HARVARD | BUSINESS | SCHOOL

9-515-090
REV: APRIL 20, 2016

JOHN DEIGHTON
LEORA KORNFELD

Target Stores: The Hunt for "Unvolunteered Truths"

New York Times reporter Charles Duhigg's research for an article on predictive analytics in marketing had led him to Target Stores' corporate headquarters at Target Plaza in Minneapolis, Minnesota. He approached the reception desk. How would, and indeed, how should, Target respond to his inquiries?

Background

Duhigg's article in the *New York Times Magazine* in 2012 reported on an angry father's words to a Target store manager outside Minneapolis. The article read:

> "My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?" The manager didn't have any idea what the man was talking about. He looked at the mailer. Sure enough, it was addressed to the man's daughter and contained advertisements for maternity clothing, nursery furniture and pictures of smiling infants. The manager apologized and then called a few days later to apologize again. On the phone, though, the father was somewhat abashed. "I had a talk with my daughter," he said. "It turns out there's been some activities in my house I haven't been completely aware of. She's due in August. I owe you an apology."[1]

The article, under the headline "How Companies Learn Your Secrets," started a media storm that would churn for several years. It was written by Duhigg after he had viewed a year-old video titled, "How Target Gets the Most out of Its Guest Data," delivered by a Target staff statistician, Andrew Pole, at a marketing industry event called Predictive Analytics World.[2] *Forbes* magazine retold Duhigg's story under a more arresting headline, "How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did,"[3] and other media followed suit under progressively more lurid headlines. The technology website Business Insider trumpeted, "The Incredible Story of How Target Exposed a Teen Girl's Pregnancy."[4] Simply typing "Target p" into Google three years later produced the auto-completed suggestion "Target Pregnant" and images of pregnant bellies painted with the Target logo. (See **Exhibit 1**.)

HBS Professor John Deighton and Leora Kornfeld (Schulich School of Business) prepared this case. This case was developed from published sources. Funding for the development of this case was provided by Harvard Business School and not by the company. HBS cases are developed solely as the basis for class discussion. Cases are not intended to serve as endorsements, sources of primary data, or illustrations of effective or ineffective management.

Eric Siegel, an industry analyst, provided some perspective on how the *New York Times* article had come to be written. In a book published in 2013, he told how he had invited Pole to speak at Predictive Analytics World and had introduced him to Duhigg. Siegel wrote:

> Pole manages dozens of analytics professionals who run various PA [predictive analytics] projects at Target. In October of that year, Pole delivered a stellar keynote on a wide range of PA deployments at Target. He took the stage and dynamically engaged the audience, revealing detailed examples, interesting stories, and meaningful business results that left the audience clearly enthused. Free to view, here it is: www.pawcon.com/Target.
>
> A few months after Pole's presentation, *New York Times* reporter Charles Duhigg interviewed me. Exploring, he asked for interesting discoveries that had come from PA. I rattled off a few and included pregnancy prediction, pointing him to the online video of Pole's talk, which had thus far been receiving little attention, and introducing him to Pole. I must admit that by now the privacy question had left my mind almost entirely.
>
> One year later, in February 2012, Duhigg published a front-page[a] *New York Times Magazine* article . . . alleging an anonymous story of a man discovering his teenage daughter is pregnant only by seeing Target's marketing offers to her, with the unsubstantiated but tacit implication that this resulted specifically from Target's PA project.
>
> This well-engineered splash triggered rote repetition by press, radio, and television, all of whom blindly took as gospel what had only been implied and ran with it. Not incidentally, it helped launch Duhigg's book, *The Power of Habit: Why We Do What We Do in Life and Business* . . . which hit the *New York Times* best seller list.[5]

Target Stores

Target traced its roots to a Minneapolis department store, R.S. Goodfellow & Company, founded in 1878. George Draper Dayton acquired it in 1902, and it became the first of a chain of Dayton Department Stores, later renamed Dayton-Hudson and later still, Marshall Field. In 2005, the chain was subsumed into the Macy's department store chain.

Target Stores was founded in 1962 as a discount division of Dayton's. It grew rapidly and soon outstripped the revenues of its parent. It launched an e-commerce website, Target.com, in 1999. In 2004, Target Stores was spun off. In 2010, the year in which its predictive analytics practices were discussed by Andrew Pole, it had revenues of $65 billion from 1,740 stores. By 2015, revenues had grown to $73 billion. (See **Exhibit 2** for Target revenues, 2001–2014.)

