2

ONLINE BEHAVIORAL ADVERTISING

This chapter provides an introduction to the concept of online behavioral advertising (OBA). This domain is the backdrop against which experimental hypotheses in this dissertation are tested. Since I began writing this dissertation, OBA has grown from relative obscurity to notoriety. With the rising prominence of mobile browsing, cookie tracking is predicted to disappear within a few years. The surge of public awareness of OBA has recently led to new default settings in some browsers where third-party cookies are blocked by default.[1] Though particular issues discussed in this chapter will quickly stale, research in this dissertation should not as it relies upon theoretical tools and methods generalizing to phenomena not yet observed.

This chapter is organized in the following way. First I discuss the phenomenon of OBA and circumstances that led to its meteoric rise. Second, I introduce privacy issues in terms of how data is collected and used by third party advertisers. This has played an important role in shaping policy, leading to the design of interactions examined in this dissertation. Therefore, I follow by outlining the scope of user confusion as observed by privacy researchers. Finally, I conclude by summarizing the sorts of interactions behavioral advertising has engendered and speculate on what we may see in the future. Advertising is becoming increasingly becoming less of a broadcast art and more a personalized one. For this reason, we need to take a closer look at how advertisers may manipulate our thinking and beliefs during the course of interaction.

2.1 INTIMATE CAPITAL

A new market-driven ecosystem of targeted advertising has emerged, spanning the divide of Internet and brick-and-mortar

1 Mozilla and Microsoft have moved to this in the past year, while Apple Safari has blocked cookies by default for some time. Meanwhile, Microsoft, Google, and Apple are all exploring new technologies amenable to tracking across platforms to include mobile, gaming, and video services (D'Orazio, 2013).

business. This ecosystem is fueled by the unprecedented availability of data, algorithms, storage, and computing power. The forbidden fruit, and yet golden chalice, is the ability for the advertiser to know the target intimately. Of serious concern outside the advertising industry is that behavioral data may be combined with data owned by other content publishers or advertisers to actually identify a particular user. The more an advertiser knows the user — and not just as a demographic profile but as an individual — the more fine-grained the targeting. So it would not be surprising to learn that publishers and advertisers treat information about consumers as a form of capital — an "intimate capital... likely to be worth something to others" (Locke, 2010, p. 127). This section focuses on the sorts of data advertisers track and how this data is used in behavioral advertising. The availability of such data, coupled with advances in computing technology have given rise to new market economies struggling to maximize a return-on-investment (ROI) while the very measures defining ROI and ad effectiveness are in flux.

2.1.1 *Science in Advertising*

The quote below from a 1901 article in Publicity seems almost prescient by today's standards:

> The time is not far away when the advertising writer will find out the inestimable benefits of a knowledge of psychology. The preparation of copy has usually followed the instincts rather than the analytical functions. An advertisement has been written to describe the articles which it was wished to place before the reader; a bit of cleverness, an attractive cut, or some other catchy device has been used, with the hope that the hit or miss ratio could be made as favorable as possible. But the future must needs be full of better methods than these to make advertising advance with the same rapidity as it has during the latter part of the last century. And this will come through a closer knowledge of the psychological composition of the mind. The so-called 'students of human nature' will then be called successful psychologists, and the successful advertisers will be likewise termed psycholog-

> ical advertisers. The mere mention of psychological terms, habit, self, conception, discrimination, association, memory, imagination and perception, reason, emotion, instinct and will, should create a flood of new thought that should appeal to every advanced consumer of advertising space (as cited in Scott, 1904).

That these concepts should transfer so easily to the 21st century Internet is startling: what held true in advertising more than 100 years ago still rings true today.

In 1957, when Vance Packard wrote the runaway best seller The Hidden Persuaders, he exposed advertising not as a huckster bag of tricks, but as a subtle and calculated science with deep roots in psychoanalysis, sociology, and ethnographic anthropology. He honed in on a group of psychologists known as "the depth boys" who believed that to understand the consumer, you needed to find out what they really wanted at an unconscious level. In previous decades, advertisers had found little success interviewing and asking people what they wanted in a product. From *Advertising Age*, "In very few instance do people really know what they want, even when they say the do" (as cited in Packard, 1957, p. 37). According to Packard, what marketers and psychologists had been learning is that what people tell interviewers has little bearing on how they would actually behave. "What you are more likely to get, they decided, are answers that will protect the informants in their steadfast endeavor to appear to the world as really sensible, intelligent, rational beings" (Packard, 1957, p. 35). What advertisers ultimately learned from motivational research was how to find psychological and emotional levers that would generate an affinity to a particular product over a myriad of close alternatives — and how to use those levers to trigger an action to buy.

Recently, author Duhigg (2012b), *The Power of Habit: Why We Do What We Do in Life and Business*, described advances in the brain and behavioral sciences relating specifically to habit and learning. Habits, he says, are essentially a mechanism by which the brain encodes basic behaviors (in behavior chunks) in order to improve the efficiency of our brains. Marketers and advertisers, he says, leverage a three-step habit loop in order to reach "in market" customers and make sales. In this process, there is a *cue*, or trigger, for a particular pattern or behavior chunk. As Duhigg (2012b) relates, the basal ganglia

plays a central role in recalling patterns and acting on them as stored habits. A stored habit, or *routine*, which can be physical, mental or emotional, is linked to both the cue and subsequent *reward*. The reward itself plays a vital role in remembering such that a habit becoming "intertwined until a powerful sense of anticipation and craving emerges" (Duhigg, 2012b, p. 19). To no surprise, habit plays a central role in market strategy and advertising success. But the tools with which advertisers discover habits and influence habit loops, have become quite sophisticated.

2.1.2 *The Power of Data*

Discovering and tracking customer behavior is not new. Brick and mortar institutions have been doing this for the last century using purchase history to develop and refine models as well as coupons and mail to target specific customers. What has changed is technology. When Netscape engineer Lou Montulli invented the cookie and wrote the original specification, he was solving a state management problem in web applications (Kristol, 2001). Using cookies, websites could remember users and pages they had visited. Moreover, cookies were simply part of a browser protocol hidden from the user. Between 1994 and 2000, working groups were established in the Internet Engineering Task Force (IETF) centered on standards for browser state management: cookies had become the central mechanism by which all browsers managed state. By this time, America Online was a huge presence on the Internet and many businesses scurried to establish an online presence.

In 2000, Google began selling advertisements associated with search keywords (Google, 2010). Google also claimed the world's largest search index — exceeding one billion pages. By 2008, it had reached a trillion pages: it had solved a myriad of technical problems to reach this scale (Google, 2010). Chiefly, through innovations in computer hardware and processing, Google designed a network infrastructure capable of supporting the sharing of trillions of pages of web content, as well the means to aggregate and index that content. One such innovation was the means to process very large data sets using a model of distributed computing on large clusters of computers. Concurrently, advances were made in algorithm development, particularly in the area of data mining and machine learning.

Even as technical advancements have lead to innovation at the corporate level, so they have became more accessible to small companies and individuals. "I can get a hundred machines and I can sequence my own DNA for $150, or analyze the rise of Justin Bieber links across the Web over the last two years really quickly," says Hilary Mason, chief scientist at bit.ly (as cited in Woods, 2012). "If we didn't have that kind of commodity access to computer power and commodity access to analytics tools, we wouldn't be able to do the things we're able to do, and we certainly wouldn't be able to do them at startups with small budgets" (Woods, 2012).

Thus, in the past few years, there has been a surge of industry interest in the burgeoning field of "data science." According to Mason, data science is a blend of analytics ("counting things") and statistical machine learning algorithms. And a really good data scientist is a master at asking the right questions (Mason, 2012). In the world of advertising, its not surprising that the "right" questions are deeply concerned with understanding people.

2.1.3 *The Rise of Behavioral Advertising*

In the early years of the Internet, web advertisements were characterized by large banner ads and ad spaces reminiscent of print news and magazine media. A publisher sold space (inventory) on its site to advertisers, who filled that space with banner (or pop-up) ads. Cookies quickly changed the landscape of web advertising by making banners clickable and trackable, and this basic form of online advertising remains prevalent today. However, cookies also made it possible for advertisers to target ads to particular users, user agents, and devices. This has stimulated third-party (non-publisher) tracking and has led to changes in the economy and format of advertising, as well.

Though traditional advertising revenue via newspapers, television, radio and cable have been steadily dropping for several years, Internet advertising revenues continue to grow (IAB, 2012). In 2011, the Interactive Advertising Bureau (IAB) reported Internet advertising revenues of 31.74 billion, up 21.9% from 2010, and increasing at a linear rate from 2002 (IAB, 2012).

In terms of ad formats, search revenue in 2011 accounted for 46.5% of the total, with display advertising at 34.8%. Mo-

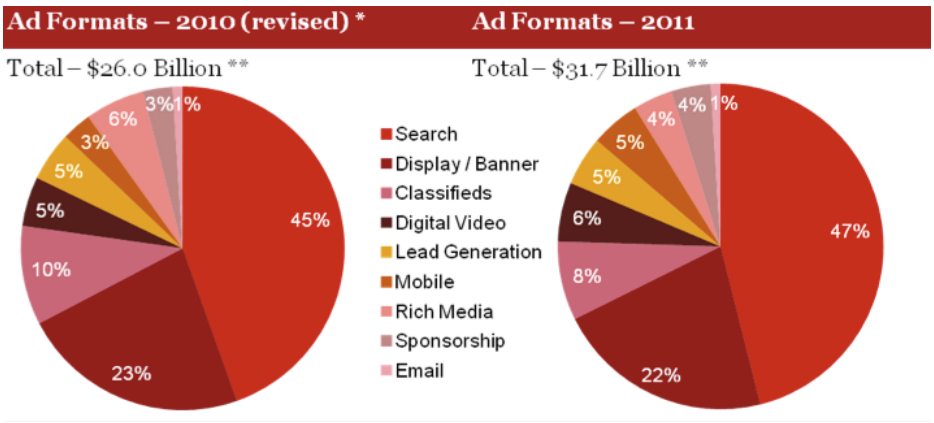

Figure 1: Revenue According to Ad Format (Image credit: IAB, 2012)

bile has emerged as a relevant category at 5%, while classifieds, lead generation, and email account for the remaining amount (IAB, 2012).

Because advertisers are most concerned with efficiently reaching high value consumers (those most likely to buy), targeting is important across all forms of advertising. On the surface, targeting is about breaking the market down into groups, or segments, which share certain characteristics. Typical market segments includes geographic (e.g., region, climate, urban / surburban / rural), demographic (e.g., age, gender, ethnicity, education), contextual (e.g., web page content) and behavioral (e.g., browsing history, content accessed over time). Ideally, advertisers target an advertisement to users most likely to be influenced by it. To accomplish this, advertisers attempt to collect and correlate information about users in order to best segment them.

Intrinsically, targeting seems linked with profiling and the potential for the accumulation, aggregation, and storage of personal data. Advertising networks claim that such information is anonymous, while consumers and policy makers struggle with the question of whether advertising networks should be entitled to assemble profiles and whether that should entail gaining user permission first. In this balance of this section of Chapter 2, we will put aside privacy concerns and instead focus on:

1. What it means to track users and why this is important;
2. The nature and use of behavioral data in advertising; and,

3. New market opportunities and potential effects.

2.1.4 *Tracking Users*

Defining "tracking" is a contentious exercise. The World Wide Web Consortium (W3C) Tracking Protection Working Group has been struggling with a definition for over a year (W3C, 2012). But, in its broadest sense, tracking is the collection, correlation, or transfer of data about Internet activities of a particular users, user agent, or device. Tracking may or may not be consensual and the users may or may not be aware of who is tracking, how tracking occurs, or what data is tracked. Tracking in the browser is most commonly associated by the use of HTTP cookies, but other means include IP address tracking, flash cookies (using the Adobe Flash plug-in), and web bugs or beacons (e.g., images retrieved from a third party website). Of potential future concern may be browser fingerprinting (Eckersley, 2010), though it does not seem that advertisers are using this technique now.

