How Your Data Are Being Deeply Mined

Alice E. Marwick

JANUARY 9, 2014 ISSUE

Stuart Conway/Camera Press/Redux

A CCTV camera in Times Square, New York City, 2005

The recent revelations regarding the NSA's collection of the personal information and the digital activities of millions of people across the world have attracted immense attention and public concern. But there are equally troubling and equally opaque systems run by advertising, marketing, and data-mining firms that are far less known. Using techniques ranging from supermarket loyalty cards to targeted advertising on Facebook, private companies systematically collect very personal information, from who you are, to what you do, to what you buy. Data about your online and offline behavior are combined, analyzed, and sold to marketers, corporations, governments, and even criminals. The scope of this collection, aggregation, and brokering of information is similar to, if not larger than, that of the NSA, yet it is almost entirely unregulated and many of the activities of data-mining and

digital marketing firms are not publicly known at all.

Here I will discuss two things: the involuntary, or passive, collecting of data by private corporations; and the voluntary, or active, collection and aggregation of their own personal data by individuals. While I think it is the former that we should be more concerned with, the latter poses the question of whether it is possible for us to take full advantage of social media without playing into larger corporate interests.

Database Marketing

The industry of collecting, aggregating, and brokering personal data is known as "database marketing." The second-largest company in this field, Acxiom, has 23,000 computer servers that process more than 50 trillion data transactions per year, according to *The New York Times*.[1] It claims to have records on hundreds of millions of Americans, including 1.1 billion browser cookies (small pieces of data sent from a website, used to track the user's activity), 200 million mobile profiles, and an average of 1,500 pieces of data per consumer. These data include information gleaned from publicly available records like home valuation and vehicle ownership, information about online behavior tracked through cookies, browser advertising, and the like, data from customer surveys, and "offline" buying behavior. The CEO, Scott Howe, says, "Our digital reach will soon approach nearly every Internet user in the US."[2]

Visiting virtually any website places a digital cookie, or small text file, on your computer. "First-party" cookies are placed by the site itself, such as Gmail saving your password so that you don't have to log in every time you visit the site. "Third-party cookies" persist across sites, tracking what sites you visit, in what order. For those who have logged in, Google Chrome and Firefox sync this browsing history across devices, combining what you do on your iPad with your iPhone with your laptop. This is used to deliver advertising.

For example, a few nights ago I was browsing LLBean.com for winter boots on my iPhone. A few days later, LLBean.com ads showed up on a news blog I was reading on my iPad. This "behavioral targeting" is falling out of fashion in favor of "predictive targeting," which uses sophisticated data-mining techniques to predict for L.L.Bean whether or not I am likely to purchase something upon seeing an LLBean.com ad.

Acxiom provides "premium proprietary behavioral insights" that "number in the thousands and cover consumer interests ranging from brand and channel affinities to product usage and purchase timing." In other words, Acxiom creates profiles, or digital dossiers, about millions of people, based on the 1,500 points of data about them it claims to have. These data might include your education level; how many children you have; the type of car you drive; your stock portfolio; your recent purchases; and your race, age, and education level. These data are combined across sources—for instance, magazine subscriber lists and public

records of home ownership—to determine whether you fit into a number of predefined categories such as "McMansions and Minivans" or "adult with wealthy parent."[3] Acxiom is then able to sell these consumer profiles to its customers, who include twelve of the top fifteen credit card issuers, seven of the top ten retail banks, eight of the top ten telecom/media companies, and nine of the top ten property and casualty insurers.

Acxiom may be one of the largest data brokers, but it represents a dramatic shift in the way that personal information is handled online. The movement toward "Big Data," which uses computational techniques to find social insights in very large groupings of data, is rapidly transforming industries from health care to electoral politics. Big Data has many well-known social uses, for example by the police and by managers aiming to increase productivity. But it also poses new challenges to privacy on an unprecedented level and scale. Big Data is made up of "little data," and these little data may be deeply personal.

Alone, the fact that you purchased a bottle of cocoa butter lotion from Target is unremarkable. Target, on the other hand, assigns each customer a single Guest ID number, linked to their credit card number, e-mail address, or name. Every purchase and interaction you have with Target is then linked to your Guest ID, including the cocoa butter.

Now, Target has spent a great deal of time figuring out how to market to people about to have a baby. While most people remain fairly constant in their shopping habits—buying toilet paper here, socks there—the birth of a child is a life change that brings immense upheaval. Since birth records are public, new parents are bombarded with marketing and advertising offers. So Target's goal was to identify parents *before* the baby was born. The chief statistician for Target, Andrew Pole, said, "We knew that if we could identify [new parents] in their second trimester, there's a good chance we could capture them for years."[4] Pole had been mining immense amounts of data about the shopping habits of pregnant women and new parents. He found that women purchased certain things during their pregnancy, such as cocoa butter, calcium tablets, and large purses that could double as diaper bags.

