into different kinds of “types” or “segments,” which can be used for targeting of advertisements.

In 2010, the *Wall Street Journal* focused a series of articles on this monitoring, finding that the “nation’s 50 top websites on average installed 64 pieces of tracking technology onto the computers of visitors, usually with no warning.”¹³ The *Wall Street Journal* series *What They Know* has been one of the most important expositions of Internet tracking and has piqued the interests of regulators¹⁴ and the U.S. Congress.¹⁵

As with the telemarketing debate, online advertisers have strongly resisted universal choice mechanisms for consumers that would allow users to avoid tracking. Under pressure from the Obama administration, advertisers recently relented, agreeing in principle to a universal “Do Not Track” mechanism. However, advertisers have made key caveats that may render the mechanism ineffective.¹⁶ For instance, some have argued that social widgets, such as the Facebook “Like” button and the Google “+1” feature should not be blocked by Do Not Track.¹⁷ This means that even those who

¹² JOSHUA GOMEZ ET AL., KNOWPRIVACY 5 (2009), available at http://www.knowprivacy.org/report/KnowPrivacy_Final_Report.pdf.

¹³ Julia Angwin, *The Web’s New Gold Mine: Your Secrets*, WALL ST. J., July 30, 2010, at W1.

¹⁴ See Julia Angwin & Amir Efrati, *Google Settles With FTC Over Buzz*, WALL ST. J. (Mar. 30, 2011), <http://online.wsj.com/article/SB10001424052748703806304576232600483636490.html> (“At a Senate hearing earlier this month, FTC Chairman Jon Leibowitz said that The Wall Street Journal’s ‘What They Know’ series on online privacy prompted the agency to ‘step up our enforcement efforts.’”).

¹⁵ Memorandum to Members of the Subcomm. on Commerce, Trade, and Consumer Protection From Subcomm. on Commerce, Trade, and Consumer Protection Democratic Staff Regarding Hearing on “Do Not Track Legislation: Is Now the Right Time?” (Nov. 30, 2010), available at <http://democrats.energycommerce.house.gov/documents/20101201/Briefing.Memo.12.01.2010.pdf>.

¹⁶ See Rainey Reitman, *White House, Google, and Other Advertising Companies Commit to Supporting Do Not Track*, EFF DEEPLINKS BLOG (Feb. 23, 2012), <https://www.eff.org/deeplinks/2012/02/white-house-google-and-other-advertising-companies-commit-supporting-do-not-track>.

¹⁷ *What Does Tracking Mean?*, MOZILLA DEVELOPER NETWORK (last updated Sept. 8, 2011), https://developer.mozilla.org/en/The_Do_Not_Track_Field_Guide/Introduction/What_does_tracking_mean.

enable Do Not Track will be followed online by Google and Facebook.¹⁸ Google, as explained more fully below, already has an unrivaled capacity to monitor users online.

One way that websites track users is through “cookies,” small text files that typically contain a string of numbers that can be used to identify a computer. For instance, a website might set a tracking cookie on a user’s computer with a key (a fancy word for the cookie name) such as “id” and value (the unique identifier assigned to a user) such as “123456789.” Advertisers can then access the “id” cookie and track how user 123456789 visits different websites.¹⁹

A common distinction is drawn between first-party and third-party cookies (TPCs). The former are issued by the website the user is visiting, the latter by some other website.²⁰ TPCs are commonly used to track users across different websites²¹ by companies that have no relationship with consumers. Thus for privacy sensitive users, blocking TPCs is seen as a convenient and effective way of preventing tracking by advertising and other companies without disabling the basic functionality of the web.²² By 2005, over twelve percent of users were rejecting TPCs.²³ In addition, with colleagues, author Hoofnagle found in 2009 that thirty-nine percent of American Internet users delete all their cookies “often”; only twenty-one percent never deleted cookies or did not know what they were.²⁴

The privacy problem from cookies comes from the aggregation of this tracking across different websites into profiles and through attempts at linking this profile to the user’s identity. By tracking these identifiers across websites that users visit, advertisers can infer users’ interests,²⁵ perhaps sensitive ones, such as medical conditions, political opinions, or even sexual fetishes.²⁶ While one might dismiss this as not problematic, arguing that the tracking is performed without using personal information, there are many popular mechanisms to link identifying information to a formerly pseudony-

¹⁸ See Arnold Roosendaal, *Facebook Tracks and Traces Everyone: Like This!* 3 (Tilburg Law Sch. Legal Studies Research Paper Series No. 03/2011, 2010), available at <http://ssrn.com/abstract=1717563>.

¹⁹ See *What They Know: A Glossary*, WALL ST. J., July 30, 2010, at 13.

²⁰ *Understanding Cookies*, MICROSOFT.COM, http://www.microsoft.com/resources/documentation/windows/xp/all/proddocs/en-us/sec_cook.mspx (last visited Apr. 14, 2012).

²¹ *Id.*

²² See Rob Pegoraro, *How to Block Tracking Cookies*, WASH. POST, July 17, 2005, at F7 (“I’ve had my browsers set to block third-party cookies for the past few years. I haven’t met the slightest inconvenience as a result.”).

²³ Mickey Alam Khan, *Rising Cookie Rejection Bites Into Metrics*, DIRECT MARKETING NEWS (July 11, 2005), <http://www.dmnews.com/rising-cookie-rejection-bites-into-metrics/article/88103/>.

²⁴ Chris Jay Hoofnagle, Jennifer King, Su Li, & Joseph Turow, *How Different are Young Adults From Older Adults When it Comes to Information Privacy Attitudes and Policies?* 5 (Working Paper, 2010), available at <http://ssrn.com/abstract=1589864>.

²⁵ Angwin, *supra* note 13.

²⁶ See Adrian Chen, *Use Facebook’s Targeted Ads to Find Out How Many People Are Into Kinky Sex in Any Workplace*, GAWKER (Jan. 13, 2012), <http://gawker.com/5875937/heres-how-many-facebook-employees-are-into-kinky-sex-according-to-facebook>.

mous cookie.²⁷ For instance, by signing up for some “free” offer, advertisers can link the information provided by the user to the existing cookies on that user’s machine.

Users may be able to avoid some tracking by blocking cookies, but that approach assumes that advertisers will respect individuals’ choices, and that advertisers will not employ alternative methods for tracking. Recall that in the telemarketing debate, technologies adopted by consumers to avoid sales calls were circumvented through clever new approaches by telemarketers.

Our research at Berkeley examines those assumptions through investigations into new and existing tracking technologies.

In 2009, we surveyed popular websites to empirically document how such sites were tracking users. Our study showed that advertisers do adapt to user cookie blocking through alternative trackers. In that study, we found widespread use of “Flash cookies.”²⁸ Flash cookies, technically called “local shared objects,” are files used by Adobe Flash developers to store data on users’ computers. Developers can use Flash cookies to store information about users’ preferences, such as volume settings for Internet videos, or they can be used to store unique identifiers for tracking users.