Publishers and advertisers alike track user data. Publishers, or first-party, tracking may include very fine-grained information about a user such email address, name, navigation behavior, etc. Personally identifiable information (PII) may be collected with user consent. And companies with both a brick-and-mortar and significant online presence likely combine information about what they know from shopping habits in the store with other data obtained online. More recently, even purely online businesses are able to tap offline data. For example, FaceBook has recently partnered with Datalogix to link online and offline consumer data (Reitman, 2012).

First party tracking may also be combined with data collected and aggregated by third party trackers. Third party trackers do not communicate directly with a user but monitor a user's actions in other ways. Generally, third party trackers are explicitly given consent to track by a website publisher. But they may also acquire user data via data exchanges and from Internet Service Providers (ISPs; e.g., via deep packet inspection such as that described in Sesto & Frankel, 2008) or other vendor sources. It is also possible for trackers to acquire data by taking advantage of security vulnerabilities and via information leakage (as described in Krishnamurthy & Wills, 2009, for example).

2.1.5 *The Nature and Use of Behavioral Data in Advertising*

In fact, how behavioral tracking is accomplished is anything but transparent. The image below in Figure 2 depicts the many HTTP requests called from the New York Times (NYT) homepage in November of 2012.

A large number of the domains included are third party trackers engaged in analytic or personalization services and advertising. But what makes behavioral tracking feel so insidious is that user activity on one site *can directly influence content displayed in completely un-related sites*.

For example, on October 5, 2012, after visiting the NYT, Washington Post, and Mozilla home pages, I generated a search query for "waterproof boots" on Google search. On October 6, 2012, I refreshed the NYT homepage and was presented an ad for Marc Jacob waterproof boots prominently displayed in the top banner ad space in Figure 3 below. This may or may not be co-incidence. There is no way to know.

A graph diagram generated from the Firefox Collusion extension in Figure 4 offers few clues about how my Google query might have caused this ad to be displayed on the NYT homepage.

Possibly, Google's advertising subsidiary DoubleClick may have been involved. When seen in the context of this collusion diagram, DoubleClick appears to know of visits to the Washington Post and NYT, but not what search query was made. Yet this is still possible. Business relationships outside of the communicative context also account for data transactions and exchanges.

Behavioral tracking concerns technologies and methods centered on capturing user data when users interact with web content. For example, information captured may include: user visits to a web site, specific page content, visit recency / frequency of visits, links clicked, searches, form data, and other interactive content. This data, plus metadata such as IP address may be combined to create a 'profile' linked to that user, browser user agent, or device. The goal of behavioral targeting is to show ads only to users of high value (those who are likely to buy the product) combined with a large number of opportunities to show such ads to that user.

An informal survey of agencies by the IAB suggests that behavioral advertising is widespread: up to 80% or more of advertising campaigns conducted in 2009 involved some form

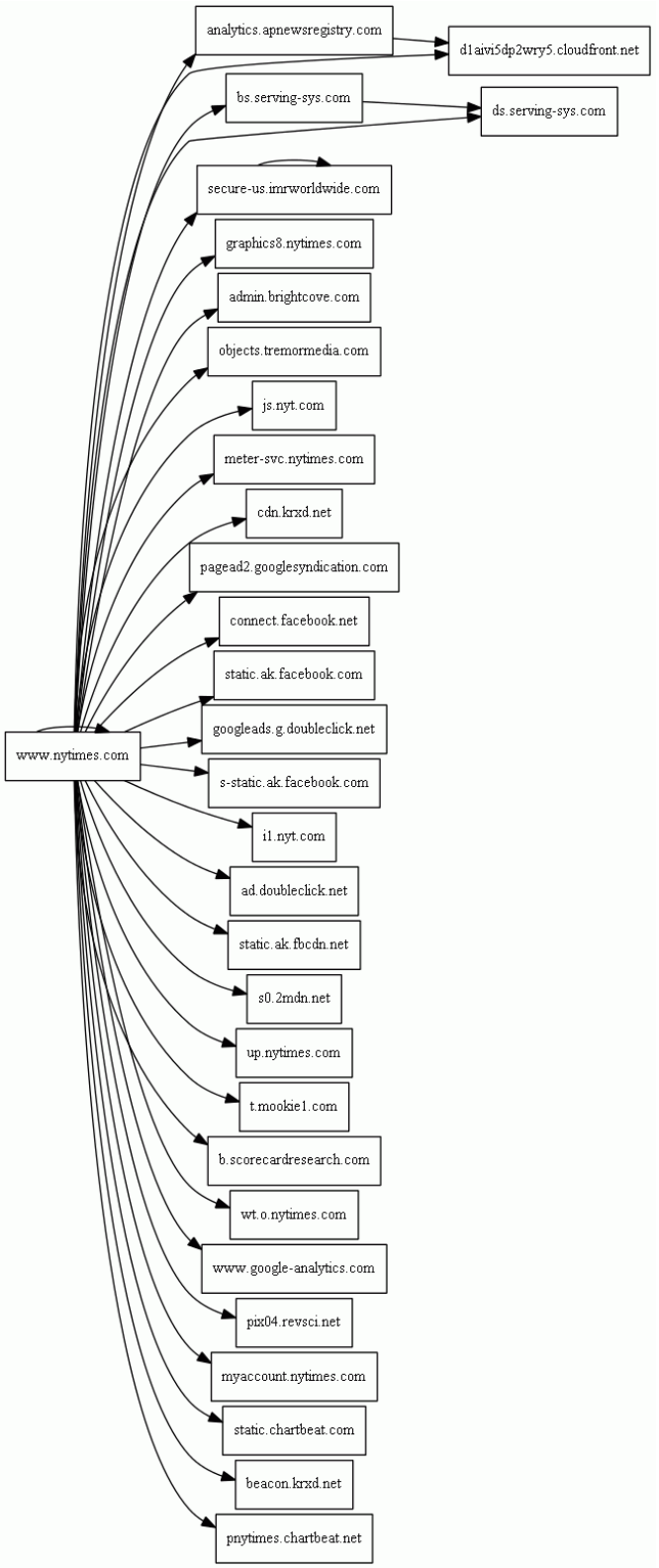

Figure 2: HTTP Requests from the New York Times Homepage

Figure 3: New York Times Homepage with Advertisement for Marc Jacob Boots

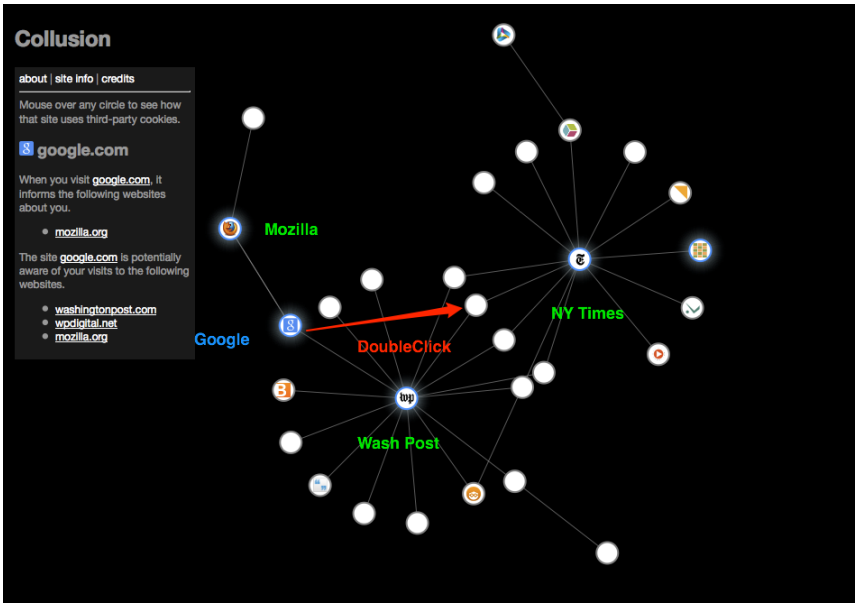

Figure 4: Firefox Collusion Diagram

of tracking (IAB, 2010). But until recently, we've had little real insight into how businesses, to include advertisers, used behavioral tracking.

In Feb 2012, The New York Times published an article written by Duhigg (2012a) telling the story of how Andrew Pole, a data scientist in Target's marketing analytics department, developed a model that could successfully predict whether a female customer was in the second trimester of pregnancy.

According to Duhigg (2012a), there are periods in life when routines and buying patterns change. Advertisers are quick to target customers during major life changes such as the purchase of a house or vehicle, or birth of a child. Retailers are particularly interested in acquiring and retaining new customers during such life changes when habits are more malleable. In the case of a birth event, parents will buy all sorts of items such as maternity clothing and prenatal vitamins — transitioning to baby care products soon after. Pole began analyzing consumer spending to identify 25 products, in aggregate, predictive of pregnancy and within a narrow window of time. As Duhigg (2012a) notes, "it's difficult to attribute how much behavior modeling contributed to Target's revenue, but between 2002 — when Pole was hired — and 2010, Target's revenues grew from $44 billion to $67 billion" (Duhigg, 2012a). Beyond anecdotal accounts such as pregnancy prediction by Target, there is relatively little information available about how businesses and advertisers profile users through online tracking.

Not all businesses have such fine-grained information as customer data to enrich predictive models. While Target may have access to name, address, demographics, geography, contact history and many other pieces of information (Pole, 2010), search engines typically have access to different sorts of data. In Yan et al. (2009), a team from China reportedly published the first academic, and empirical study, to address the question of the whether online behavioral targeting (OBT) can help in online advertising. In this paper, a basic OBT assumption was addressed. The basic assumption is this: "users who have a similar search or browsing behavior will have similar interests and thus have higher probability to click the same ad than the users who have different online behaviors" (Yan et al., 2009).

The team used seven days' ads click-through log data from a commercial search engine recording user search click be-

havior to include both web page clicks and ad clicks. They represented user behavior by page views and created a behavioral profile by considering all terms appearing in a user's query as previous behaviors. Both users and queries were represented as numerical vectors so that similarity between users could be easily calculated. Using simple clustering techniques, users were segmented according to behavior. Finally, user within-ad similarity was compared with user between-ads similarity. The result was that the within ads similarity, represented by user search behavior, was around ninety times larger than between ads similarity. Thus Yan et al. (2009) concluded that users with similar search behavior would indeed be distinguishable from other users and this could be used to predict the clickability of an ad.

Chen, Pavlov, and Canny (2009) described "massive improvements" in the ability to do offline training of OBT models. They noted, "behavioral data is intrinsically in large scale (e.g., Yahoo! logged 9 terabytes of display ad data with about 500 billion entries on August, 2008)" with a sparse click through rate (CTR) of about 0.05% (Chen et al., 2009). To build the entire 450 OBA category models from Yahoo! on fine-grained (ad clicks and search queries) at this scale would take about a week before the innovations they describe were implemented. By using a MapReduce learning algorithm, improved feature vector generation algorithm, better in-memory caching, and more efficient data structures and models they were able to reduce offline model building to about one day.