In 2013, in the run-up to the Christmas shopping season, Target Stores experienced a significant data breach. Hackers installed malware in Target's security and payments system and captured 40 million credit card numbers and 70 million addresses, phone numbers, and other personal information. Breach-related expenses by the end of January 2015 were $252 million, partly offset by insurance payments of $90 million.

---

[a] The article appeared on p. 30 of the magazine.

Target was America's second-largest big-box discount retailer behind Walmart ($279 billion U.S. sales). It used stylish merchandise to distinguish itself from Walmart, commissioning clothing lines from fashion designers such as Jason Wu, Zac Posen, and Isaac Mizrahi, and home hardware designs from the architect Michael Graves. As far back as 1970, it had played with the illusion that the name might be vaguely French and pronounced Tar-*zhay*, selling a line of shoes branded Miss Targé and registering the brand Targét Couture.

Target's data-gathering and analysis practices were sophisticated. Long before big data was fashionable, it used large volumes of data generated by its customers' behavior. The data came from a number of sources, including:

- Credit cards
- Guest cards that could be used only in Target stores
- Check-cashing cards
- Coupon usage
- Gift registry
- Customer service calls
- Website visits

Duhigg's investigations found that Target, in common with most retailers, also purchased third-party data about its shoppers, some of it personally identifiable, some of it identifiable only at the aggregate level. As he explained in the *New York Times* article:

> [T]hey can buy data about your ethnicity, job history, the magazines you read, if you've ever declared bankruptcy or got divorced, the year you bought (or lost) your house, where you went to college, what kinds of topics you talk about online, whether you prefer certain brands of coffee, paper towels, cereal or applesauce, your political leanings, reading habits, charitable giving and the number of cars you own.[6]

Important life events—marriage, the birth of a child, the move to a new home—were moments when shoppers were unusually responsive to the marketing efforts of retailers because new shopping habits formed at these times. Target wanted to try to get its shoppers to patronize Target for more than just clothing and hard goods: it was out to capture sales of items such as health and beauty products and grocery products from drugstores and supermarkets. Pole had explained it to Duhigg this way:

> We want the information earlier than when it's made public, which is through birth records. Ideally we'd like to identify when pregnant women were in their second trimester, as that's when they start buying things such as prenatal vitamins and maternity clothing. As soon as we get them buying diapers from us, they're going to start buying everything else too. If you're rushing through the store, looking for bottles, and you pass orange juice, you'll grab a carton. . . . Soon, you'll be buying cereal and paper towels from us, and keep coming back.[7]

Pole and his team knew that being conspicuous in the way ads and offers were targeted could backfire, so they devised systems that made targeted offers seem more random than they actually were. Pole described the process to Duhigg:

> [W]e started mixing in all these ads for things we knew pregnant women would never buy, so the baby ads looked random. We'd put an ad for a lawn mower next to diapers . . . and we found out that as long as a pregnant woman thinks she hasn't been spied on, she'll use the coupons. She just assumes everyone else on her block got the same mailer for diapers and cribs.[8]

Pole's Presentation to Predictive Analytics World

In October 2010, Pole delivered a keynote address at Predictive Analytics World, a conference for specialists in the field. He was head of a team of 50, of whom 25 were in Minneapolis and 25 in India, who performed analytics, modeling, campaign execution, and custom reporting for a number of marketing functions—online marketing, direct marketing, point-of-sale couponing, e-mail, banner advertising, and search engine marketing. They also gave input to the design of the printed circulars that were inserted into newspapers and available to shoppers in stores, and to broadcast mass media advertising. They developed and estimated models that attempted to show the effect of online and offline media elements on sales, termed "media mix modeling." He explained the team's overriding mission:

> When my team sits down to do an analysis, we try to keep the *end* in mind, and that is to be *relevant* for our guests. We want to meet those guests' expectations. They've changed. It's probably not a surprise to anybody but today's consumers have taken control.[9]

He explained that there were three primary kinds of guest expectation:

- To have her needs understood.
- To receive relevant messages: "I don't want to send her a coupon for baby diapers if she doesn't have a baby."
- To be contacted by the right communication vehicles: "If she does not open email, I won't email her."[10]

He said that, wherever possible, he wanted to know the expectations of particular guests, not segments. So he tried to attach each expectation to a Guest ID. "Our key," he explained, "is the Guest ID. It's just a big, gigantic number in our database. We try to tie everything back to this key." The Guest ID was the stable identifier that sought to recognize people across multiple encounters, whether store visits, online transactions, or browsing, and, relying on the guest's contact history and response history across all media, to select the best way to communicate with the guest in the future. He explained how the Guest ID could be linked to a name and address if the guest paid at checkout with a credit or debit card, or in a variety of other ways to an e-mail, a cookie, and/or a mobile ID.