Our 2009 study elucidated the advantages of Flash cookies from a developer perspective, and documented that some advertisers adopted Flash cookies because they were relatively unknown, more difficult for consumers to delete, and more effective in tracking than standard or “HTTP” cookies.²⁹ We noted other tracking advantages of Flash cookies as well—they are more persistent than standard cookies, they can store 100 KB of information by default (standard cookies only store 4 KB), and they are stored in such a way that all browsers on a computer can access them, meaning that even if a user switches browsers, Flash cookies enable the user to be tracked.³⁰

Consumers can avoid some online tracking and aggregation by deleting their cookies. By deleting cookies, the user breaks the link between the

²⁷ Arvind Narayanan, *There Is No Such Thing as Anonymous Online Tracking*, CENTER FOR INTERNET & SOCIETY BLOG (July 28, 2011), <http://cyberlaw.stanford.edu/node/6701>.

²⁸ Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas, & Chris Jay Hoofnagle, *Flash Cookies and Privacy* (Working Paper, 2009), available at <http://ssrn.com/abstract=1446862>.

²⁹ For a discussion of the benefits of the Flash-cookies-based web tracking utility developed by United Virtualities, see Press Release, United Virtualities, United Virtualities Develops ID Backup to Cookies (Mar. 31, 2005), available at <http://web.archive.org/web/20050410041854/http://www.unitedvirtualities.com/UV-Pressrelease03-31-05.htm> (“United Virtualities, the leading innovator of creative marketing and technology solutions for the digital marketplace, today announced it has developed a backup ID system for cookies set by web sites, ad networks and advertisers, but increasingly deleted by users. UV’s ‘Persistent Identification Element’ (PIE) is tagged to the user’s browser, providing each with a unique ID just like traditional cookie coding. However, PIEs cannot be deleted by any commercially available anti-spyware, malware, or adware removal program. They will even function at the default security setting for Internet Explorer.”).

³⁰ For an in-depth discussion of the various advantages of different tracking vectors, see Sonal Mittal, *User Privacy and the Evolution of Third-party Tracking Mechanisms on the World Wide Web* (May 10, 2010) (unpublished honors thesis, Stanford University), available at http://www.stanford.edu/~sonalm/Mittal_Thesis.pdf.

identifier assigned to her computer and the tracking mechanisms on advertisers' servers. In the example of the "id" cookie above, if user 123456789 deletes the cookie, the server will assume that a new person has visited the site, and assign a cookie with another value, let's say 987654321.

Flash has the capacity to circumvent cookie deletion. Flash enables the "respawning" of cookies—that is, the ability to reinstate standard cookies that are deleted or otherwise lost by the user.³¹ Using Flash cookie respawning, advertisers can continue to track individuals uniquely even if the user deliberately tries to avoid web tracking. Thus the new user 987654321 can be matched with the older user 123456789. Flash respawning occurs subtly—the user is not alerted to the rewriting of the cookies and the reenabling of tracking.

These findings occur against a political backdrop where interventions to balance consumer privacy interests are described as "paternalistic." As one critic recently commented, do-not-track proposals "implement paternalistic judgments that subjects of targeted marketing cannot make proper judgments for themselves."³² This line of criticism suggests that privacy advocates and regulators think that online tracking is harmful or otherwise inappropriate for consumers. Merely giving consumers some legal or technical mechanism to block such tracking is paternalistic because it intervenes in the natural market ecosystem of consumers, websites, and advertisers.

History is repeating itself. In the telemarketing debate, sales callers used both policy and technology to force marketing upon consumers. Although consumers hated telemarketing, from a pragmatic perspective, sales calls worked. According to the telemarketing industry, hundreds of billions of dollars in sales were completed through sales calls annually.³³ The telemarketing industry wished to keep the sales channel open, even if many consumers found the practice unpopular. To enable choice, consumers needed legal rules that protected them against highly motivated and sophisticated actors determined to keep the phones ringing. In that context, government rules enabled choice, as opposed to marketplace approaches, which invalidated choice.

We challenge the notion that government intervention is paternalistic. Marketplace approaches effectively make it impossible to avoid tracking online. Our current work shows that advertisers are using technologies that consumers are not familiar with, specifically in order to override consumers' preferences. Behavioral advertising—and the tracking that goes with it—is the offer you cannot refuse, not necessarily because you are tempted by it, but because sophisticated, market-dominant actors control the very platforms you use to access the web. Advertisers are so invested in the idea of a per-

³¹ Flash cookie respawning is sometimes referred to as cookie "backups," or reinstating cookies.

³² Thomas R. Julin, Sorrell v. IMS Health *May Doom Federal Do Not Track Acts*, 10 BNA PRIVACY & SECURITY LAW REP. (PVL) 1262 (2011).

³³ *Telemarketing Industry & Stats*, DMA, <http://www.the-dma.org/telemarketing/telemarketingfaq.shtml> (last visited Apr. 18, 2012).

sonalized web that they do not think consumers are competent to decide to reject it.

This article proceeds in three parts. First, we discuss the landscape of research on Internet tracking and the findings from our 2011 study. The landscape's contours show that there has been increasing interest in studying how companies track consumers online. These studies show that there is much more tracking now than at the inception of the commercial web, among a smaller group of tracking companies. Second, we turn to the privacy problems raised by this tracking. Increased tracking means that a small number of companies have a window into most of our movements online. Inferences derived from that tracking can be sold to third parties or used in ways that users find transgressive. Finally, we conclude by returning to the theme of consumer choice. Advocates of market approaches vigorously object to consumer privacy rules, sometimes labeling them "paternalistic." We suggest that this objection more aptly applies to market approaches. Policy-makers can remedy this problem by enabling consumer choice and protecting those choices from technical circumvention.

RECENT RESEARCH ON WEB TRACKING

"Web privacy measurement"—the study of the methods employed by websites to track users—is a nascent field, with significant contributions developed by academic computer scientists and others interested in discovering tracking vectors and quantifying them.³⁴

The Electronic Privacy Information Center (EPIC) made the earliest attempts to enumerate privacy practices in a systematic fashion. In June 1997, it released *Surfer Beware: Personal Privacy and the Internet*, a survey of the top 100 websites.³⁵ EPIC found that only seventeen of the top 100 websites had privacy policies. Twenty-three sites used cookies, although it appears that EPIC used a "surface crawl" to detect those cookies, meaning that it only visited the homepage of the site and did not explore the site more deeply.