Target then began sending targeted mail to women during their pregnancy. This backfired. Women found it *creepy*—how did Target know they were pregnant? In one famous case, the father of a teenage girl called Target to complain that it was encouraging teen pregnancy by mailing her coupons for car seats and diapers. A week later, he called back and apologized; she hadn't told her father yet that she was pregnant.[5]

So the Target managers changed their tactics. They mixed in coupons for wine and lawnmowers with those for pacifiers and Baby Wipes. Pregnant women could use the coupons without realizing that Target knew they were pregnant. As Pole told *The New York Times Magazine*, "Even if you're following the law, you can do things where people get

queasy."

These same techniques were used to great effect by the Obama campaign before the 2012 election. Famously, the campaign recruited some of the most brilliant young experts in analytics and behavioral science, and put them in a room called "the cave" for sixteen hours a day.[6] The chief data scientist for the campaign was an analyst who had formerly mined Big Data to improve supermarket promotions. This "dream team" was able to deliver microtargeted demographics to Obama—they could predict exactly how much money they would get back from each fund-raising e-mail. When the team discovered that East Coast women between thirty and forty were not donating as much as might be expected, they offered a chance to have dinner with Sarah Jessica Parker as an incentive.[7] Every evening, the campaign ran 66,000 simulations to model the state of the election. The Obama analysts were not only using cutting-edge database marketing techniques, they were developing techniques that were far beyond the state of the art.

The Obama campaign's tactics illuminate something that is often missed in our discussions of data-mining and marketing—the fact that governments and politicians are major clients of marketing agencies and data brokers. For instance, the campaign bought data on the television-watching habits of Ohioans from a company called FourthWallMedia. Each household was assigned a number, but the names of those in the household were not revealed. The Obama campaign, however, was able to combine lists of voters with lists of cable subscribers, which it could then coordinate with the supposedly anonymous ID numbers used to track the usage patterns of television set-top boxes.[8] It could then target campaign ads to the exact times that certain voters were watching television. As a result, the campaign bought airtime during unconventional programming, like *Sons of Anarchy*, *The Walking Dead*, and *Don't Trust the B—in Apt. 23*, rather than during local news programming as conventional wisdom would have advised.

The "cave dwellers" were even able to match voter lists with Facebook information, using "Facebook Connect," Facebook's sign-on technology, which is used for many sign-ups and commenting systems online. Knowing that some of these users were Obama supporters, the campaign could figure out how to get them to persuade their perhaps less motivated friends to vote. Observing lists of Facebook friends and comparing them with tagged photos, the campaign matched these "friends" with lists of persuadable voters and then mobilized Obama supporters to convince their "real-life" friends to vote.

Social Media

In view of these sophisticated data- mining and analyzing techniques, is there any way we can use social media—or the Internet itself—without adding to our profiles collected by companies like Acxiom, Experian, or Epsilon?

Social media allow us to collect and track data about ourselves. For instance, I have been using a website called Last.fm since 2005 to track every piece of digital music I have listened to when using iTunes or Spotify. As a result, I have a fascinating picture of how my musical tastes have changed over time, and Last.fm is able to recommend obscure bands to me based on this extensive listening history.

Using social media allows us to connect with friends; to learn more about ourselves; even to improve our lives. The Quantified Self movement, which builds on techniques used by women for decades, such as counting calories, promotes the use of personal data for self-knowledge. Measuring your sleep cycles over time, for instance, can help you learn to avoid caffeine after 4:00 pm, or realize that, if you want to fall asleep, you can't use the Internet for an hour before bedtime.

But these data are immensely beneficial to data brokers. Imagine how a health insurer might react to viewing your caloric intake on MyFitnessPal, the number of steps you walk per day tracked by Fitbit, how often you check in to your local gym using Foursquare, and what you eat based on the pictures of your meals that you post on Instagram. Each piece of information, by itself, may be inconsequential, but the aggregation of this information creates a larger picture. Data trackers can centrally access such information and add it to their databases. Two large consequences of this collection of data deserve more attention.

The first is data discrimination. Once customers are sliced and diced into segmented demographic categories, they can be sorted. An Acxiom presentation to the Consumer Marketing Organization in 2013 placed customers into "customer value segments" and noted that while the top 30 percent of customers add 500 percent of value, the bottom 20 percent actually cost 400 percent of value. In other words, it behooves companies to shower their top customers with attention, while ignoring the bottom 20 percent, who may spend "too much" time on customer service calls, and may cost companies in returns or coupons, or otherwise cost more than they provide.