> You get an email, which says we've got a great new clothing line you might want to check out. You click through and you browse around. Guess what just happened? You clicked through an email, which tied that email ID to a cookie, which then could be tied to a Guest ID.[11]

A Guest ID could be found for over half of all store transactions, for almost all online transactions, but for less than one quarter of online cookies found on the browsers of people who visited the Target website. When a cookie could be matched to a Guest ID, the browsing history could be stored. He referred to the pattern of needs, message preferences, and communication preferences as the guest's portrait. He went on:

> Our guests expect us to use the information on what they have done to give them relevant offers. If 24 hours ago you went and did a lot of looking at a particular brand of clothing line at Target.com, and next day I send you a coupon for potato chips, guests are going to get ticked off. "I went to Facebook, I talked about how much I liked this clothing line." They expect us to do something about it and make those messages relevant.[12]

Step-by-step, Pole built a slide (**Exhibit 3**) that showed how the portrait of a particular guest was constructed over time across multiple points of customer contact. When it was complete, he advanced to a slide titled "A Word About Compliance." He paused, and said:

> Some of you may be cringing a little bit from that previous slide, right? There's a lot of big brother? I want to be honest and say, right, there's a lot of information. [Choosing his words, he went on.] Target is very, very conservative about compliance. We do not want to go out and get information that you have told us explicitly that we can't have, or that state or federal laws say we cannot have. . . . We'd rather be safe than sorry. We don't want to take a chance that someone would interpret us as breaking the law or not respecting those data privacy rules. And it's tough, right, because as data miners more data is better. And there are times when we are out there, we build all these models with all this great information, and all of a sudden my data quality partner comes over and says, "You know, Andrew, all that driver's license data we were using, we can't use any of it now." [His body slumped in a caricature of despair.] We'll just go rebuild all the models that we built, and without the driver's license information.
>
> Target wants to make sure that we don't end up in the newspapers or on TV because we went out and used something that we were not supposed to be using.[13]

His presentation emphasized three technical challenges in building predictive models:

- They had to join disparate data sources quickly, such as browser data from Target.com, media contact history data, and store transaction data.
- They had to be flexible enough to accept a new source of data as soon as it became available.

  That's been a big challenge to us at Target lately. We turn to our IT partners, who are great and doing an awesome job, but they go, "Andrew! Every time I turn around and finish a project you come to us and you say you've got a new data source and I've got to rebuild my solution." And my answer is you've got to build one that's flexible.

- They had to use standard formats across channels. The product hierarchy on Target.com had to be reconcilable with the hierarchy in the store.[14]

Pole wrapped up his talk with examples of how he used guest profiles to improve Target's marketing. One example, the Mom and Baby Acquisition Mailer, began on a flippant note:

> We're not trying to acquire babies, right? This is not an illegal thing, right. We're trying to acquire guests. Ideally we want to acquire and convert prenatal mothers *before* they have their baby. How the heck do we do that? Well, we develop a model to predict whether the mother is likely pregnant with child. We've done this, and it's fairly accurate. It's not our best model, but it's fairly accurate. What do we look at? Well, we've found that prenatal mothers, expectant mothers, they start nesting. They start buying things in preparation for their child, and they buy them at fairly regular intervals. So if I can say that (I'll make something up here) mothers will buy the baby crib 90 days before the due date, and I see you buy it on 1st November, what is the estimated due date? Now imagine I've got 25 key products. Now I have 25 due dates, and they're all in this range. What do you do with that? The result is in our database. We have people who have registered (so they're definitely pregnant) and we have other people through other programs who have told us that they're pregnant. But now all of a sudden I've got 30% more guests that I'm pretty sure are pregnant. In fact I know this one's in the second trimester and this one's in the third trimester, so I even know what to offer them. This is a very profitable acquisition mailer, and to find 30% more guests was a *big* win for us.[15]