In May 2000, the Federal Trade Commission (FTC) released a survey of sites that detected third-party cookies. In its study, the FTC drew from two groups of websites: those with over 39,000 visits a month and a second sample of popular sites (ninety-one of the top 100). The FTC found that

57% of the sites in the Random Sample and 78% of the sites in the Most Popular Group allow the placement of cookies by third par-

³⁴ See, e.g., Jonathan R. Mayer & John C. Mitchell, *Third-Party Web Tracking: Policy and Technology*, 33 IEEE SYMP. ON SECURITY & PRIVACY (forthcoming May 2012), available at <https://www.stanford.edu/~jmayer/papers/trackingssurvey12.pdf>; Berkeley Ctr. for Law & Tech., *May 2012 Web Privacy Measurement*, BERKELEY L., <http://www.law.berkeley.edu/12633.htm> (last visited Apr. 18, 2012).

³⁵ *Surfer Beware: Personal Privacy and the Internet*, ELECTRONIC PRIVACY INFO. CENTER (June 1997), <http://epic.org/reports/surfer-beware.html>.

ties. . . . The majority of the third-party cookies in the Random Sample and in the Most Popular Group are from network advertising companies that engage in online profiling.³⁶

There were few efforts at web privacy measurement until nine years later, when our 2009 report found cookies on ninety-eight of the top 100 websites.³⁷ In 2011, we found cookies on all top 100 sites. Thus, the web has had a dramatic change in web tracking, with a large shift in the prevalence of tracking cookies.

Recent research has also focused on other aspects of web tracking, including website “leakage,” concentration of tracking companies, and new vectors for tracking. This Section discusses that research.

The Problem of Information Leaking to Third-Party Websites

In recent years, there has been great interest in online tracking. In their ongoing investigations of web privacy issues, Bala Krishnamurthy, Konstantin Naryshkin, and Craig Wills studied how personal information flows from first- to third-party sites. They found that a majority of the popular sites they analyzed “directly leak sensitive and identifiable information to third-party aggregators.”³⁸

Practically, this means that the design of these sites is such that personal information entered by the consumer is exposed to third-party advertising companies. For instance, users entered their email addresses in order to sign up for a newsletter; in processing the request, the website would make the email addresses available to third-party advertisers, probably inadvertently. This would occur despite promises in privacy policies to not share data with such third parties.

In a multiple-year study of 1200 websites, Krishnamurthy and Wills found greater collection of information about users from an increasingly concentrated group of tracking companies.³⁹

Krishnamurthy and Wills also describe how third-party tracking sites disguise themselves as first parties. We call it “DNS aliasing,” a practice where “what appeared to be a server in one organization (e.g. w88.go.com) was actually a DNS CNAME alias to a server (go.com.112.2o7.net) in another organization (Omniure).”⁴⁰ Practically, this renders consumers’ attempts to block third-party cookies ineffective because first parties (such as

³⁶ FED. TRADE COMM’N, PRIVACY ONLINE: FAIR INFORMATION PRACTICES IN THE MARKETPLACE 21 (2000), available at <http://www.ftc.gov/reports/privacy2000/privacy2000.pdf>.

³⁷ Soltani et al., *supra* note 28.

³⁸ Balachander Krishnamurthy, Konstantin Naryshkin & Craig E. Wills, Privacy Leakage vs. Protection Measures: The Growing Disconnect 10 (May 2011) (unpublished manuscript), available at <http://www.cs.wpi.edu/~cew/papers/w2sp11.pdf>.

³⁹ Balachander Krishnamurthy & Craig E. Wills, *Privacy Diffusion on the Web: A Longitudinal Perspective*, in PROCEEDINGS OF THE 18TH INTERNATIONAL WORLD WIDE WEB CONFERENCE 541 (2009), available at <http://www2009.eprints.org/55/1/p541.pdf>.

⁴⁰ *Id.* at 543.

go.com in the above example) have built their servers to allow third parties (Omniure in the above example) to instate cookies as first parties.

Krishnamurthy and Wills found a doubling in such DNS aliasing: “[T]he percentage of first-party servers with multiple top third-party domains has risen from 24% in Oct’05 to 52% in Sep’08. . . . This increase is significant because it shows that now for a majority of these first-party servers, users are being tracked by two and more third-party entities.”⁴¹

Through decoding aliased domains, Krishnamurthy and Wills found that third-party trackers were becoming more concentrated. Sampling from five periods, concentration grew from forty percent in October 2005 to seventy percent in September 2008. Further, they found that “[t]he overall share of the top-five families—Google, Omniure, Microsoft, Yahoo and AOL—extends to more than 75% of our core test set with Google alone having a penetration of nearly 60%.”⁴² This means that a small number of companies can track much of what users do online.

The Move From Standard Cookies to New Tracking Vectors

Cookies have been the standard technology for uniquely enumerating Internet users. But in recent years, advertisers have adopted new methods that are more difficult for users to detect and block. At the same time, researchers have identified these technologies and explained how they implicate privacy.

These techniques fall into two categories. The first, explained below, primarily relies upon writing files to the user’s computer. These files contain unique identifiers that advertisers can use to track individuals as they use the web. These technologies are ETags, Flash cookies, HTML5 local storage, and Evercookies.

In the second, advertisers rely upon attributes of the user’s computer. For instance, the advertiser may detect that a user is employing a certain kind of web browser and has certain kinds of fonts installed on the computer. By combining these attributes, advertisers can “fingerprint” the user’s computer, and then rely upon the fingerprint to uniquely track the user across sites.

ETags

Researchers have also focused upon new vectors for tracking. As early as 2003, Dean Gaudet described unique user tracking through using “ETags,” a feature of the cache in browsers.⁴³ The cache helps speed up the

⁴¹ *Id.* at 546.

⁴² *Id.* at 549.

⁴³ See Dean Gaudet, *Tracking Without Cookies*, ARCTIC (Feb. 17, 2003), <http://www.arctic.org/~dean/tracking-without-cookies.html> (“[O]ther than cookies, there’s typically only one other type of data a webserver can cause a browser to store on its local hard-drive—cacheable web content. [T]his technique attempts to get the browser to store unique id

user's web browsing experience by detecting whether the user has previously visited a webpage. If she has, the browser can show the user a saved version of the site, rather than requesting another copy from the server. Advertisers can use this mechanism to store unique identifiers on the user's machine. Such enumeration is very inconvenient to block, and if users did so, they would substantially slow their Internet browsing.

Flash Cookies

In particular, recent research has focused upon the privacy implications of plug-ins such as Flash. As noted above, Flash enables developers to place small files on users' computers that can store information and identify users as they use the web. Recall that Flash also enables websites to "respawn" or back up standard web cookies that the user deletes. Flash cookies are installed "outside the browser," meaning that even if users switch their web browser (for instance, from Internet Explorer to Firefox), websites can still access the same Flash cookies.