Scarcely a year later, Pandey et al. (2011) demonstrated the ability to include even more behavioral context to achieve, what they describe, as better results on live data. They created a general purpose model which allowed for optimizing three strategies of behavioral targeting: property (document context), segment (user demographics), and behavior (user past behavior). Each strategy was estimated to successively encode deeper context, thereby potentially improving the richness of the model. User events were modeled as an event stream of three types of events: pages visited, search queries, and graphics ads. By modeling in this manner, the research team examined the relative performance of specific event types (using extracted features), within specific temporal windows with respect to a variety of advertising campaigns of various types and sizes. In a live experiment spanning three months, they generated user models spanning eight weeks of user history.

For each of four ad campaigns, old models (like those described in the previous two paper described above) and new models received at least 1M ad impressions on a monthly basis. The conversion rate for new models was calculated to be considerably higher than for the older models, ranging from 57% to 264% higher.

Pandey et al. (2011) note, the problem of predicting clicks and predicting conversion have been split along clear lines of information available to publishers and information available to advertisers. There has been growing pressure from within the ad industry to change this. Moving to payment by conversion is an attractive option for advertisers because: 1) it helps to prevent click fraud; 2) can be used to analyze the effectiveness of the advertisement (was there an actual sales conversion and not just a click); 3) can prevent the user from being inundated by too many of the same ad; and, 4) and is compatible with re-targeting across sites (Pandey et al., 2011). Accordingly, advertisers have been recently more willing to share individual responses to ads to publishers since it facilitates the use of conversion-optimized models.

2.1.6 *New Markets*

The market of online advertising today is, in fact, much more complex than simple supply (publisher) and demand (advertiser) economics. This model, known as direct buy, was dominant in the early days of online advertising (Mayer & Mitchell, 2012). By the late 1990's advertising networks emerged allowing advertisers to place ads with many publishers — and publishers to work with many advertisers — through a common network. In such a model, it is easier for advertisers to target users along multiple dimensions (e.g., page context, demography, geography, behavior) for ad slots.

The advertising community as a whole sees targeting as an opportunity funnel (see Figure 5). Relating this back to psychology, advertisers are aware that to target effectively, then need to consider consumer:

1. Daily activities;
2. Online activities and habits; and
3. Research time, whether purchase is online or offline (Jupiter Research, 2010).

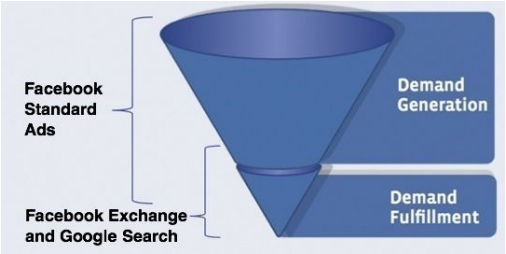

Figure 5: Facebook Re-targeting (Image credit: Constine, 2012)

There is a clear "in market" time of opportunity when users are researching purchase decisions. During this period of time, their behavior changes in predictable ways. For example, initially, the user may broadly review a general product space. Then move to a more comparative or winnowing down of choices. And finally, that consumer will look for a store and make a purchase.

Cookie re-targeting takes advantage of this purchase funnel. Recently, Facebook announced a new Facebook Exchange program, a real-time bidding ad system, where visitors with exchange party cookies can then be shown ads related to their web browsing when they return from those sites to Facebook (Constine, 2012). The basic idea is to re-target customers that visited a commercial site but did not purchase at that visit. This effectively positions Facebook from working in the broad part of the funnel (demand generation) to the more narrow part of the funnel (demand fulfillment; Constine, 2012).

To do this, Facebook doesn't share biographical information with advertisers, but takes cookies and combine them with Facebook data (Constine, 2012). The emergence of an *exchange market* — where buyers and sellers converge — is attributable to Google's DoubleClick Ad Exchange in 2009 (Duggal, 2012). The perceived value of an exchange over an advertising network is that the exchange operates on the behalf of any number of buyers, sellers, and middlemen alike. Krux (2012) estimates that real-time bidding exchanges now account for 40% online data collection. This is up from 0% three years ago.

Another sort of emergent market is a data market. *Data management platforms* serve as "a unified technology platform that intakes disparate first-, second-, and third-party data sets, provides normalization and segmentation on that data, and allows a user to push the resulting segmentation into live interactive channel environments" (O'Connell, 2011). Krux

(2012) found that more than 300 companies collected data on its users, up from 167 companies in 2010.

Many data collectors also piggyback on each other. By piggy-backing, one tracker invites other trackers to the trough. Because of this, publishers and users may not be aware of how much data is tracked on a given site (Angwin, 2012).

Not surprisingly, it is difficult to measure the effect of behavioral advertising. Referencing a study from Kazienko and Adamski (2007), Farahat and Bailey (2012) observe that click-through-rates (CTR) have declined from 3% to less than 1% (since the 1990s). CTR have traditionally been a measure of ad performance, though ultimately, advertisers prefer a more concrete measure such as sales conversions. In the Yan et al. (2009) experiment described earlier, ad CTR improvements through behavioral tracking were measured as high as 670% on the basis of their simple user segmentation strategy from search queries and clicked pages.

However, Farahat and Bailey (2012) question these ad CTR results stating that any study that naively looks at response lifts between targeted and un-targeted groups will greatly overestimate the effects of advertising since there is an inherent selection bias: the targeted users' behavior is likely to be highly correlated with the measured response ("click"). Farahat and Bailey (2012) ask, how do you know that the targeted segment isn't also the most clicky? Since what advertisers care more about are sales conversions than clicks, this is an important question.

In a field experiment on the Yahoo! front page, Farahat and Bailey (2012) try to weed out selection effects in terms of clickiness of the targeted population, brand interest, and category interest. They find strong evidence that user clickiness and brand interest are determinants of variation in CTR. This has bearing on the projected ROI of behavioral advertising for a given campaign. Furthermore, they question the ad CTR improvements of Yan et al. (2009), suggesting that the targeted lift of brand-related searches may be over-estimated by almost 1000%.

Farahat and Bailey (2012) conclude that advertisers would benefit from comparing how targeted and non-targeted populations respond to advertising in any particular campaign in order to gauge the cost effectiveness of their targeting strategy. "Given an industry average of a three times price premium for targeting, we might conclude targeting is more cost ef-

fective, but of course this depends greatly on the targeting product" (Farahat & Bailey, 2012). To date, relatively little is published about proprietary behavioral tracking model development nor the effectiveness of such models in actual ad campaigns. Nonetheless, the value of behavioral data is vigorously defended by marketers in public media as "the fuel the drives the Internet".[2]

What we really have little understanding is how to deal with the socio-technological effects of the connectedness of many different sorts and bits of data across vast networks of inter-connected participants. Currently, the heavy emphasis is on *who* to target. But there is mounting research to suggest that *when* (in the purchase funnel), *where* (on what site and how the ad should be placed), *what time of day*, *at what geographic location*, *how often*, and *how much* to pay (via auction strategy) are also important to advertisers. Furthermore, with the market moving away from CTR (click through rate) and CTM (cost per mile /"impression") towards a sales conversion rate (CVR), there will pressure for even greater information flow between producers and advertisers. Impediments to behavioral tracking become no longer technological, but sociological. And of course, these barriers will fall if both producers and advertisers believe they can profit.

In purely statistically-driven systems, the amount of data, not the algorithm, is king. However, there are many examples of where social network topology provides information that dramatically improves performance over pure volume. When Google created the PageRank algorithm it used page outbound links as implicitly encoded information about what page creators thought important. So what might the effect be when social networks such as Facebook harness vast networks of inter-connected consumers to the problem of behavioral advertising? How likely are we to behave like our friends? Or to what extent can they be used to influence us?[3] What if producers and advertisers find secure and effective means of sharing private information about users and their purchasing and other behavior? If the time has not yet come that trans-

2 This is an oft repeated phrase in the media and not attributable to a particular person or source.

3 A NYT article (published Oct. 13, 2012) reports on how the current presidential political campaign is heavily leveraging social and behavioral information to both reach "in market" voters as well as pressure them to vote. http://www.nytimes.com/2012/10/14/us/politics/campaigns-mine-personal-lives-to-get-out-vote.html?pagewanted=1&hp&_r=0

actional, behavioral, social, geographic location, and other offline third party data sources are not combined and used by business to more closely target individual consumers — that time will come shortly.

2.2 PRIVACY AND POLICY ISSUES

Every day, major newspapers across the globe report on some concern over digital data and privacy. Included in this global discussion are talk of law, policy, and technology — to include threats to privacy as well as proposed solutions. According to Gomez, Pinnick, and Soltani (2009), consumer reports and polls repeatedly show "overwhelming concern by users about the collection of personal information and behavioral profiling" (Gomez et al., 2009, p. 17). While website publishers typically offer a privacy policy to inform consumers what sorts of data may be collected and how that data may be used or shared, there is nonetheless much confusion and debate regarding what data is considered private and how it is handled. Though how personal data is collected and stored by first-party websites is in question, this section focuses more on the problem of data collected and used by third party trackers as well as public attitude toward this collection. However, in current privacy debates, there has been intrinsically little difference in the stance and attitude between large publishers and advertisers.

2.2.1 *The Internet is Free*

For some years, the Internet has been a free and open place. Access to services and information has been largely free. While many were initially skeptical about online commerce, convenience and ease won us over. With the advent of mobile technology and GPS, our applications became time and location aware. We can scan barcodes in stores and have applications tell us whether cheaper products are found close by. We embrace the wonder of connecting our virtual lives with our physical lives. But we feel uneasy, as well. The idea of being personally *surveilled* as we move about in the physical world is deeply unsettling — whether at home or in public. But to be monitored in a web browser leaves open the possibility that strangers will learn, in an intimate sense, *who we are*. They will do this by knowing what we read, what activ-

ities we pursue, where we live and travel, where we bank, where we work, what sorts of medical issues we research — and, naturally, what we buy. This information can be aggregated and stored by strangers — and then shared with other strangers: without consent, without review.

The delicate balance between public and private life may, in fact, be a relatively recent phenomenon. Locke (2010) examines closely the inherent tendencies of biological organisms to eavesdrop on others. Privacy is likened to a "regulatory process that serves to selectively control access of external stimulation to one's self or the flow of information to others" (Klopfer & Rubenstein, 1977, p. 53). One of the key functions of privacy is the withholding of information that might provide an opponent with a competitive advantage (Klopfer & Rubenstein, 1977).

In small, egalitarian societies, the ability to constantly monitor and eavesdrop "rendered trust unnecessary" (Locke, 2010, p. 74). Everyone shares access to the same resources. According to Locke, the pervasiveness of eavesdropping serves as a form of social control with an inhibitory effect on behavior. In more complex societies, where walls and buildings function as a sort of social technology, it has become less possible to overhear others and the notion of privacy has became mired with secrecy. At the same time, since one does not know everyone in the community, eavesdropping has become a means for knowing and understanding the private lives of others. Ultimately, "walls and population increases made it necessary to take in information about others when they were physically absent, relying on the representations — the gossip — of intermediaries" (Locke, 2010, p. 192).