Duhigg's interviews with Pole supplied more details about the modeling process. Analysis had identified about 25 products that, when incorporated into the prediction equation, let it assign each shopper a "pregnancy prediction" score. More important, it could estimate her due date to within a small window, so that Target could send coupons timed to very specific stages of her pregnancy. Pole's colleagues noticed that women on the baby registry were buying larger quantities of unscented lotion around the beginning of their second trimester. Another analyst noted that sometime in the first 20 weeks, pregnant women loaded up on supplements like calcium, magnesium, and zinc. Many shoppers purchased soap and cotton balls, but when someone suddenly started buying lots of scent-free soap and extra-big bags of cotton balls, in addition to hand sanitizers and washcloths, it signaled that they were close to their delivery date.[16]

Unvolunteered Truths

Writing about the Target incident, Eric Siegel coined the term "unvolunteered truths" to distinguish things that a consumer knew but was keeping private from future things that no one knew with certainty.[17] The methodology for making predictions, known to marketing professionals as predictive analytics, involved forecasting future behavior from data on past behavior augmented with data on demographics such as age, gender, income level, geographical location, and socioeconomic status. Discovering unvolunteered information used an identical methodology, so the distinction was not in the method of the procedure but in the ethics of doing so.

The distinction began to enter public discussion on the ethics of data collection and surveillance when a journalist invoked it in an article under the headline, "What's the NSA Picking Out of Your Phone Calls? Just 'Unvolunteered Truths.'"[18] In 2013, Edward Snowden, a computer programmer, came forward with the information that, as a subcontractor for the National Security Agency (NSA), he had seen the NSA collect data on people that he believed was outside the jurisdiction of the agency. He reported deals with communications technology companies that enabled routine surveillance of domestic technologies ranging from cellphones and laptop computers to social media, Skype, and online chat services. Plainly put, people's cellphone calls, e-mail conversations, and online chats were being listened to by the government without their knowledge. A common justification for these data-collection practices was that people with nothing to hide had nothing to fear.

The journalist explained that digital inference "allows the NSA or law enforcement agencies to derive actionable intelligence about targets that those targets have not disclosed, as well as information that may be more valuable than direct surveillance—it's spying 2.0."[19]

A Princeton computer science professor, Edward Felten, in a court filing written to support the American Civil Liberties Union opposition to the NSA's gathering of phone metadata (numbers called, and time, date, and duration of calls) argued:

> Sophisticated computing tools permit the analysis of large datasets to identify embedded patterns and relationships, including personal details, habits, and behaviors. As a result, individual pieces of data that previously carried less potential to expose private information may now, in the aggregate, reveal sensitive details about our everyday lives—details that we had no intent or expectation of sharing.[20]

Felten went on to contend that to analyze the content of phone calls:

> [T]he government would first have to transcribe the calls and then determine which parts of the conversation are interesting and relevant. . . . Assuming that a call is transcribed correctly, the government must still try to determine the meaning of the conversation: When a surveillance target is recorded saying "the package will be delivered next week," are they talking about an order they placed from an online retailer, a shipment of drugs being sent through the mail, or a terrorist attack? Parsing and interpreting such information, even when performed manually, is exceptionally difficult.[21]

By contrast, the mere pattern of calls could be more informative than the content of the calls themselves:

> A young woman calls her gynecologist; then immediately calls her mother; then a man who, during the past few months, she had repeatedly spoken to on the telephone after 11pm; followed by a call to a family planning center that also offers abortions. A likely storyline emerges that would not be as evident by examining the record of a single telephone call.[22]

Pregnancy and terrorism were by no means the only kinds of unvolunteered information constructed from data gathered by data brokers, processed by analysts, and linked back to digital and physical addresses. It was straightforward to buy lists of online identifiers (including e-mail addresses) and postal addresses of people who had, for example, read romance novels, donated to arts or cultural causes, searched online for allergy relief or certain prescription drugs, were registered Democrat or Republican, or had donated to political campaigns and in what amounts. Models existed to infer information for which no lists existed, such as monthly income. This information was used mainly for marketing, to target prospects for offers for example, but some was used to verify a person's claimed identity to prevent fraud or during the course of employment background checks.