These "low-value targets" are known in industry parlance as "waste." Joseph Turow, a University of Pennsylvania professor in communications who studies niche marketing, asks what happens to those people who fall into the categories of "waste," entirely without their knowledge or any notification. Do they suffer price discrimination? Poor service? Do they miss out on the offers given to others? Such discrimination is still more insidious because it is entirely invisible.

Second, we may be more concerned with government surveillance than with marketers or data brokers collecting personal information, but this ignores the fact that the government regularly purchases data from these companies. ChoicePoint, now owned by Elsevier, was an enormous data aggregator that combined personal data extracted from public and private

databases, including Social Security numbers, credit reports, and criminal records. It maintained 17 *billion* records on businesses and individuals, which it sold to approximately 100,000 clients, including thirty-five government agencies and seven thousand federal, state, and local law enforcement agencies.[9]

For instance, the State Department purchased records on millions of Latin American citizens, which were then checked against immigration databases. Choicepoint was also investigated for selling 145,000 personal records to an identity theft ring. More recently, Experian, one of the three major credit bureaus, mistakenly sold personal records to a Vietnamese hacker. Scammers refer to these records, which include Social Security numbers and mothers' maiden names, as "fullz," because they contain enough personal information for crooked operators to apply for credit cards or take out loans.

A few years ago, I toured the experimental lab of a large advertising agency. They showed me the cutting edge of consumer-monitoring technologies. Someday, not too far in the future, if you're at Duane Reade, aimlessly staring at a giant shelf of shampoo trying to figure out which to buy, the shelf will track your eye movements and which bottles you pick up and examine in more detail. Using this data, Duane Reade can algorithmically generate a coupon for a particular brand of shampoo, which you can then print from the shelf. I watched an experimental application that tracks the movements of individuals through a mall, based on the unique identifiers, or MAC addresses, of their cell phones, kept in purses or pockets but available to wireless tracking devices. Again, in all of these cases, the individuals are unaware that they are being tracked. A description of such procedures may be hidden at the end of a byzantine privacy policy people may not have noticed when they bought their devices, or written on a notice next to a CCTV camera. Though they may not be technically illegal, they seem ethically dubious.

While the easy answer to these problems is to opt out of loyalty cards, Internet use, or social media, this is hardly realistic. In fact, it is practically impossible to live life, online or offline, without being tracked—unless one takes extreme measures of avoidance. Cities track car movements; radio-frequency identification (RFID) tags are attached to clothing and dry cleaning; CCTV cameras are in most stores.[10] The technology is developing far more rapidly than our consumer protection laws, which in many cases are out of date and difficult to apply to our networked world.

The Federal Trade Commission and the Senate Commerce Committee are currently investigating data brokers and calling for more transparency in the collection and dissemination of personal information. Those of us concerned with privacy must continue to demand that checks and balances be applied to these private corporations. People should be encouraged to investigate the various opt-out tools, ad-blockers, and plug-ins that are available for most platforms. While closer scrutiny of the NSA is necessary and needed, we

must apply equal pressure to private corporations to ensure that seemingly harmless targeted mail campaigns and advertisements do not give way to insidious and dangerous violations of personal privacy.

---

1 See Natasha Singer, "Acxiom, the Quiet Giant of Consumer Database Marketing," *The New York Times*, June 16, 2012. ↩

2 See Judith Aquino, " Acxiom Prepares New 'Audience Operating System' Amid Wobbly Earnings," *AdExchanger.com*, August 1, 2013. ↩

3 See Natasha Singer, "A Data Broker Offers a Peek Behind the Curtain," *The New York Times*, August 31, 2013. ↩

4 See Charles Duhigg, "How Companies Learn Your Secrets," *The New York Times Magazine*, February 16, 2012. ↩

5 See Kashmir Hill, " How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did," *Forbes*, February 16, 2012. ↩

6 See Jim Rutenberg, " Data You Can Believe In: The Obama Campaign's Digital Masterminds Cash In," *The New York Times Magazine*, June 20, 2013. ↩

7 See Michael Scherer, " Inside the Secret World of the Data Crunchers Who Helped Obama Win," *Time*, November 7, 2012. ↩

8 See Lois Beckett, " Everything We Know (So Far) About Obama's Big Data Tactics," *ProPublica*, November 29, 2012. ↩

9 See " ChoicePoint," at the Electronic Privacy Information Center. ↩

10 See Sarah Kessler, " Think You Can Live Offline Without Being Tracked? Here's What It Takes," *Fast Company*, October 15, 2013. ↩