2.2.2 *Privacy and Personally Identifiable Information (PII)*

Dan Solove, a George Washington University Law professor and author of *Understanding Privacy* (2008), frames privacy in terms of legal history and case law. The Fourth Amendment provides for the fundamental right to a "reasonable expectation of privacy" to the United States Constitution (as cited in Solove, 2008, p. 2). In Katz v. United States, 1967, the Fourth Amendment protection was extended to include Fourth Amendment protection to public places, such as private conversations in public phone booths (as cited in Solove, 2008, p. 22). However, the Supreme Court observed, "What

a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection" (as cited in Solove, 2008, p. 22).

In fact, it is the inherent difficulty in resolving what "privacy" is, that drove Solove to write his book. He notes "there is no overarching conception of privacy — it must be mapped like terrain, by painstakingly studying the landscape" (Solove, 2008, p. ix). This is particularly relevant when considering user browser-based interaction with websites over the Internet.

As Solove (2008) notes, there are literally hundreds of laws at state and federal levels to protect privacy:

> Congress has passed several dozen statutes to protect the privacy of government records, student records, financial information, electronic communications, video rental data, and drivers' records, among other things. Furthermore, privacy is recognized as a fundamental human right. According to the United Nations Declaration of Human Rights of 1948, "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation." (Solove, 2008, p. 3).

At particular risk or bits of information known as "personally identifiable information" (PIIs) typically associated with harms such as identify theft and mis-use. In May 2007, the Deputy Directory of the Office of Management and Budget (OMB) sent a memorandum on the subject of "safeguarding against and responding to the breach of personally identifiable information" noting,

> The term "personally identifiable information" refers to information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information. (OMB, 2007)

The purpose of this memorandum was to reinforce requirements of the Privacy Act of 1974 to both safeguard and establish rules of conduct for the handling of personally identifiable information. The particular challenge was in response to

the ease to which technology has made it possible to personally identify particular individuals by combining small pieces of information. It is not that this is a new challenge: for example, brick-and-mortar businesses have long asked customers for a zip code when a customer pays with a credit card. These two pieces of information can be easily combined to find a home address. What's changed is that technology and the Internet has made it far easier to identify and, potentially harm, users at vast scale. And the idea of PII now seems to extend to seemingly innocuous bits information such as Internet browsing behavior. What was once viewed as categorically deterministic is now probabilistic in terms of re-identifiability and risk.

2.2.3 *Growing Public Awareness*

> **Steward Cheifert** (Interviewer): How concerned should I be as a user about what I put up on the Net?
>
> **John Markoff** (West coast correspondent, New York Times): I think it depends on what you do. There's better security coming for economic business transactions, but I'd be careful about putting my password on the net, I'd pick a password that's a safe password, and I wouldn't put my credit card up until there is security software up that will protect my credit card. (*The Internet*, 1995)

In the early 1990's, the Internet exploded into public awareness. 1993 signaled the first graphical web browser and the world was mesmerized. There was a growing excitement for interaction on the World Wide Web, and public talk was dominated by the promise of possibility. The large concern was whether it was safe to shop online — though the focus was on the encryption of credit card transactions, rather than the seemingly benign micro-exchanges of a user clicking through web pages.

In November of 1999, DoubleClick, an advertising network of over 11,000 publishers, purchased Abacus Direct a company maintaining a database of detailed consumer profiles on approximately 90% of American households ("In re: DoubleClick Inc.", 2001). DoubleClick announced that it planned to merge "anonymous" online data (100 million user profiles) with Abacus profiles (including names, addresses, phone

numbers, etc.). The media roiled and public sentiment was overwhelmingly negative. Ultimately, DoubleClick put these plans on hold: it's stock dropped dramatically after the merger was announced since neither the public nor Wall Street had reacted favorably to this plan (Chief Marketer, 1999). It was about this time that the public became aware of how their behavior on the Internet might be used by commercial entities. A Federal Trade Commission (FTC) investigation in 2001 ended its investigation with no finding that DoubleClick had engaged in unfair or deceptive trade practices. It concluded:

> Based on this investigation, it appears to staff that DoubleClick never used or disclosed consumers' PII (personal identifiable information) for purposes other than those disclosed in its privacy policy. Specifically, it appears that DoubleClick did not combine PII from Abacus Direct with clickstream collected on client Web sites. In addition, it appears that DoubleClick has not used sensitive data for any online preference marketing product, in contravention of its stated online policy ("In re: DoubleClick Inc.", 2001).

This case, and others like it, now had the public's eye. By the year 2000, 54% of Americans used the Internet at home (Simms, 2000). In a March 2000 poll, 82% of web users polled indicated that they were uncomfortable if their browsing habits and shopping patterns were linked to their identities (Business Week / Harris Poll, 2000). 74% indicated that they were very uncomfortable if their information were sold to other organizations. Though only 40% of users had heard of browser "cookies", many were aware that data was being collected silently and that data could be both sold and merged with identifiable information. Public negative sentiment toward tracking has changed little over the years. Over 70% in 2012 say they find little value in online ads (Hoofnagle, Urban, & Li, 2012). And Microsoft reports that 75% of its US and European users would like to opt-out of online advertising (Hill, 2012).

2.2.4 *Third-Party Data Collection*

Browser tracking is not the only means by which websites identify and collection user information. But it's, perhaps, the

sort of data collection people are intuitively least comfortable with. Third-party trackers are trackers on websites that collect information about users for a third party. They are not officially part of the website visited, but are generally associated with advertisers or web analytics. Typically, they take the form of cookies. Cookies can be set by embedded JavaScript and allow a site to essentially "remember" arbitrary bits of information from a previous session. Advertisers can utilize cookies in a cross-domain manner by creating a unique ID that is set by one website and then picked up by that same advertiser on another website. In this way, the advertiser knows that you've visited both sites.

The problem with cookies is that it is a browser-mediated information exchange *hidden from view*. Some cookies serve useful functions for sites with which we have trusted relations. But many are set and tracked by organizations that we may not recognize and for purposes that are unclear. Of great concern is that these entities may be buying, selling, and trading browser profiles which can be linked to personally identifiable information (PII). Narayanan (2011) proposed a five-fold taxonomy in which browsing history might become pseudoanonymously[4] identifiable to a particular device or user.

1. **The third party is sometimes a first-party.** A third party tracker such as Facebook may be a first-party in other contexts. If you visit Facebook and login, it sets a cookie that can be used to track you as you browse the web. Roesner, Kohno, and Wetherall (2012) describe the particular situation where sites have buttons such as Facebook, Twitter, or Google embedded on their web pages. These buttons are trackers and used to track you even if you never click on them. If you visit a website, and are also logged into Facebook in the same browser, Facebook will know that you've visited that site.

2. **A first party leaks data to third-parties.** Krishnamurthy, Naryshkin, and Wills (2011) found that 56% of 100 popular, non-social network sites leaked private information. This number grows to 75% if a site userid is included as a PII.

4 "Psuedoanonymous" unique IDs may be used to identify a particular device or user agent. These may potentially be linked to other, personally identifying information later.

3. **A first party sells the user's identity.** The previously mentioned case of Abacus and DoubleClick serves as a good example. Also, Narayanan (2011) notes that survey sites (e.g., "win a free iPod!") can also act as an identity provider to sites on which they have a third-party presence.

4. **Hacks and exploits.** A third-party might exploit a cross-site vulnerability on a first-party website to learn the user's identity (Narayanan, 2011).

5. **Deanonymization (re-identification).** A third party could match browsing histories against datasets to re-identify particular persons, as Narayanan and Shmatikov (2010) did with the Netflix Prize dataset. In this experiment, they demonstrated that it's possible to identify a particular subscriber with only a little bit of information.[5]

Though Netscape never intended cookies to be privacy-invasive, enterprise-minded businesses quickly realized that cookies could be used for identifying browsers. In fact, the ecology of tracker technology in the wild is surprising complex and, potentially, adaptable.

Roesner et al. (2012) performed a study in which they observed tracker behavior from the point-of-view of the client browser. They classified this behavior into five behavioral patterns in which a given tracker site might exhibit variable behavior depending on contexts. They note complex behaviors such as that of an aggregate tracker referring data to other trackers. A few trackers stored data in multiple locations simultaneously (e.g., cookies, flash storage, HTML5 local storage) or exhibited spawning behavior when cookies were cleared. When they asked the question of how much data can any one tracker could collect, using 2006 AOL search query data, they found that, on average, DoubleClick could track a user across 39% of pages visited, and a maximum of 66% of pages visited.

Though cookies are the most prevalent trackers, "web bugs" serve a similar function. They are traditionally embedded as small (often 1 pixel) GIF or PNG images and can also be used in HTML mail, informing advertisers whether an email

5 In fact, this is only one example of many. Re-identification is another huge challenge for privacy law. For discussion in greater depth, see Ohm (2010).

has been opened and when. In March 2010, Ghostery identified 117 unique web bugs on nearly 400,000 unique domains (Gomez, Pinnick, & Soltani, 2009). Though privacy policies may state that information will not be shared with third parties, many of these sites nonetheless allow third-party tracking through web bugs (Gomez et al., 2009).

Given the complex behavior of trackers, coupled with complex business practices in which first-parties and third-parties may mingle or exchange data, it is not surprising that users may be confused about who has collected, stored, traded, or used their browsing profile.

2.2.5 *Cookies and Privacy Law*

Recent litigation over OBT invokes Wiretapping acts both federal and state. At the federal level, the Electronic Communications Privacy Act ("ECPA") Wiretap Act, Stored Communications Act ("SCA"), and the Computer Fraud and Abuse Act ("CFAA") are frequently invoked.

The federal Wiretap Act (as amended by the ECPA) protects the privacy of wire, oral, and electronic communication. The latter is defined as "any transfer of signs, signals, writing, images, sounds, data, or intelligence of any nature transmitted in whole or in part by a wire, radio, electromagnetic, photoelectronic or photooptical system" (Title 18, Part 1, Chapter 119 §2510(12)). Under §2511(3), "a person or entity providing an electronic communication service to the public shall not intentionally divulge the contents of any communications... while in transmission on that service to any person or entity other than an addressee or intended recipient..."

The ECPA has been criticized lately by the media:

> Many Internet companies and consumer advocates say the main law governing communication privacy — enacted in 1986, before cellphone and e-mail use was widespread, and before social networking was even conceived — is outdated, affording more protection to letters in a file cabinet than e-mail on a server." (Helft & Miller, 2011)

Such criticism had been directed at Government data requests, yet numerous cases solely involving commercial enti-

ties abound.[6] Of issue for OBT is an exception known as the *consent exception*: so long as one of the parties of electronic communication has given prior consent to the interception, and the interception was not intended for criminal or "tortuous act", then it is exempt from the Wiretap Act (Center for Democracy and Technology, 2008).

One notable case which invoked ECPA, SCA, and CFAA was the case of "In re Doubleclick, Inc.". In 2000, a class action suit was brought against Doubleclick Inc., which used cookies for behavioral advertising. With respect to ECPA Wiretap law, Doubleclick successfully argued that its affiliated websites were "users" of the Internet and that communications accessed by Doubleclick's cookies were "of or intended for" these websites ("In re: DoubleClick Inc.", 2001). The Stored Communications Act was found not to apply to Doubleclick, as well. The SCA was intended to address temporary and transactional records of internet service providers. Cookies stored on personal hard drives and were out of scope for the intent of SCA ("In re: DoubleClick Inc.", 2001). Finally, with respect to the Computer Fraud and Abuse Act, the plaintiff failed to provide evidence that they had incurred harm.