As early as 2006, Corey Benninger noted that Flash cookies could be set without any visible sign to the user that Flash was running.⁴⁴ As Sipior, Ward, and Mendoza recently noted, addressing this risk by simply disabling Flash is unrealistic from a user perspective because an enormous amount of web content is delivered in formats requiring a plug-in.⁴⁵ This makes plug-ins such as Flash a very attractive technology for user tracking. Most users have Flash installed on their computer, and the price of forgoing the technology means that the user will not be able to view many web videos. Advertisers then can place Flash programs on websites to write files on users' computers and track them across websites.⁴⁶

Aleecia McDonald and Lorrie Faith Cranor of Carnegie Mellon University have conducted the most important research relevant to Flash cookies.⁴⁷ Their 2011 investigation of Flash cookies found a dramatic decline in their use. For instance, McDonald and Cranor found that only twenty of top 100 websites used Flash cookies (down from fifty-four in our 2009 study), and that only two sites respawned using Flash cookies.⁴⁸ McDonald and Cranor were also careful to determine whether Flash cookie values were unique or

information in its cache in a manner which will be communicated to the server at a later date. ([T]he later communication will be via a GET If-Modified-Since, or If-None-Match.)").

⁴⁴ COREY BENNINGER, *AJAX STORAGE: A LOOK AT FLASH COOKIES AND INTERNET EXPLORER PERSISTENCE 2* (2006), available at <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.128.2523> ("In fact, it would be difficult to reliably detect if an application were using flash cookies.").

⁴⁵ See Janice C. Sipior et al., *Online Privacy Concerns Associated With Cookies, Flash Cookies, and Web Beacons*, 10 J. INTERNET COM. 1, 4 (2011).

⁴⁶ See *id.* at 10–11.

⁴⁷ ALEECIA M. McDONALD & LORRIE FAITH CRANOR, *A SURVEY OF THE USE OF ADOBE FLASH LOCAL SHARED OBJECTS TO RESPAWN HTTP COOKIES* (2011), available at <http://www.casos.cs.cmu.edu/publications/papers/CMUCyLab11001.pdf>.

⁴⁸ *Id.* at 14.

not—eight of the top 100 sites had Flash cookies that were not unique and thus probably not used to track individuals.⁴⁹

The McDonald and Cranor team used different methods from our 2009 study. They visited the landing page of the top 100 sites, plus a selection of random sites. We thought that this approach did not adequately simulate typical use of websites. Users typically visit a homepage and then click on links in order to see other content, such as news stories or social media profiles. Thus, in our current and 2009 studies, we made ten arbitrary clicks on the same website, to simulate a user session. Because McDonald and Cranor only visited the homepage of websites, they acknowledged that their scan represented a “lower bound” in the enumeration of Flash cookies.⁵⁰

McDonald and Cranor also emphasized the normative implications of Flash use for user tracking. The use of Flash cookies for unique user tracking is problematic because it is functionally equivalent to respawning, or backing up and reinstalling standard HTTP cookies. Flash cookies were developed in 2000, with the release of Adobe Flash version 6. But users are still not generally aware of them, and until 2011, web browser controls did not include settings to control Flash cookies. Whether or not a website respawns HTTP cookies deleted by the user, if it uses Flash cookies, it can uniquely and persistently track individuals even in situations where the user has taken reasonable steps to avoid online profiling.

In 2009, we focused on the practice of respawning and, in the process, failed to adequately articulate this problem elucidated by McDonald and Cranor. In fact, our rhetorical choice to use the term “Flash cookies” seems to have backfired. We referred to Flash local shared objects as “Flash cookies” in order to make the issue more accessible to policymakers and others. But this caused many to speciously argue that Flash cookies are really no different than HTTP cookies.

Local shared objects are not just like HTTP cookies—they are far more flexible than HTTP cookies, and the infrastructure that gave rise to them enabled an obscure and persistent tracking mechanism that largely is still in place today. Table 1 below sets forth the basic differences among the cookies analyzed in this paper.

HTML5 Web Storage

Flash cookies may be just an ephemeral approach for web tracking. Web programming is advancing, and the new standard for writing websites is “HTML5.” HTML5 has many advantages for website design, especially for mobile devices.

HTML5 enables a new kind of cookie known as “HTML5 local storage.” HTML5 storage offers many advantages for website developers over ordinary cookies. Like Flash cookies, HTML5 local storage is more persis-

⁴⁹ *Id.* at 12.

⁵⁰ *Id.* at 8.

tent than standard web cookies. Standard cookies expire by default when the user closes her browser. In order to make standard cookies persistent, developers must use complex programming. HTML5 data are persistent until affirmatively deleted by a website or user. Storage size is important too. While Flash cookies have a default limit of 100 KB, standard cookies store just 4 KB, compared to 5 Mb for HTML5 storage.⁵¹

HTML5 local storage is a more universal storage mechanism than Flash cookies because it does not require that users have plug-ins, such as Flash, installed on their computers. Increasingly, device manufacturers such as Apple are releasing products without support for Flash. Thus we expect to see less reliance on Flash as a technology for tracking users.

Table 1: Key Characteristics of HTTP Cookies, Flash Cookies, and HTML5 Storage

| | HTTP Cookies | Flash cookies | HTML5 storage |
|-------------------|---|--------------------------------------|----------------------|
| Storage | 4 KB limit | 100 KB by default | 5 Mb by default |
| Expiration | Deleted by default when the browser is closed | Permanent by default | Permanent by default |
| Access | Only by one browser | By multiple browsers on same machine | Only by one browser |

Several commentators have highlighted the privacy risks that HTML5 presents. Others have argued that HTML5 has great potential to enable more privacy-preserving advertising models.⁵²

However, to our knowledge, no one has performed a survey of HTML5 privacy practices. Thus, as part of our update to our original Flash cookies investigation, we also captured and analyzed HTML5 data.

The Evercookie

Samy Kamkar has created the “Evercookie,” a tracking mechanism that uses Flash storage, standard cookies, and a variety of other techniques (including ETags) in order to make it resistant to user attempts to delete cookies and other unique identifiers.⁵³ The Evercookie approach relies upon redundancy. If one identifier—for instance, the cookie—is deleted, other

⁵¹ BRUCE LAWSON & REMY SHARP, INTRODUCING HTML5 142–43 (2011).

⁵² See generally Arvind Narayanan & Jonathan Mayer, Presentation at Workshop on Internet Tracking, Advertising, and Privacy (July 22, 2011) (on file with author); *The Do Not Track Cookbook*, DO NOT TRACK, <http://donottrack.us/cookbook>.

⁵³ See Samy Kamkar, *Evercookie*, SAMY KAMKAR (Sept. 20, 2010), <http://samy.pl/evercookie/>; see also Tanzina Vega, *New Web Code Draws Concern Over Privacy Risks*, N.Y. TIMES, Oct. 11, 2010, at A1.

mechanisms can reinstate the cookie. Because it is redundant, the Evercookie is difficult to eliminate.