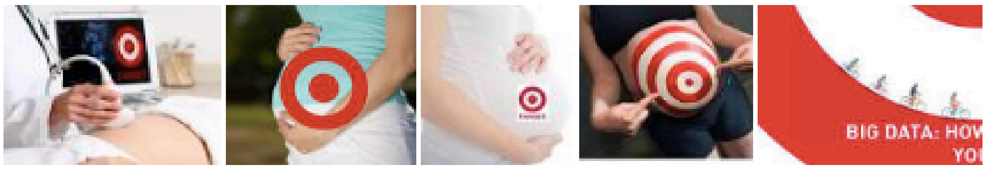
Conclusion

When Charles Duhigg asked Target for comment on the story he was writing, Target responded by denying him further access to the company or to Pole. Duhigg wrote:

> When I approached Target to discuss Pole's work, its representatives declined to speak with me. "Our mission is to make Target the preferred shopping destination for our guests by delivering outstanding value, continuous innovation and exceptional guest experience," the company wrote in a statement. "We've developed a number of research tools that allow us to gain insights into trends and preferences within different demographic segments of our guest population."
>
> When I sent Target a complete summary of my reporting, the reply was more terse: "Almost all of your statements contain inaccurate information and publishing them would be misleading to the public. We do not intend to address each statement point by point." The company declined to identify what was inaccurate. They did add, however, that Target "is in compliance with all federal and state laws, including those related to protected health information."
>
> When I offered to fly to Target's headquarters to discuss its concerns, a spokeswoman emailed that no one would meet me. When I flew out anyway, I was told I was on a list of prohibited visitors. "I've been instructed not to give you access and to ask you to leave," said a very nice security guard named Alex.[23]

Exhibit 1 A Google Search for the Keywords "Target Pregnant"

Images for target pregnant Report images



More images for target pregnant

How Companies Learn Your Secrets - NYTimes.com
www.nytimes.com/2012/02/19/magazine/shopping-habits.html?...all
Feb 16, 2012 - Using data to predict a woman's **pregnancy**, **Target** realized soon after ... But for **pregnant** women, **Target's** goal was selling them baby items ...
You visited this page on 23/02/15.

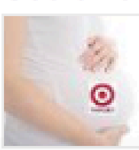
Lessons from Target's pregnancy prediction PR fiasco ...
https://www.linkedin.com/.../20140616204813-2554671-lessons-from-**ta**...
Jun 16, 2014 - For example, when **Target** started the **pregnancy** prediction project, it faced a bit of a backlash when customers started receiving direct mail or ...

How Target knows when its shoppers are pregnant - and ...
www.dailymail.co.uk/.../How-**Target**-knows-shoppers-**pregnant**--figured...
Feb 18, 2012 - Clues such as vitamin supplements, large quantities of lotion, and hand sanitizers, typical to many **pregnant** women according to the **Target** ...

How Target Knew a High School Girl Was Pregnant Before ...
techland.time.com/.../how-**target**-knew-a-high-school-girl-was-**pregnant**-...
Feb 17, 2012 - A father found himself in the uncomfortable position of having to apologize to a **Target** employee. Earlier he had stormed into a store near ...

In-depth articles

Could Target Sell Its 'Pregnancy Prediction Score'?



Forbes - Feb 2012
Target assigns every one of its customers a "**pregnancy** prediction score," with an estimate of the due date, so that coupons can be timed to the right stage of **pregnancy** (e.g., maternity ...
Explore: pregnancy