As it stands today, federal laws do not address concerns raised by online behavioral tracking on the Internet. Furthermore, given the single-party consent exemption of Federal Wiretap law, communications between Internet users and websites is not private if the website owner has given its consent for monitoring.

2.2.6 *Policy Legislation - Do Not Track*

Though the United States does not have specific federal law that applies to behavioral tracking, the Federal Trade Commission (FTC) can enforce the terms of privacy policies under Section 5 of the FTC Act prohibiting "unfair or deceptive" marketing practices (*United States Code Title 15 §45*). This applies to FTC handling of behavioral tracking. In 2011, the FTC found three parties in violation for the FTC act relating to third-party tracking (Mayer & Mitchell, 2012).[7]

6 For example, "In re Doubleclick, Inc." and others referenced at http://www.internetlibrary.com/topics/electronic_cpa.cfm.

7 A violation is determined if a company states in a website notice or privacy policy that it does not engage in a particular practice, when in fact it does.

In 2010, the Federal Trade Commission proposed "do not track" technology as "likely a persistent setting on consumers' browsers — so consumers can choose whether to allow the collection of data regarding their online searching and browsing activities" (Federal Trade Commission, 2010). This represents a form of self-regulation insofar that trackers are expected to respond appropriately to browser requests not to track. In its March 2012 report Federal Trade Commission (2012), the FTC outlined a broader privacy framework to address a myriad of concerns surrounding data brokering, mobile, internet service providers, and enforceable self-regulatory codes. The scope of this framework applies to:

> [...] all commercial entities that collect or use consumer data that can be reasonably linked to a specific consumer, computer, or other device, unless the entity collects only non-sensitive data from fewer than 5,000 consumers per year and does not share the data with third parties. (Federal Trade Commission, 2012)

The FTC notes that this framework applies to offline as well as online data and applies to data that is reasonably linkable to a specific consumer, computer, or device.

As proposed by FTC (Federal Trade Commission, 2012), Do Not Track should include five key principles:

1. It should be implemented universally to cover all parties that would track consumers.
2. The choice mechanism should be easy to find, easy to understand, and easy to use.
3. Any choices offered should be persistent and should not be overridden if, for example, consumers clear their cookies or update their browsers.
4. It should be comprehensive, effective, and enforceable: it should opt consumers out of behavioral tracking through any means and not permit technical loopholes.
5. It should go beyond simply opting consumers out of receiving targeted advertisements; it should opt them out of collection of behavioral data for all purposes other than those that would be consistent with the context of

the interaction (e.g., preventing click-fraud or collecting de-identified data for analytics purposes; Federal Trade Commission, 2012).

Since the FTC released its preliminary findings, the W3C Internet standards body convened a working group toward the adoption of an industry-wide Do Not Track standard. This group includes representatives from the advertising industry, online businesses, academia, privacy advocates — both nationally and internationally. The challenge for this group is to devise a standard that meets FTC guidelines, while satisfying a diverse array of agendas.

2.2.7 *Building Privacy Policies into Browser Protocols*

In the late 1990s, there were a couple of attempts to address the problem of privacy invasion introduced by third-party cookies via technical means. Introduced in 1997, RFC 2109 was a proposal aimed at putting strict controls on cookie usage. Commercial entities participating in standards discussion objected to portions of this proposal and it was, ultimately, not supported by browser developers. "Where the browser manufacturers had been asked to redesign their software to reject cookies automatically, Netscape and Microsoft Internet Explorer instead included options for users to reject cookies if they so chose" (Eichelberger, n.d.). As cookie inventor Lou Montulli stated, "browser default would still be set to accept cookies, since it was felt by the designers that if we were to unilaterally disable this feature, existing content on the Web would no longer work" (Bruner, 1997).

From 1998 - 2003, the W3C organized a standards committee to propose a Platform for Privacy Preferences (P3P) standard (Schwartz, 2009). The standard aimed to create a machine-readable policy that captured the complexity of human-readable privacy policies and boiled them down to a multiple-choice set of options. What resulted from long and heated discussion was highly controversial and touted as overly complex — and unlikely to be adopted as a self-regulatory option with no degree of enforcement.

About the same time, advertisers began to take a more active role in self-regulation. The National Advertising Initiative (NAI) set up a website to allow users to download opt-out cookies for participating networks. Their principles have been

to notify users and allow them a choice. Opt-out does not mean that you will not be tracked or that ads will not be served. It means that ads will not be tailored to your browsing behavior. Opt-out choices are stored in browser cookies, so they recommend to visit their opt-out page periodically for updates.

Other advertising networks and companies have since also offered similar opt-out cookie services.[8] Large advertising industry groups formed a new parallel self-regulatory program for behavioral advertising, the Digital Advertising Alliance (DAA). The DAA and the Interactive Advertising Bureau (IAB), a self-regulatory for advertising networks, advocate the use of an advertising option icon to be displayed on web pages with tracking scripts. A user who clicks on the icon sees a disclosure statement and can click through to an opt-out page. However, according to a 2011 study at the Center for Internet and Society at Stanford University, only 11.3% of members to date offer cookie-based opt-out. (Mayer, 2011).

In 2007, the Center for Democracy and Technology (CDT) sent a note to the Federal Trade Commission (FTC) requesting that "the online tracking and targeting of consumers — both in its current form and as it may develop in the future — needs to be limited so that consumers can exercise meaningful, granular preferences based on timely and contextual disclosures that are understandable on whichever devices consumers choose to use" (Center for Democracy and Technology, 2007). They called for a national Do Not Track list similar to the national Do Not Call list. Their proposal called for a machine readable list of the domain names which used cookies and other means to track users. Browsers and third party software could then use this list in order to limit tracking (Center for Democracy and Technology, 2007).

Thus, the W3C chartered the Tracking Protection Working Group (TPWG) "to improve user privacy and user control by defining mechanisms for expressing user preferences around Web tracking and for blocking or allowing Web tracking elements" (W3C TPWG, n.d.).

In effect, discussion in this group mirrors, and perhaps fuels, public debate in media and society.[9] Chairs for the group

8 For example, see http://www.worldprivacyforum.org/cookieoptout.html.

9 The public-tracking list is archived at: http://lists.w3.org/Archives/Public/public-tracking/.

Figure 6: Tweet, 7 September 2012

serve to organize and moderate debate. As questions become more defined, formal "issues" are raised and "actions" assigned to particular members. For example, ISSUE–5, "What is the definition of tracking?" was opened on September 21, 2011. As of October 2013, there are 446 emails pertaining to this issue alone, and it is still open more than two years later.

Because each member has his or her own platform or stance on any particular matter, debate sometimes bleeds into other venues. For example, as of this writing, the current DNT draft standard says the following (regarding determining a person's DNT preference): "A user agent must have a default tracking preference of 'unset' (not enabled) unless specific tracking preference is implied by the decision to use that agent" (W3C TPWG, 2011). When TPWG member Roy T. Fielding, who works for Adobe but who also contributes to the open source Apache webserver codebase, committed a patch to this codebase which disabled DNT if the browser requesting data was Internet Explorer 10 (which at the time set DNT to 1 by default), he caused a ripple tide of effects. His patch, labeled "Apache does not tolerate deliberate abuse of open standards" (Fielding, 2012a), clearly reflects his own position on the implementation of DNT. This hit C/NET news and Twitter almost simultaneously. On Twitter, Jonathan Mayer (Mayer, 2012b), fellow TPWG member responded with Figure 6.

About the same time, C/NET reported the event, noting Roy T. Fielding's position as an author of the DNT standard and principal scientist at Adobe Systems (Shankland, 2012). Of course, there was an immediate back-lash on the public-tracking list between Mayer and Fielding, in which Mayer (Mayer, 2012a) partially retracted his comment on Twitter Figure 7.

From Fielding on the public-tracking list:

> Our charter forbids us from specifying UI requirements. That does not mean any of the following excerpts are ambiguous:

Jonathan Mayer @jonathanmayer 7 Sep
Adobe's lead DNT negotiator blocked the most popular browser in the most popular server. PR: hey, it was on his own time. Wow that's lame.
Expand

Figure 7: Tweet, 7 September 2012

> A user is an individual human. When user-agent software accesses online resources, whether or not the user understands or has specific knowledge of a particular request, that request is made "by" the user.
>
> The goal of this protocol is to allow a user to express their personal preference regarding tracking to each server and web application that they communicate with via HTTP ...
>
> Key to that notion of expression is that it MUST reflect the user's preference, not the choice of some vendor, institution, or network-imposed mechanism outside the user's control. The basic principle is that a tracking preference expression is only transmitted when it reflects a deliberate choice by the user. In the absence of user choice, there is no tracking preference expressed. (Fielding, 2012b)

On the basis of this personal conviction, Fielding argued that if DNT was switched on by default, it should be ignored, regardless of other considerations. Since Microsoft had announced that it would turn on DNT by default, Fielding retaliated using the technological means at hand — direct modification of code. Naturally, this led to heated debate within the Apache http server developer community. As of this writing, there were 364 comments on this commit. Fielding's changes are reflected in the current configuration of the main branch of code, but are commented out.

Lorrie Craner, one of the original developers of P3P is skeptical that DNT will succeed. To her it seems a simpler, but not better, replay of the P3P proposal (Federal Trade Commission, 2012; Fulton, 2012).

> As we walk about in the physical world, we raise and lower our voice and we raise and lower our window shades and we turn our faces, and we are all constantly adjusting to regulate our exposure and our privacy, Dr. Cranor tells RWW. "And

> it comes naturally; we don't spend a lot of time thinking about it. We just sort of naturally do it. But when we go online, it's no longer natural, because we don't have these readily apparent, physical things where you can just easily close that shade, and it's obvious what you're doing. So we have to rely on software tools to help us with this privacy regulation process. (Fulton, 2012)

2.2.8 *Choice and Self-Regulation*

While policy debate rages, the advertising industry continues to advocate self-regulation. Central to self-regulation is the notion of choice. The DAA and NAI naturally view behavioral advertising in a very positive sense such that better targeted ads are preferable to ads that do not match a consumer's interests. But if a user does not wish to participate, then they should have the choice to opt-out. However, privacy advocates cite two large concerns with respect to industry self-regulation.

The first argument against the effectiveness of self-regulation is non-compliance. A Carnegie Mellon University (CMU) report by Komanduri, Shay, Norcie, and Ur (2012) finds that, despite attempts at self-regulation by bodies such as the DAA and NAI, there are still many examples of non-compliance. Stanford University website `http://donottrack.us` believes that further steps are required: namely, regulation enforcing compliance with user choice.

A second, perhaps more compelling argument is that the advertising industry relies on technologies that seem to ensure that users remain unaware of them. When consumers attempt to block tracking, other more resistant methods are developed which circumvent these defenses. While DNT would make it very easy for users to decide not to be tracked, market choice seems to make it easier for them to be tracked. As Hoofnagle, Soltani, Good, Wambach, and Ayenson (2012) note, "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 the consumer can leave without leaving behind a data trail. But in the online world, invisible attributes leave marks that are easy to follow. They further note that surveys of top websites in 2009 and 2011, revealed new tracking mechanisms resistant

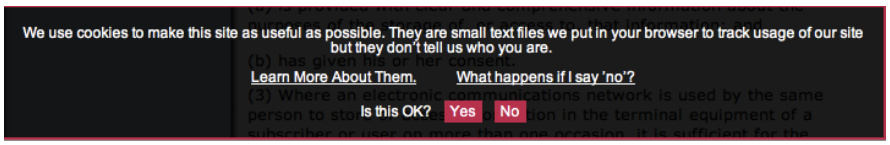

Figure 8: Example cookie opt-in notification from: http://www.jonwallacedesign.com/privacypolicy.html

to the strongest of privacy settings. Or, perhaps, advertisers simply know that once online shopping is habit, consumers will adapt.