Fingerprinting

The second category of tracking mechanisms enumerates users by recording attributes of the user's computer. Peter Eckersley has demonstrated the privacy risks associated with browser fingerprinting, where server-side programs can query a browser for enough information about its configuration to identify a computer.⁵⁴ For instance, advertisers may observe the type of browser a person is using, the fonts they have installed, and the plug-ins they have installed on their computer. These discrete attributes may not be unique, but in combination they tend to uniquely identify users.

Because the mechanism occurs on servers, fingerprinting may be difficult for users to detect and block. To avoid fingerprinting, one must disable key functionality of websites, such as JavaScript and Adobe's Flash. As of 2010, BlueCava, an online tracking company, claimed to have fingerprinted 200 million devices.⁵⁵

METHODS

We largely followed the methods of our 2009 paper. We crawled the top 100 U.S. websites based upon QuantCast.com's ranking of July 13, 2011. The data collection occurred on July 21, 2011. Using Firefox version 5, we visited each site and then made ten arbitrary clicks on each website. We collected standard cookies, HTML5 local storage, and Flash cookies from these crawling sessions. We never "signed in" to a website in this process.

Because of the dynamic nature of websites and online advertising, any given survey may produce different advertisements and correspondingly different standard cookies, HTML5 local storage, and Flash cookies. Thus, our snapshot may differ from another user's experience. However, we feel that this provides a reasonable sample for study.

We used several methods to detect and confirm respawning cookies, including manually deleting standard cookies to see whether they reappeared. We also manipulated the identifiers inside cookies to see whether those same identifiers would later appear in other cookies.

⁵⁴ See Peter Eckersley, *How Unique Is Your Web Browser?*, 6205 LECTURE NOTES COMPUTER SCI. 1 (2010).

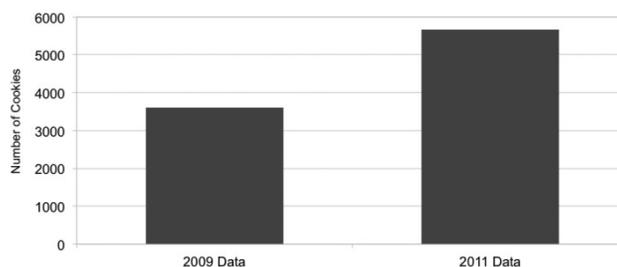
⁵⁵ Julia Angwin & Jennifer Valentino-DeVries, *Race Is On to 'Fingerprint' Phones, PCs*, WALL ST. J., Nov. 30, 2010, at A1.

RESULTS

A Dramatic Increase in the Use of Standard Cookies

We detected standard cookies on all top 100 websites. In total, we detected 5675 standard cookies. This is dramatically higher than the 3602 we detected in 2009. Twenty sites placed 100 or more cookies, including seven that placed more than 150 (wikia.com, 242; legacy.com, 230; foxnews.com, 185; bizrate.com, 175; drudgereport.com, 168; myspace.com, 151; and time.com, 151).

NUMBER OF HTTP COOKIES IN 2009 AND 2011



The most frequently appearing cookie names were: uid, id, PREF, __utmz, __utma, __utmb, and UID. Many of these cookie names are commonly associated with user tracking. For instance, cookies named “__utma” are used by Google for identifying unique visitors.⁵⁶ “[U]id” and “id” typically refer to unique identifier and identifier, respectively.

Most Cookies Were Placed by Third-Party Hosts—Typically Tracking Companies

First-party cookies are placed by the website that the consumer is visiting, for instance, nytimes.com. Third-party cookies are placed by advertisers and others who are in partnership with the first party, for instance, DoubleClick. We found that most cookies—4915 of them—were placed by a third party. We detected over 600 third parties among the 4915 third-party cookies. This suggests that there are approximately 600 companies involved in tracking users online.

Google had cookies on eighty-nine of the top 100 sites; the company’s ad tracking network, doubleclick.net, had cookies on seventy-seven. Combined, Google has a presence on ninety-seven of the top 100 websites. This includes popular government websites such as usps.com, irs.gov, and nih.gov.

⁵⁶ *Cookies & Google Analytics*, GOOGLE DEVELOPERS, <https://code.google.com/apis/analytics/docs/concepts/gaConceptsCookies.html> (last visited Apr. 15, 2012).

This means that the browsing that one does on irs.gov for tax information and advice, or on nih.gov for information about health conditions, is silently being tracked by Google. Google is free to make inferences from the use of these sites and to combine those observations with data it obtains from tracking users on other sites.

Only microsoft.com, ups.com, and wikipedia.org lacked some type of Google cookie.

Other third-party trackers with a strong presence in the top 100 included scorecardresearch.com (sixty-one) and atdmt.com (fifty-six). Among the top 100 sites, wikia.com, legacy.com, foxnews.com, drudgereport.com, and bizrate.com hosted the most cookies from third-party domains.

The Use of Flash Cookies Declined

We found 100 Flash cookies on the top 100 sites, down from the 281 we found in 2009. These Flash cookies appeared on thirty-seven sites, down from the fifty-four sites we found in 2009.

Recall that Flash cookies can store much more information than a standard cookie. We found that some sites coded a large amount of information into their Flash cookies. For instance, MTV.com had eight Flash cookies, one of which stored over 140 values. This means that MTV.com's eight Flash cookies store about the same amount of information as 140 standard cookies.

Two sites had shared values between Flash cookies and HTTP cookies: hulu.com and foxnews.com. In the case of foxnews.com, the value was shared in HTML5 local storage as well. Shared values are a signal that the website is using multiple technologies to track users. As explained above, this means that if a user deletes a single cookie, one of the other technologies may serve as a backup and reinstate the cookie.

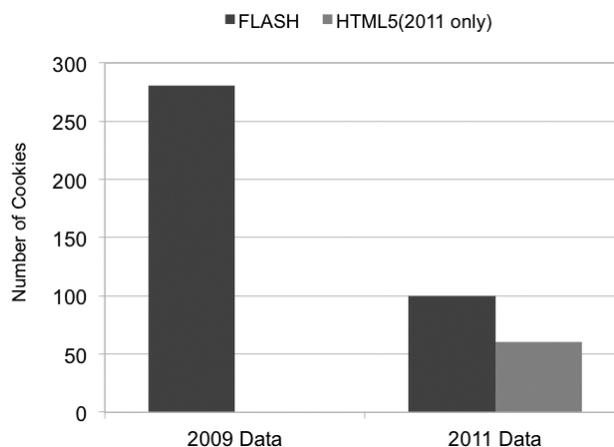
HTML5 Storage Was Present on Some Top Websites

HTML5 local storage is a relatively new technology, and we did not scan sites for its presence in 2009. We included it in our 2011 study and found that seventeen of the top 100 sites were using HTML5 local storage.