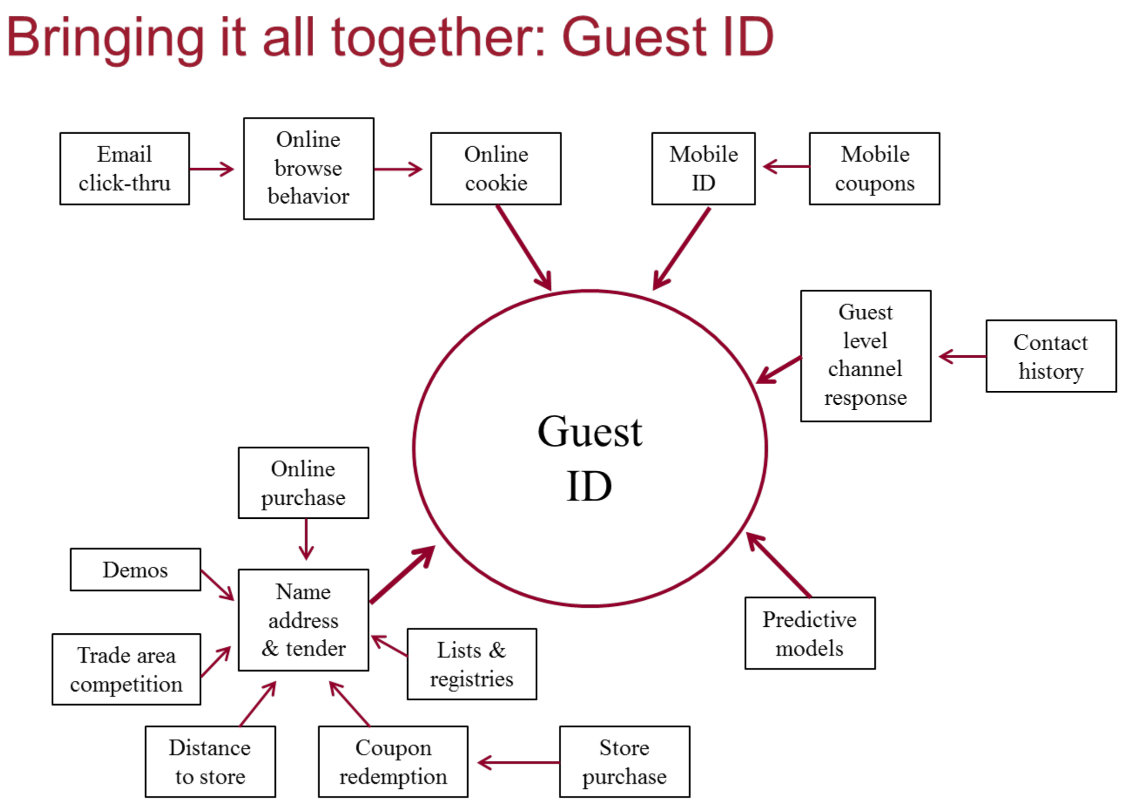
Source: Google search results for "target pregnant," accessed February 25, 2015.

Exhibit 2 Target Revenues, 2001–2014

Year	Revenues (millions)	Cost of Sales (millions)	Earnings from Continuing Operations before Interest and Taxes
2001	$33,021	$23,030	$1,101
. . .	. . .	. . .	. . .
2009	$65,357	$44,062	$4,673
2010	$67,390	$45,725	$5,252
2011	$69,865	$47,860	$5,443
2012	$73,301	$50,568	$5,740
2013	$71,279	$50,039	$5,170
2014	$73,618	$51,278	$4,535

Source: http://investors.target.com/phoenix.zhtml?c=65828&p=irol-reportsAnnual, accessed February 27, 2015.

Exhibit 3 A Slide from Andrew Pole's Presentation



Source: Created by casewriter to replicate a slide in the presentation by Andrew Pole on October 19, 2010, www.pawcon.com/Target, accessed March 25, 2015.

Endnotes

[1] Charles Duhigg, "How Companies Learn Your Secrets," *New York Times*, February 19, 2012, http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html, accessed February 24, 2015.

[2] Andrew Pole, "How Target Gets the Most out of Its Guest Data," lecture, Predictive Analytics World, October 2010, http://www.rmportal.performedia.com/node/1373, accessed March 2, 2015.

[3] Kashmir Hill, "How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did," *Forbes*, February 16, 2012, http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/, accessed February 24, 2015.

[4] Gus Lubin, "The Incredible Story of How Target Exposed A Teen Girl's Pregnancy," Business Insider, February 16, 2012, http://www.businessinsider.com/the-incredible-story-of-how-target-exposed-a-teen-girls-pregnancy-2012-2, accessed March 2, 2015.

[5] Eric Siegel, *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* (Hoboken, NJ: John Wiley & Sons, 2013).

[6] Duhigg, "How Companies Learn Your Secrets."

[7] Ibid.

[8] Ibid.

[9] Pole, "How Target Gets the Most out of Its Guest Data."

[10] Ibid.

[11] Ibid.

[12] Ibid.

[13] Ibid.

[14] Ibid.

[15] Ibid.

[16] Duhigg, "How Companies Learn Your Secrets."

[17] Siegel, *Predictive Analytics.*

[18] Andrew Couts, "What's the NSA Picking Out of Your Phone Calls? Just Unvolunteered Truths," Digital Trends, August 31, 2013, http://www.digitaltrends.com/mobile/whats-the-nsa-picking-out-of-your-phone-calls-just-unvolunteered-truths/, accessed April 19, 2014.

[19] Ibid.

[20] Ibid.

[21] Ibid.

[22] Ibid.

[23] Duhigg, "How Companies Learn Your Secrets."