2.2.9 *Opt-In Versus Opt-Out*

On the 26th of May 2012 in the U.K., a new European Cookie Law, E-Privacy Directive 2009/136/EC, came into effect.[10] This law requires the consent of users before cookies are set (see Figure 8). In effect, this is an unequivocal "opt-in" system of notification and consent (with the exception of cookies "strictly necessary" for the operation of a website (Official Journal of the EU, 2009).

Unlike current trends in US policy and law, this puts the burden of consent on the website publisher — not the user. Though this issue permeates the entire European Union (EU), discussion of the UK "Cookie Law" requirement for consent has finally reached US audiences on the Internet.

Perhaps, this EU Directive gives a sense of what may be required if US lawmakers decide to place the burden of consent on website publishers and not users. Such a shift would also acknowledge that Internet transactions between users and websites are currently not "private" as Federal Wiretap law has held. The EU Directive seems more in line with user expectations, but given current criticism on the difficulty with compliance, falls short with respect to implementation. The hope is that the W3C Tracking Protection Working Group (TPWG) will provide strong enough guidance to satisfy this Directive. However, US members (and especially those representing business interests) may decide to act in an insular fashion and have, thus far, shown little motivation to join a more global debate (W3C TPWG, 2012).

10 There is a high-level summation here: http://gigaom.com/europe/cookie-law-explainer/.

2.3 USER CONFUSION

There has long been an asymmetry in communication between web publishers and users. Interactive page elements such as "contact us" via web forms, instant chat, web forums, and email have enabled users to connect directly to persons behind the scenes. More recently, publishers utilize personal messaging through other channels such as Twitter, Facebook, or SMS for allowing users to offer direct and personal feedback. These are direct and intentional communications between parties. But the act of visiting a webpage in itself also has communicative value.

In this section, I highlight user confusion regarding indirect communications where advertisers and other commercial actors silently monitor web interaction via the use of cookies. Not only is the phenomenon of behavioral tracking difficult to understand and see, but user defenses against ubiquitous monitoring are confusing, as well. Of particular interest are user opt-out mechanisms provided for by emerging DNT standards and also the advertising industry efforts toward self-regulation of online behavioral advertising. The vehicle for this is "consumer choice" and the Advertising Option Icon (viz. AdChoices icon; see `http://www.aboutads.info` for general information).

2.3.1 *Confusion About Third Party Tracking*

In a 2009 study of Internet user understanding of behavioral advertising, McDonald, Reeder, Kelley, and Cranor (2009) found that only 51% believed that the ability of an ad company to determine which ads to show them based on the history of prior websites they visited was something that "happens a lot right now" (McDonald et al., 2009). 46% of the same study participants found the idea of behavioral advertising "creepy", and another 40% agreed or strongly agreed they would be more careful online if they knew advertisers were collecting data.

In the virtual online world where billions of people have access to the web through the view of what may often be a personal device, it's not surprising that the experience feels, to some degree, private. As privacy advocate Christopher Soghoian remarks,

> Consumers treat the search engine box like their psychiatrist, their rabbi, their priest, their doctor. People type the most intimate things into search engines and other websites primarily because they think they're anonymous. They type in things on WebMD that sometimes they wouldn't even ask their own doctors... And in fact, we are not anonymous, these sites are tracking us." (as cited in Kessler, 2010)

The UC Berkeley School of Information KnowPrivacy Project (Gomez et al., 2009) looked specifically at user concerns and knowledge. They examined 50 of the most visited websites and their privacy policies. They also considered specific practices such as third-party tracking and sharing with affiliates. Not surprisingly, they found that most users are concerned about data collection on themselves and their control over that collection and use of information. Furthermore, they found overwhelming evidence (from prior surveys) that users lack knowledge and understanding about data collection practices and policies.

Once source of obvious confusion has to do with terminology. What is tracking? Does that include collection? Most users think not (McDonald, 2011; McDonald & Peha, 2011). What is the difference between "third party" and "affiliate"? Privacy law treats these categories differently: third-party information sharing is subject to more restriction. According to Gomez et al. (2009), websites make distinctions between sharing with affiliates and third parties. 29 of 50 companies examined state that they do not share data with unrelated third parties, 45 state that they share data with affiliates, and 36 state that they allow third-party tracking. This last category falls outside of privacy policy coverage. As Gomez et al. (2009) note, it's very unclear what it means to not share data with unrelated third parties yet permit third party tracking.

Despite assurances that affiliates and third parties are treated differently, website providers often leak personal information to the less trusted, embedded third party trackers such as username, login time, and other information. Just as troubling is that users have no practical way of knowing who affiliates are and what sorts of information are passed to them. In the Know Privacy Report (2009), many of the websites examined are owned by companies with hundreds of subsidiaries: NewsCorp has over 1500 subsidiaries while Bank of America

has over 2300.[11] The pervasive presence of third party trackers on the Internet means companies have ample opportunity to construct long term profiles of things we do online.

Privacy policies often go unread, or may be difficult to understand (Milne & Culnan, 2004; Sherman, 2008). Fernback and Papacharissi (2007) conducted an in-depth discourse analysis of privacy statements of three large companies. They also noted discrepancies regarding the use of the term "third-party". From their study of the Real Network privacy policy: "We will never sell, rent or disclose to third parties our customers' personally identifiable information ... gathered on a RealNetworks Website unless we are required to do so by law or receive your advance informed consent" (Fernback & Papacharissi, 2007). Yet, in what appears a blatant contradiction, later in the same privacy policy:

> Your personally identifiable information may be transferred in connection with a sale, merger, transfer, exchange or other disposition (whether of assets, stock or otherwise) of all or a portion of a business of RealNetworks, Inc. and/or its subsidiaries. (Fernback & Papacharissi, 2007)

In general, Fernback and Papacharissi (2007) note that privacy policies are seen to protect the company and typically avoid mention of specifics such as precisely what information might be collected and for what purpose. Given that the FTC has a history of prosecuting companies that violate their terms of agreement,[12] this is not terribly surprising.

McDonald, Reeder, Kelley, and Cranor (2009), examined the readability of online privacy statements. They examined standardized text formats and compared these to free text policies. They found that experiment participants could more readily read and comprehend standardized formats, but at the expense of accuracy. And users, in general, did not like either standardized or free text formats. But when it comes to purchasing privacy-sensitive items, Egelman, Tsai, Cranor, and Acquisti (2009) found that users actually paid attention to graphical privacy indicators.

11 These numbers reportedly do not include subsidiaries of subsidiaries.

12 See http://www.ftc.gov/os/caselist/index.shtm for example. Google and Facebook have both earned penalties.

2.3.2 *Confusion about Defenses*

When it comes to defending against third party tracking and privacy, users have a number of choices. Roesner et al. (2012) identified the following options:

- **Pop-up blockers:** Pop-up blockers can stop trackers from forced pop-ups, but there are similar methods such as site re-directs that may not be stopped;
- **Browser third-party cookie blocking:** Many third party cookies can be blocked if users select third-party cookie blocking in their browers. However, this will not stop the sort of tracking when a Facebook tracker is accepted as a first-party tracker and its role changes later to a third-party tracker;
- **Private browsing:** Private browsing mode was designed to protect users from having their browse state examined by physical access. But it does not keep browsing state from being examined online;
- **Opt-out cookies (and AdChoice icon):** The Digital Advertising Alliance (DAA) is an industry funded policy group that hosts an opt-out web page. From this page, users can click a button to set opt-out cookies;[13]
- **Clearing browser state:** Clearing cookies when closing the browser is a simple means to reduce the effects of tracking. However, this may also remove opt-out cookies and is also not effective against re-identification by trackers such as the Facebook like icon when logged in;
- **Do not track (DNT):** The proposed FTC "Do Not Track" DNT policy is designed to give users a way to opt out of web tracking. This is accomplished via an http request header with a DNT=1 header to inform the remote server that the client wishes to opt out. DNT is not mandatory and requires no compliance. The DAA has committed only to stop content personalization, if it receives a DNT signal. (`http://www.aboutads.info/choices/`); and,

13 There are other similar sites that offer opt-out cookies. For example, Evidon Global, and others have web pages with such links.

- **Blocking plug-ins:** There are a number of browser plug-in which are designed to block trackers.

Lorri Craner directs the CyLab Usability Privacy and Security (CUPS) Laboratory at Carnegie Mellon University. She and her students have conducted a series of studies centered on user understanding of online behavioral advertising and usability of blocking tools (Cranor, 2012; Leon, Ur, et al., 2012; McDonald, 2010; Ur, Leon, Cranor, Shay, & Wang, 2012). Leon, Ur, et al. (2012) studied the usability of Ghostery, Adblock Plus, and the Internet Explorer Tracking Protection List blocking tools. They found that self-help blocker tools have significant issues in terms of user understanding:

- Users don't recognize the names of the majority of companies that they can opt-out;
- Some of the tools use terms that were meaningless to participants: for example, "web tracker, web bug, flash cookie, silverlight cookie, tracking cookie, script, iframe, and targeted ad network.";
- Participants testing opt-out tools did not understand what the tools would opt them out of, mistakenly believing that they were protected against tracking;
- Opt-out tool users thought deleting cookies would protect their privacy even more, not realizing that deleting their cookies would also delete their opt-out cookies and undo their opt-out;
- Users were left unaware whether or not most tools were working, and oblivious to what was happening behind the scenes;
- None of the opt-out tools tested notify users while they are browsing that their preferences are being respected; and,
- Participants who tested the browser cookie settings also had no mechanism for understanding what exactly was happening behind the scenes unless websites didn't work (Leon, Ur, et al., 2012).

In a related study conducted by Leon, Cranshaw, et al. (2012), 45% of participants who saw "AdChoices" believed

that it was intended to sell advertising space, while only 27% believed that it was intended to stop tailored ads. A significant part of the problem is that 1) users don't really understand the mechanisms behind tracking and so misunderstand the nature of the problem; and 2) users don't have a clear idea of what options are available to help alleviate the problem (McDonald, 2010; Ur et al., 2012). Moreover, users have no practical way of knowing how effective blocking is when it works. Mayer and Mitchell (2012) examined the effectiveness of 11 blocking tools and found significant variability in performance. Most of these self-help tools work about the same way. They consist of a black list that is modifiable by the user. The lists vary widely and account for much difference seen in performance. Since it's not obvious that these tools essentially work in the same way, users may be tempted to install more than one. However, doing so may be risky: Mayer and Mitchell (2012) report that TRUSTe's tool actually over-rode blocking lists by other tools allowing tracking by several large third party trackers.