Like Flash cookies, HTML5 local storage can accommodate more information than a standard cookie. We found that the seventeen sites with HTML5 local storage stored the equivalent of sixty standard cookies.

We found shared values among HTML5 local storage and standard cookies in several cases. This suggests that the sites were using HTML5 to back up standard cookies. Twitter.com, tmz.com, squidoo.com, nytimes.com, hulu.com, foxnews.com, and cnn.com had such matching values. In most of these cases, the matching value was with a third-party service, such as meebo.com, kissanalytics.com, and polldaddy.com.

NUMBER OF FLASH AND HTML5 COOKIES IN 2009 AND 2011



Hulu.com Respawned Cookies With Flash and HTML5, Using Two Different Methods

In 2009, we reported that a QuantCast cookie was respawned on hulu.com. After our 2009 paper, QuantCast executives contacted authors Hoofnagle and Soltani almost immediately, and quickly acted to change the behavior of their service in order to prevent respawning.⁵⁷

Nevertheless, hulu.com, QuantCast, and other companies were sued for the practice.⁵⁸ Plaintiffs invoked a number of state and federal statutes to argue that respawning cookies was a form of computer hacking. The case settled in 2011.⁵⁹ In a summary of Flash cookies filed with the court, companies such as Hulu claimed that they did not know that third-party services provided by QuantCast and Clearspring tracked users through Flash.⁶⁰ This assertion effectively shifted the blame from consumer-facing websites to the third-party tracking companies involved. In the settlement, QuantCast and Clearspring explicitly promised not to respawn cookies using Flash, or to

⁵⁷ See Ryan Singel, *Online Tracking Firm Settles Suit Over Undeletable Cookies*, WIREDCENTER (Dec. 5, 2010, 2:02 AM), <http://www.wired.com/epicenter/2010/12/zombie-cookie-settlement/>.

⁵⁸ *In re Quantcast Adver. Cookie Litig.*, No. 2:10-cv-05484-GW-JCG (C.D. Cal. June 13, 2011).

⁵⁹ Singel, *supra* note 57.

⁶⁰ Joint Submission of Supplemental Information Regarding Plaintiffs' Motion for Preliminary Approval of Class Action Settlement at 13, *In re Quantcast Adver.*, No. 2:10-cv-05484-GW-JCG ("The Customer Defendants, on their own behalf and on behalf of their corporate parents and affiliates, have represented to Quantcast and Clearspring that the Customer Defendants were unaware that LSOs were being used to store information regarding consumers who accessed their websites and web content. Quantcast and Clearspring do not dispute that representation and, to the extent of their knowledge, information, and belief, adopt and incorporate it here.").

use Flash as an alternative to HTTP cookies for tracking purposes.⁶¹ These obligations did not apply to consumer-facing websites, such as hulu.com.

We found two different methods of cookie respawning on hulu.com. As explained above, these methods back up standard cookies, thus preserving the ability of advertisers to track users even if they delete their cookies.

First, hulu.com used standard Flash respawning to reinstate a standard cookie with the key “guid,” mirroring a Flash cookie with the key “computerguid.” There are two important points to raise about this: Unlike the situation in 2009, where a third party respawned the cookies, this use of Flash is in-house at hulu.com. And while Adobe points out that local storage enables the delivery of rich content, hulu.com’s use of Flash appears to fall into the category of unique user tracking condemned by Adobe. Adobe argues that such uses of Flash should be subject to express user consent.⁶²

Second, we found first-party standard and HTML5 cookies respawned on hulu.com through a service hosted at kissmetrics.com. This respawning employed ETags to back up the cookies. To our knowledge, this is the first demonstration of this ETag tracking “in the wild.”

ETag tracking and respawning is particularly problematic because the technique generates unique tracking values even where the consumer blocks standard, Flash, and HTML5 cookies. In order to block this tracking, the user would have to clear the cache between each website visit. Even in private browsing mode, ETags can track the user during a browser session. The script for this function, hosted at <http://doug1izaerwt3.cloudfront.net>, included other code that indicated its author was aware of tracking and the risk of data collection about the user. For instance, it included a function to detect the collection of information that credit card companies require websites to control more carefully.

On June 30, 2011, hulu.com updated its privacy policy to include disclosures surrounding Flash cookies.⁶³ This update appears to have been driven by obligations in a recent settlement from a lawsuit sparked by our 2009 paper. This settlement required any consumer-facing website to include, “in its online Privacy Policy, a disclosure of its use of LSOs [Flash cookies] and a link to at least one website or utility offering users the ability to manage LSOs, if such website or utility is available.”⁶⁴ This policy was in effect when we scanned popular sites for cookies and other tracking technologies.

In the June 30, 2011 policy, hulu.com included a link to Adobe’s Flash cookie manager and disclosed that it used Flash cookies, but downplayed

⁶¹ See Settlement Agreement § 4.19, *In re Quantcast Adver.*, No. 2:10-cv-05484-GW-JCG.

⁶² See ADOBE SYS. INC., COMMENTS FROM ADOBE SYSTEMS (2010), available at <http://www.ftc.gov/os/comments/privacyroundtable/544506-00085.pdf>.

⁶³ See *Privacy Policy*, HULU (June 30, 2011), <http://www.hulu.com/privacy> (“We have updated our Privacy Policy to provide more details about our information practices, including . . . our use of ‘Local Shared Objects’ in connection with Adobe’s Flash Player.”) (on file with the Harvard Law School Library).

⁶⁴ Settlement Agreement, *supra* note 61, § 4.20.4.

their potential for tracking: “Local Shared Objects are similar to browser cookies, but can store data more complex than simple text. By themselves, they cannot do anything to or with the data on your computer.”⁶⁵

We object to this last sentence in particular. While it is technically true that *by themselves* Flash cookies cannot do anything to the data on a user’s computer, in reality, Flash cookies never are used *by themselves*. It is the code accompanying Flash cookies that enables them to mirror other data and can be used to back up that data when deleted by the user.

The June 2011 hulu.com privacy policy does not mention respawning of any kind, and even claims: “You can configure your Internet browser to warn you each time a cookie is being sent or to refuse cookies completely. However, unless you accept cookies, you will not have access to certain Hulu Services.”⁶⁶

Hulu.com’s June 2011 policy also describes “Web beacons.” It is unclear whether this section of the policy describes kissmetrics.com cache respawning. The description would not lead an average user to understand that the cache was being used to undo cookie deletion.

We find it surprising that months after settling a suit involving unique user tracking through third parties, hulu.com moved Flash tracking and respawning in-house. Furthermore, the use of KissMetrics cache cookie respawning is very similar to the respawning we found in 2009—hulu.com used a third party to engage in tracking that users do not know about, cannot detect, and effectively cannot block.

THE OFFER YOU CANNOT REFUSE

Government interventions to protect consumer privacy are often framed as paternalistic. For instance, as noted above, one critic of Internet privacy legislation claimed that do-not-track proposals “implement paternalistic judgments that subjects of targeted marketing cannot make proper judgments for themselves.”⁶⁷ We argue that this criticism is misplaced. Modern privacy regulations do not make choices for consumers. Instead, they enable choices. A key example is the Telemarketing Do Not Call Registry, which enables consumers to easily opt out of telemarketing. Prior to the creation of the extremely popular Registry, consumers had few effective tools to address telemarketing intrusions.

The paternalism label is much more convincingly attached to the industry itself—the individuals pushing personalization even where consumers express preferences against it. The thrust of our 2009 and 2011 works strongly points to an industry that does not believe that consumers can make choices about web tracking on their own. They have restricted the freedoms of users by incorporating little-known technologies into websites for track-

⁶⁵ *Privacy Policy*, *supra* note 63.

⁶⁶ *Id.*

⁶⁷ Julin, *supra* note 32, at 1262.

ing, and by making these vectors resistant to choice mechanisms. Here, we describe how cookie respawning invalidates consumer choice, and more broadly how the KissMetrics ETag tracking system we discovered on hulu.com implicates privacy. The industry has used obscure technologies to circumvent user choices, and they have developed other techniques to undermine consumers' key tool for protecting privacy—the ability to withhold information from sites.

Using Technology to Circumvent User Autonomy

There are three principal privacy problems with the kind of cookie respawning we observed on hulu.com that was being performed by KissMetrics. First, users cannot fairly be said to have notice of these activities. The entire point of new tracking methods seems to be to ensure that users are ignorant of them. The websites that used Flash respawning and cache ETag tracking did not disclose those practices in their privacy policies.

Second, because these vectors are resistant to blocking, they rob consumers of choice. This undermines the advertising industry's representations about respecting individuals' choices and leaves consumers in a technical arms race with advertisers.

Marketers think that the benefits of being tracked outweigh consumer preferences, and thus have developed tools to frustrate cookie deletion and blocking. This attitude is probably best presented by the CEO of United Virtualities, a company that was a leader in promoting Flash cookies as a tracking technology:

All advertisers, websites and networks use cookies for targeted advertising, but cookies are under attack. According to current research they are being erased by 40% of users creating serious problems. . . . From simple frequency capping to the more sophisticated behavioral targeting, cookies are an essential part of any online ad campaign. PIE will give publishers and third-party providers a persistent backup to cookies effectively rendering them unassailable.

. . . .

The erasing of cookies threatens many cookie dependent server-side applications from registration to targeting to traffic counting PIEs are a cookie support product that ensures persistent identification of the users.⁶⁸

Our finding of cookie spawning in 2009 and 2011 is consistent with other researchers' findings that advertisers are using technology to invalidate consumer choice. Lorrie Cranor's team at Carnegie Mellon University recently found that thousands of websites were unblocking cookies that consumers ordinarily would not receive because of settings in their browser.

⁶⁸ Press Release, United Virtualities, *supra* note 29 (internal quotation marks omitted).

Cranor's team exposed an issue in the Microsoft Internet Explorer browser (MSIE). By default, MSIE blocks third-party cookies from any site that lacks a machine-readable privacy policy.⁶⁹ Website developers discovered that they could reenable cookie tracking by posting any kind of machine-readable policy—even an invalid one. Cranor's team found “thousands of sites using identical invalid CPs that had been recommended as workarounds for IE cookie blocking.”⁷⁰ In addition, they found that “98% of invalid CPs resulted in cookies remaining unblocked by IE under its 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.”⁷¹

Like MSIE, Apple's Safari browser blocks third-party cookies by default. Recently, Jonathan Mayer of Stanford University found that Google and other network advertisers had found a way to circumvent this cookie blocking.⁷² The method used by Google was particularly brazen—it opened a webpage invisible to the user and used a program to simulate the user clicking on it.⁷³ It is as if a Google engineer grabbed the user's mouse and clicked on a “track me” button while the user was not watching.

Our research and the work done by Cranor and Mayer show that advertisers are willing to use technology to circumvent settings on individuals' computers, leading to more tracking online.

Circumventing Selective-Revelation as a Check on Collection

The KissMetrics system presents another problem, in addition to a lack of notice and invalidation of choice. It allows companies to aggregate information about users in new ways that consumers are unlikely to understand. Consumers are aware of the sale of information to third parties. But the first-party tracking mechanism implemented by KissMetrics inverts the issue: How does tracking enable websites to *buy* information about their users from others?

The KissMetrics system uniquely enumerated users, and shared the same identifier with different first-party sites (for instance, the same identifier beginning with “GuTj890” enumerated our browsing sessions at Hulu, Spotify, Etsy, Spokeo, and Gigaom). This enabled these subscribers to Kiss-

⁶⁹ PEDRO GIOVANNI LEON ET AL., TOKEN ATTEMPT: THE MISREPRESENTATION OF WEBSITE PRIVACY POLICIES THROUGH THE MISUSE OF P3P Compact Policy Tokens 1 (2010), available at http://www.cylab.cmu.edu/files/pdfs/tech_reports/CMUCyLab10014.pdf.

⁷⁰ *Id.*

⁷¹ *Id.*

⁷² See Julia Angwin & Jennifer Valentino-DeVries, *Google's iPhone Tracking*, WALL ST. J., Feb. 17, 2012, at A1; Jonathan Mayer, *Safari Trackers*, WEB POL'Y (Feb. 17, 2012), <http://webpolicy.org/2012/02/17/safari-trackers/>.

⁷³ See Mayer, *supra* note 72 (“We discovered four advertising companies that surreptitiously submit a *form* in an invisible *iframe* and place trackable cookies in Safari: Google, Vibrant Media, Media Innovation Group, and PointRoll.”).

Metrics to share information about users with other sites. Any of the above-mentioned sites could share registration data about “GuTj890.”

This development is important because it breaks the trust model enabled by “selective revelation.” A bedrock privacy principle holds that information should be collected through fair means and, where possible, with the informed consent of the data subject.⁷⁴ This allows the individual to be directly involved in data collection practices.

Advocates of market-based approaches to privacy have often echoed this principle in theory. They argue that consumers selectively reveal information to businesses they “trust.” For instance, user “GuTj890” may fear that hulu.com would send spam, and thus provide a throw-away email address when signing up. At the same time, “GuTj890” may trust etsy.com more, and provide more personal information and her main email address there. This selective revelation is the way that consumers choose in the marketplace. Companies with strong levels of trust and privacy thus prevail without the need for burdensome regulation, while companies with low trust values will fail from lack of consumer participation.