2.3.3 *Are Privacy Concerns Affecting User Behavior?*

When questioned in polls and studies, users overwhelmingly share a negative attitude toward online tracking. In a 2010 CUPS study:

> 64% found 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 found 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. We found participants are comfortable with the idea that advertising supports free online content, but they do not believe their data are part of that exchange. (McDonald & Cranor, 2010)

Though, a large proportion of users claim that they would change their online behavior if they believed advertisers were collecting data, there appears to be no general usage statistics to indicate that so large a population employs existing opt-out

technology beyond what the browser may natively provide. In 2011, Fowler (2011) reported that only 5.6 of Firefox desktop users had turned on DNT. In terms of more aggressive blocking, when I started this dissertation in late 2012, 14 million Firefox users had installed the most popular blocker extension (Ad Blocker). But this was only approximately 3% of the 450 million Firefox users which represented only 20–24% of desktop browsers in use (Mozilla, n.d.; Statcounter, 2012). However, trends do indicate that blocking trackers is growing in popularity (Acohido, 2011). In late 2013, DNT adoption is approximately 17% (Fowler, 2013b). And the new default setting for Firefox (as of Feb 2013) is not to allow all third party domains, but to allow cookies only from visited domains (Fowler, 2013a).

In some contexts, users may be willing to make very clear choices when confronted with a privacy trade-off. Egelman et al. (2009) found that laboratory study participants of an online shopping task were willing to pay more for sensitive purchases when confronted with a choice between a site with less privacy but cheaper prices, and a site with more privacy but more expensive prices. Hoofnagle, King, Li, and Turow (2010) likewise found that more than half of online poll participants reported changing their minds about the purchase of a product online because of privacy concerns.

As users awareness grows, advertisers find new strategies to evade blocking (Hoofnagle, Urban, & Li, 2012; Leon, Ur, et al., 2012; Soltani, Canty, Mayo, Thomas, & Hoofnagle, 2009). What users want is not in accordance with what publishers and advertisers are currently willing to do. In a survey from McDonald and Peha (2011), 72% expected that regulatory "do not track" efforts limit data collection, while 34% of respondents expected that "do not track" would prevent data from being collected by websites and advertisers. Since industry proponents currently interpret "do not track" as not affecting data collection but simply use of data for presenting advertisements, there exists an impasse between consumer advocates and industry proponents in policy-oriented regulatory efforts by the W3C.

Central to this debate is the notion of "consumer notice and choice". Advertising self-regulation is firmly based on this notion. As we will see in chapters to come, there are strong reasons why this is so.

Figure 9: Interactive Print Advertisement

2.4 THE FUTURE OF INTERACTIVE ADVERTISING

Though the goal of advertising is ultimately to sell products, advertisers also have other goals such as gaining wider audiences and positive brand engagement. Interactive advertising plays a role in the space by providing consumers opportunities to interact with advertisements directly.

I live in the country. Recently, I was finally able to have high-speed Internet installed at my house. For the first time, I could stream broadcast content over my AppleTV. While I was poking around, I found I could access a PBS channel for free. In order to activate it, I was shown a code and asked to visit the PBS website to register. After registering, content was made available over my AppleTV. Now PBS knows who I am, what I watch, and when I watch it. Though, I am not engaging with ads over PBS yet, Apple has transformed the broadcast experience by opening a window between PBS and me, whereby we can affect each other's actions.

In the past year, we've seen other, more unusual, examples of interactive advertising. Lexus has created an ad which "comes to life" when an iPad is placed behind the printed magazine ad.[14]

Car manufacturers had already engaged a potential audience by creating game-like experiences where consumers could design their own vehicle and examine it in 3D. Cross-media experiences such as this, transform print into something new.

14 http://www.lexus.com/stunning

Figure 10: Print-Embedded Smartphone

On October 5, 2012, *Entertainment Weekly* placed a smartphone inside its print magazine. The phone contained a digital ad running video and live tweets (Ulanoff, 2012).

It has become easier to imagine a world where print seems fluid and magic — as envisioned by J.K. Rowling portrays the Daily Prophet in the world of Harry Potter.

Not surprisingly, there is a strong interest by advertisers to engage over mobile platforms. Utilizing location-based coupons, the iButterfly app engages customers by gamifying coupons.[15] Users flick their phone to hunt virtual 3D butterflies.

While interactive advertising may appear tangential to behavioral advertising, there is a common theme underlying both: *interactivity*. Behavioral advertising relies on the ability to observe users engaging in everyday activities while interactive ads provide a means for advertisers to invite consumers to engaged with brands and products.

Why should this parallel concern us? It does not seem farfetched to imagine the content of interactive ads become more tailored to the interests and expectations of individuals. Advertisers already argue that OBA benefits the consumer by delivering a more tailored (and presumably less annoying) experience. The argument posed by advertisers is fairly simple: wouldn't you rather see advertisements about products that interest you rather than advertisements about things that do not?

15 http://www.cherrypicks.com/products/ibutterfly

Suppose that advertisers collect images of your friends and family and create facsimiles which appear in advertisements.[16] In June 2013, Alessandro Acquisti spoke of the following research in progress.

> Imagine that an organization has access to your list of Facebook friends, and through some kind of algorithm they can detect the two friends that you like the most. And then they create, in real time, a facial composite of these two friends. Now studies prior to ours have shown that people don't recognize any longer even themselves in facial composites, but they react to those composites in a positive manner. So next time you are looking for a certain product, and there is an ad suggesting you to buy it, it will not be just a standard spokesperson. It will be one of your friends, and you will not even know that this is happening. (Aquisti, 2013)

The problem is that, even with policy efforts toward transparency of collection and use, it is remarkably easy to manipulate people to behave in predictable ways. As advertisements become increasingly interactive — and as methods for collecting and using online behavioral data become more sophisticated — we will need to become much more attuned to ways in which this data may be used to drive behavior.

2.5 SUMMARY

This chapter introduced the phenomenon of online behavioral advertising to include issues leading to the formation of law and policy as well as issues concerning user notice and choice — principles espoused by policy makers and advertiser self-regulatory bodies alike.

Two mechanisms for user control were introduced: "do not track" and the AdChoices advertising option icon. Both present a means for users to opt-out of behavioral advertising. Both present user interactive mechanisms of choice. And both present opportunities for manipulating behavior in subtle ways.

16 FaceBook may already use your posts and personal data for advertising (Goel, 2013). Google also includes Google user names, faces, and content in ads (Kelly, 2013).

The remainder of this dissertation is largely concerned with the study the cause of user confusion in specific contexts. The next chapter first introduces a theoretical framework for meaning and understanding in interaction. In addition, I briefly cover previous work in the areas of graphical communication and language and advertising.

BIBLIOGRAPHY

Acohido, B. (2011, February). *Most Google, Facebook users fret over privacy.* Retrieved from http://www.gallup.com/poll/146159/facebook-google-users-skew-young-affluent-educated.aspx

Angwin, J. (2012, Jun 17). Online tracking ramps up. *Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB10001424052702303836404577472491637833420.html

Aquisti, A. (2013). Why privacy matters [video]. *TED Talks*. Retrieved from http://www.ted.com/talks/alessandro_acquisti_why_privacy_matters.html

Bruner, R. E. (1997). Cookie proposal could hinder online advertising. *Advertising Age*. Retrieved from http://adage.com/article/news/advertisers-win-debate-cookies/405/

Business Week / Harris Poll. (2000, March). *A Growing Threat* (Tech. Rep.).

Center for Democracy and Technology. (2007). *Ehavioral advertising: Tracking, targeting, and technology.* Retrieved from https://www.cdt.org/privacy/20071031consumerprotectionsbehavioral.pdf

Center for Democracy and Technology. (2008, July). *An overview of the Federal Wiretap Act, Electronic Communications Privacy Act, and State two-party consent laws of relevance to the NebuAd system and other uses of Internet traffic content from ISPs for behavioral advertising* (Tech. Rep.).

Chen, Y., Pavlov, D., & Canny, J. F. (2009, June). Large-scale behavioral targeting. In *KDD '09: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM Request Permissions.

Chief Marketer. (1999, January). *Doubleclick, abacus stock drop after merger news.* Retrieved from http://chiefmarketer.com/news/doubleclick-abacus-stocks-drop-after-merger-news

Constine, J. (2012, Aug 15). Facebook's new retargeted ads performing "very well", adds partners to run them. *Tech Crunch*. Retrieved from http://techcrunch.com/2012/

08/25/facebook-exchange-retargeting/

Cranor, L. (2012). Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *Journal of Telecommunications and High Technology*.

D'Orazio, D. (2013, Oct 10). Microsoft said to work on technology to replace cookies, track across Windows, Bing, and Xbox. *The Verge*, 1–2. Retrieved from http://www.theverge.com/2013/10/10/4823944/microsoft-reportedly-working-to-replace-cookies-ad-tracking

Duggal, J. (2012, May). *Promise Unfulfilled? Lessons from the Revolution (Quantcast)* (Tech. Rep.).

Duhigg, C. a. (2012a, Feb 16). How Companies Learn your Secrets. *New York Times Magazine*. Retrieved from http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=1&pagewanted=all

Duhigg, C. a. (2012b). *The Power of Habit*. Random House.

Eckersley, P. (2010). How unique is your web browser? In M. Atallah & N. Hopper (Eds.), *Privacy and Enhancing Technologies* (pp. 1–18).

Egelman, S., Tsai, J., Cranor, L. F., & Acquisti, A. (2009, April 4 - 9). Timing is everything? The effects of timing and placement of online privacy indicators. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009* (pp. 319–328).

Eichelberger, L. (n.d.). *The cookie controversy: privacy issues.* Retrieved from http://www.cookiecentral.com/ccstory/cc4.htm

Farahat, A., & Bailey, M. C. (2012, April). How effective is targeted advertising? In *WWW '12: Proceedings of the 21st International Conference on World Wide Web.* ACM.

Federal Trade Commission. (2010, December). *FTC staff issues privacy report, offers framework for consumers, businesses, and policymakers* (Tech. Rep.).

Federal Trade Commission. (2012). *Protecting Consumer Privacy in an Era of Rapid Change–A Proposed Framework for Businesses and Policymakers* (Tech. Rep.).

Fernback, J., & Papacharissi, Z. (2007). Online privacy as legal safeguard: the relationship among consumer, online portal, and privacy policies. *New Media & Society*, 9(5), 715–734.

Fielding, R. (2012a, September). *Apache does not tolerate deliberate abuse of open standards [Apache / httpd change request].*

Retrieved from https://github.com/apache/httpd/commit/a381ff35fa4d50a5f7b9f64300dfd98859dee8d0

Fielding, R. (2012b, September 12). *Re: Intermediaries interfering with DNT decision making. [Online forum comment].* Retrieved from http://lists.w3.org/Archives/Public/public-tracking/2012Sep/0165.html

Fowler, A. (2011, November 2). Do not track adoption in Firefox Mobile is 3x higher than desktop. Retrieved from http://blog.mozilla.com/privacy/2011/11/02/do-not-track-adoption-in-firefox-mobile-is-3x-higher-than-desktop/

Fowler, A. (2013a, Feb 25). *Firefox getting smarter about third-party cookies.* Retrieved from http://blog.mozilla.org/privacy/2013/02/25/firefox-getting-smarter-about-third-party-cookies/

Fowler, A. (2013b, May 3). Mozilla's new Do Not Track dashboard: Firefox users contine to seek out and enable DNT. Retrieved from http://blog.mozilla.org/privacy/2013/05/03/mozillas-new-do-not-track-dashboard-firefox-users-continue-to-seek-out-and-enable-dnt/

Fulton, S. (2012, February). *Dr. Cranor on "do not track" & the improbability of complete privacy.* Retrieved from http://www.readwriteweb.com/enterprise/2012/02/dr-cranor-on-do-not-track-and.php

Goel, V. (2013, November 15). Facebook reasserts posts can be used advertise. *The New York Times*. Retrieved from http://www.nytimes.com/2013/11/16/technology/facebook-amends-privacy-policies.html?_r=0

Gomez, J., Pinnick, T., & Soltani, A. (2009, Jun 1). Know Privacy. Retrieved from http://knowprivacy.org/

Google, I. (2010). Google milestones. *Corporate Information*. Retrieved from https://www.google.com/intl/en/about/company/history/

Helft, M., & Miller, C. (2011, January). Privacy law is outrun by the web. *New York Times*. Retrieved from http://www.nytimes.com/2011/01/10/technology/10privacy.html?_r=1&hp

Hill, K. (2012, October). Microsoft is losing in a bitter battle to protect Internet users' privacy. *Forbes Magazine*.