When firms buy information from others, they circumvent consumers’ efforts to engage in selective revelation. Consumers who share any information at all—even fake information—are at risk, because sites can match up cookies and discover real information that the user “trusted” to some other site. This risk is amplified where users are encouraged to authenticate in order to use a website’s services, such as popular music or video services like Spotify or Hulu.

In the offline world, marketers have tried similar tricks for some time. Recall the time when retailers asked consumers for their addresses (Radio Shack)⁷⁵ or phone numbers. Consumers complained about those practices, and California even enacted a law restricting the collection of personal information by retailers at the register in credit-card sales.⁷⁶

Some retailers responded to this law by developing more clever and obscure ways to elicit information from consumers. Retailers learned that by collecting the name of the consumer from a credit-card swipe and asking for a zip code, they could determine the home address of their customers. This was accomplished through a sophisticated data-matching product offered by data brokers. In fact, Acxiom markets a product to accomplish this linkage, and it is explicitly marketed as a tool to identify consumers without them realizing the privacy implications of providing the zip code.⁷⁷

⁷⁴ *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, ORG. ECON. COOPERATION & DEV. (Sept. 23, 1980), http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html.

⁷⁵ Greg Saitz, *Radio Shack Aims to Be Less Annoying*, STAR-LEDGER, Nov. 26, 2002, at 29 (noting that Radio Shack ended the practice in 2002).

⁷⁶ CAL. CIV. CODE § 1747.08 (West 2012).

⁷⁷ See ACXIOM, INFOBASE® DATA FOR SHOPPER RECOGNITION 1 (2006), available at <http://isapps.acxiom.com/AppFiles/Download18/AcxiomShopperRec-3262007115722.pdf> (advertising that the product helps retailers avoid “losing customers who *feel* that you’re invading

CONCLUSION: PUBLIC POLICY THAT PRESERVES CHOICE

“[W]hen you resent a thing, you seem to recognise it.”⁷⁸

When advertisers criticize privacy protection as paternalistic, we should remember the above-quoted observation of Cremutius Cordus—we resent the things that we recognize in ourselves. Government interventions in the direct marketing field have been choice enabling. The Do Not Track proposal itself would simply make it easier for individuals to decide not to be tracked. Market interventions, on the other hand, often force choices upon the consumer.

Those who argue that consumers can negotiate the nuances of privacy and tracking online assume that the online world is similar to the offline world. In the offline world, consumers can vote with their feet and, in most circumstances, leave a business they do not wish to frequent without it collecting data about the experience. In the online world, efficiencies in identification and aggregation alter the balance of power of the relationship between the consumer and the business. This has greatly benefitted consumers in enabling comparison shopping along factors that are visible, such as price. Privacy attributes of transactions are not as visible. Collectively, website owners have organized to track individuals as they traverse the web, and few popular websites forgo such tracking.

Advocates of market approaches rarely account for the various techniques that have been developed to prevent consumers from making a choice on privacy. The use of obscure tracking methods, data enhancement, cookie respawning, and the zip code re-identification schemes discussed above circumvent user choice. These techniques are often adopted explicitly to make the consumers think they are not being tracked or identified. This combination of disguised tracking technologies, choice-invalidating techniques, and models to trick the consumers into revealing data suggests that advertisers do not see individuals as autonomous beings.⁷⁹ Once conceived of as objects, preferences no longer matter and can be routed around with tricks and technology.

their privacy” (emphasis added)). In *Pineda v. Williams-Sonoma Stores*, the plaintiff alleged that the defendant engaged in very similar conduct:

Defendant . . . used customized computer software to perform reverse searches from databases that contain millions of names, e-mail addresses, telephone numbers, and street addresses, and that are indexed in a manner resembling a reverse telephone book. The software matched plaintiff’s name and ZIP code with plaintiff’s previously undisclosed address, giving defendant the information, which it now maintains in its own database. Defendant uses its database to market products to customers and may also sell the information it has compiled to other businesses.

246 P.3d 612, 615 (Cal. 2011).

⁷⁸ TACITUS, *THE ANNALS* (109), reprinted in *ANNALS AND HISTORIES* 1, 151 (Alfred John Church & William Jackson Brodribb trans., Everyman’s Library 2009).

⁷⁹ See JOSEPH TUROW, *THE DAILY YOU: HOW THE NEW ADVERTISING INDUSTRY IS DEFINING YOUR IDENTITY AND YOUR WORTH* 7 (2011) (arguing that marketers conceive of individuals as “targets” and “waste”).

Our survey of top websites suggests several interventions. First, on a basic level, consumers' manifestations of choice should not be circumvented. In this context, a policy that prohibited backing up cookies through respawning technologies and similar technical circumventions could enable consumer choice. If advertisers wished to condition access to services on tracking, they could. But to do so, they would have to have some dialogue with the consumer, rather than resorting to sneaky technical methods to obscure the tracking.

Second, information-forcing interventions could enhance consumer autonomy as well, particularly if focused on data enhancement, the practice of buying data from third parties about consumers that the consumer herself is unlikely to provide. For instance, if a user registered on a website and omitted details such as his or her income or geographic location, the site could inform the user that it would seek that information from some third-party data broker site. This approach would also suggest that data collection fields should no longer be marked as "optional" if the site will use enhancement to fill them.

Data enhancement circumvents consumers' most basic privacy-preserving strategy: the decision to not reveal certain information. Merely disclosing the presence of data enhancement is inadequate, as many companies already do so, albeit in vague ways. More effective would be "just in time" notices that informed the user that the site might buy information not provided by the user.

Simple, choice-preserving interventions could enhance individuals' decision making and create constructive dialogue among advertisers and consumers. Such approaches could enable more choice in the market.

Table 2: Key Results and Comparison With Other Studies

| | Soltani 2009 | McDonald 2011 | Ayenson Wambach et al. 2011 |
|---|---|---------------------------------------|---|
| Number of sites with Flash cookies (top 100 sites) | 54 | 20 | 37 |
| Total number of Flash cookies (top 100 sites) | 281 | Not reported | 100 |
| Sites with respawning (top 100 sites) | 6 | 2 | 2 |
| Number of websites with HTTP Cookies (top 100 sites) | 98 | 98 | 100 |
| Total HTTP Cookies set (top 100 sites) | 3602 | Not reported | 5675 |
| Sites with shared Flash/HTTP values on top 100 | 31 | Not reported | 2 |
| Total shared Flash/HTTP values on top 100 | 41 | 8 | 2 |
| Sample | Top 100 websites and six government sites | Top 100 websites and 500 random sites | Top 100 websites |
| Method | Visited homepage and then made 10 clicks on the same domain | Visited homepage multiple times | Visited homepage and then made 10 clicks on the same domain |