Hoofnagle, C. J., King, J., Li, S., & Turow, J. (2010, April). *How different are young adults from older adults when it comes to information privacy attitudes and policies?* (Tech. Rep.).

Hoofnagle, C. J., Soltani, A., Good, N., Wambach, D., & Ayenson, M. (2012). Behavioral advertising: The offer you cannot refuse. *Harvard Law & Policy Review 273. UC Berkely Public Law Research Paper No. 2137601*. Retrieved from http://ssrn.com/abstract=2137601

Hoofnagle, C. J., Urban, J. M., & Li, S. (2012). Privacy and modern advertising: Most us internet users want do not track to stop collection of data about their online activities. *Amsterdam Privacy Conference*, 1–27.

IAB. (2010, November). *IAB Legislative and Regulatory Affairs Update March 2010* (Tech. Rep.). Retrieved from http://www.iab.net/media/file/Legislative_Update_March_2010-2.pdf

IAB. (2012, April). *Internet Advertising Revenue Report: 2011 Full Year Results* (Tech. Rep.).

In re: DoubleClick Inc. (2001). *Privacy Litigation 154 F. Supp. 2d 497*.

The Internet. (1995, January). Retrieved from https://www.youtube.com/watch?v=XluovrUA6Bk

Jupiter Research. (2010, January). *Behavioral Targeting and the Purchase Funnel Opportunity* (Tech. Rep.).

Kazienko, P., & Adamski, M. (2007, January). Adrosa-adaptive personalization of web advertising. *Information Sciences: an International Journal*, *177*(11), 1–34.

Kelly, H. (2013, October 11). Why your face might appear in google ads, and how to stop it. *CNN*. Retrieved from http://www.cnn.com/2013/10/11/tech/social-media/google-plus-ads-profiles/

Kessler, S. (2010, November). Online behavior tracking and privacy: 7 worst case scenarios. Retrieved from http://mashable.com/2010/11/03/behavior-tracking-privacy/

Klopfer, P. H., & Rubenstein, D. I. (1977). The concept privacy and its biological basis. *Journal of Social Issues*, *33*(3), 52–65.

Komanduri, S., Shay, R., Norcie, G., & Ur, B. (2012). AdChoices? Compliance with online behavioral advertising notice and choice requirements. *CMU-CyLab-11-005*, 1–22.

Krishnamurthy, B., Naryshkin, K., & Wills, C. E. (2011). Privacy leakage vs. protection measures: The growing disconnect. *Computer Communications Review*, *40*, 112–117.

Krishnamurthy, B., & Wills, C. E. (2009). On the leakage of personally identifiable information via online social networks. *Web 2.0 Security and Privacy Workshop*, 7–12.

Kristol, D. M. (2001, November). HTTP Cookies: Standards, privacy, and politics. *Transactions on Internet Technology (TOIT*, *1*(2). Retrieved from http://arxiv.org/abs/cs.SE/0105018

Krux. (2012, Jun). *State of Data Collection on the Web* (Tech. Rep.). Retrieved from http://www.krux.com/pro/broadcasts/krux_news/krux_CIS_2012/

Leon, P. G., Cranshaw, J., Cranor, L. F., Graves, J., Hastak, M., Ur, B., & Xu, G. (2012). *What do online behavioral advertising privacy disclosures communicate to users?* (Tech. Rep. No. CMU-CyLab-12-008).

Leon, P. G., Ur, B., Shay, R., Wang, Y., Balebako, R., & Cranor, L. (2012). *Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising* (Tech. Rep. No. CMU-CyLab-11-017).

Locke, J. L. (2010). *Eavesdropping: An Intimate History*. OUP Oxford.

Mason, H. (2012, July). Machine learning for hackers. In *Devs Love Bacon.* Retrieved from http://www.hilarymason.com/presentations-2/devs-love-bacon-everything-you-need-to-know-about-machine-learning-in-30-minutes-or-less/

Mayer, J. (2011, August). *Tracking the trackers: The adChoices icon.* Retrieved from http://cyberlaw.stanford.edu/node/6714

Mayer, J. (2012a, September 12). *Adobe's lead DNT negotiator blocked the most popular browser in the most popular server. PR: hey, it was on his own time. Wow that's lame. [Tweet].* Retrieved from https://twitter.com/jonathanmayer/status/244300727726510081

Mayer, J. (2012b, September 7). *Adobe tries to circumvent the W3C Do Not Track consensus process, blacklist Internet Explorer 10 in Apache. Nice try. [Tweet].* Retrieved from https://twitter.com/jonathanmayer/status/244174917615108096

Mayer, J., & Mitchell, J. C. (2012, May). Third-party web tracking: Policy and technology. In *SP '12: Proceedings of the 2012 IEEE Symposium on Security and Privacy.* IEEE Computer Society.

McDonald, A. M. (2010). Cookie confusion: do browser interfaces undermine understanding? *CHI Extended Abstracts*, 4393–4398.

McDonald, A. M. (2011, April). User perceptions of do not track. In *W3C Workshop on Web Tracking and User Privacy*. Princeton.

McDonald, A. M., & Cranor, L. (2010). Beliefs and behaviors: Internet users' understanding of behavioral advertising. *TPRC*.

McDonald, A. M., & Peha, J. (2011). Track gap: Policy implications of user expectations for the 'do not track' internet privacy feature. *TPRC*.

McDonald, A. M., Reeder, R., Kelley, P. G., & Cranor, L. (2009). A comparative study of online privacy policies and formats. In *Privacy Enhancing Technologies* (pp. 37–55).

Milne, G. R., & Culnan, M. J. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, *18*(3), 15–29.

Mozilla. (n.d.). *Mozilla at a glance.* Retrieved from http://blog.mozilla.org/press/ataglance/

Narayanan, A. (2011, July 28). There is no such thing as anonymous online tracking. In *Stanford Law School Center for Internet and Society Blog.* Retrieved from http://cyberlaw.stanford.edu/blog/2011/07/there-no-such-thing-anonymous-online-tracking

Narayanan, A., & Shmatikov, V. (2010). Myths and fallacies of "personally identifiable information". *Communications of the ACM*, *53*(6), 24.

O'Connell, J. (2011, Dec 14). The Forrester Wave: Demand-side Platforms, Q4 2011 (Tech. Rep.).

Official Journal of the EU. (2009, December). *Directive 2009/136/EC of the European Parliament and of the Council.* Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:337:0011:0036:en:PDF

Ohm, P. (2010, August). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *57 UCLA Law Review 1701*, 1–77.

OMB. (2007, May 22). Safeguarding against and responding to the breach of personally identifiable information [Memorandum for the heads of executive departments and agencies]. M-06-16. , 1–22.

Packard, V. O. (1957). *The Hidden Persuaders. An Introduction to the Techniques of Mass-persuasion Through the Unconscious*. David McKay Co., Inc.

Pandey, S., Aly, M., Bagherjeiran, A., Hatch, A., Ciccolo, P., Ratnaparkhi, A., & Zinkevich, M. (2011). Learning to target: What works for behavioral targeting. In *ACM Conference on Information and Knowledge Management (CIKM)*.

Pole, A. (2010, October). How target gets the most of its guest data. In *Predictive Analytics World*. Retrieved from http://rmportal.performedia.com/node/1373

Reitman, R. (2012, September). A deep dive into facebook and datalogix: What's actually getting shared and how you can opt-out. In *Electronic Freedom Foundation Blog*. Retrieved from https://www.eff.org/deeplinks/2012/09/deep-dive-facebook-and-datalogix-whats-actually-getting-shared-and-how-you-can-opt

Roesner, F., Kohno, T., & Wetherall, D. (2012, April). Detecting and defending against third-party tracking on the web. In *NSDI'12: Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*. USENIX Association.

Schwartz, A. (2009, November). *Looking Back at P3P: Lessons for the Future* (Tech. Rep.). Retrieved from https://www.cdt.org/files/pdfs/P3P_Retro_Final_0.pdf

Scott, W. (1904, January). The Psychology of Advertising. *Atlantic Magazine*. Retrieved from http://www.theatlantic.com/magazine/archive/1904/01/the-psychology-of-advertising/303465/

Sesto, R., & Frankel, J. (2008, September). *How deep packet inspection changed the privacy debate* (Tech. Rep.). Retrieved from http://www.bingham.com/Publications/Files/2008/09/How-Deep-Packet-Inspection-Changed-the-Privacy-Debate

Shankland, S. (2012). Apache web software overrides ie10 do-not-track setting. In *CNET, Internet and Media*. Retrieved from http://news.cnet.com/8301-1023_3-57508351-93/apache-web-software-overrides-ie10-do-not-track-setting/

Sherman, E. (2008, September). Privacy Policies are great for PhDs. *CBS News Moneywatch*.

Simms, M. (2000, January). *National opinion fact sheet: Internet use 1998-1999* (Tech. Rep.).

Solove, D. J. (2008). *Understanding Privacy*. Harvard University Press.

Soltani, A., Canty, S., Mayo, Q., Thomas, L., & Hoofnagle, C. J. (2009, August). *Flash Cookies and Privacy* (Tech. Rep.).

Statcounter. (2012). *Global stats: Top 5 browsers on Aug 2012*. Retrieved from http://gs.statcounter.com/#browser-ww-monthly-201208-201208-bar

Title 18, Part 1, Chapter 119 §2510(12). (n.d.).

Ulanoff, L. (2012, October 2). There really is a smartphone inside EW Magazine. *Mashable*. Retrieved from http://mashable.com/2012/10/02/ew-has-smartphone-inside/

United States Code Title 15 §45. (n.d.).

Ur, B., Leon, P. G., Cranor, L. F., Shay, R., & Wang, Y. (2012). *Smart, useful, scary, creepy: perceptions of online behavioral advertising* (Tech. Rep. No. CMU-CyLab-12-007).

W3C TPWG. (n.d.). Retrieved from http://www.w3.org/2011/tracking-protection/

W3C TPWG. (2011). *Tracking Preference Expression (DNT)*. Retrieved from http://www.w3.org/2011/tracking-protection/drafts/tracking-dnt.html

W3C TPWG. (2012, June). *Discussion*. Retrieved from http://lists.w3.org/Archives/Public/public-tracking/2012Jun/0257.html

Woods, D. (2012, March). bitly's Hilary Mason on "What is A Data Scientist?". *Forbes Magazine*. Retrieved from http://www.forbes.com/sites/danwoods/2012/03/08/hilary-mason-what-is-a-data-scientist/

Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y., & Chen, Z. (2009, February). How much can behavioral targeting help online advertising? In *The 18th International Conference* (pp. 261–270). New York, New York, USA: ACM Press.