

The New York Review of Books

They Have, Right Now, Another You

Sue Halpern

DECEMBER 22, 2016 ISSUE

Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy

by Cathy O'Neil
Crown, 259 pp., \$26.00

Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy

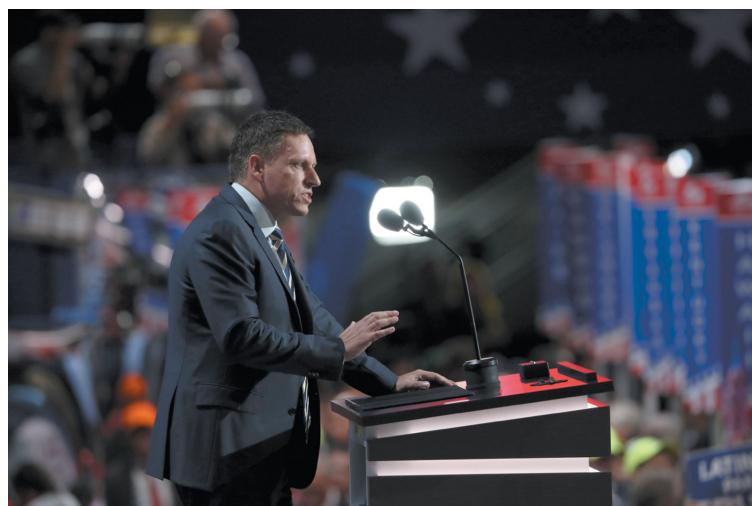
by Ariel Ezrachi and Maurice E. Stucke
Harvard University Press, 356 pp., \$29.95

A few months ago *The Washington Post* reported that Facebook collects ninety-eight data points on each of its nearly two billion users. Among this ninety-eight are ethnicity, income, net worth, home value, if you are a mom, if you are a soccer mom, if you are married, the number of lines of credit you have, if you are interested in Ramadan, when you bought your car, and on and on and on.

How and where does Facebook acquire these bits and pieces of one's personal life and identity? First, from information users volunteer, like relationship status, age, and university affiliation. They also come from Facebook posts of vacation pictures and baby pictures and graduation pictures. These do not have to be photos one posts oneself: Facebook's facial recognition software can pick you out of a crowd. Facebook also follows users across the Internet, disregarding their "do not track" settings as it stalks them. It knows every time a user visits a website that has a Facebook "like" button, for example, which most websites do.

The company also buys personal information from some of the five thousand data brokers worldwide, who collect information from store loyalty cards, warranties, pharmacy records, pay stubs, and some of the ten million public data sets available for harvest. Municipalities also sell data—voter registrations and motor vehicle information, for example, and death notices, foreclosure declarations, and business registrations, to name a few. In theory, all these data points are being collected by Facebook in order to tailor ads to sell us stuff we want, but in fact they are being sold by Facebook to advertisers for the simple reason that the company can make a lot of money doing so.

Not long ago I dug into the depths of Facebook to see what information it was using to tailor ads for me. This is a different set of preferences and a different algorithm—a set of instructions to carry out an operation—than the one Facebook uses to determine which stories it is going to display on my so-called news feed, the ever-changing assortment of photos and posts from my Facebook friends and from websites I've "liked." These ad preferences are the coin of the Facebook realm; the company made \$2.3 billion in the third quarter of 2016 alone, up from about \$900 million in the same three months last year.



Stephen Crowley/The New York Times/Redux

Peter Thiel speaking at the Republican National Convention, Cleveland, July 2016. Thiel, the first outside investor in Facebook and a cofounder of PayPal, is a founder of Palantir, a Silicon Valley firm funded by the CIA, whose algorithms allow for rapid analysis of voluminous data that it makes available to intelligence agencies and numerous police forces as well as to corporations and financial institutions.

And here is some of what I discovered about myself according to Facebook:

That I am interested in the categories of “farm, money, the Republican Party, happiness, gummy candy, and flight attendants” based on what Facebook says I do on Facebook itself. Based on ads Facebook believes I’ve looked at somewhere—anywhere—in my Internet travels, I’m also interested in magnetic resonance imaging, *The Cave of Forgotten Dreams*, and thriller movies. Facebook also believes I have liked Facebook pages devoted to *Tyrannosaurus rex*, Puffy AmiYumi, cookie dough, and a wrestler named the Edge.

But I did not like any of those pages, as a quick scan of my “liked” pages would show. Until I did this research, I had never heard of the Edge or the Japanese duo Puffy AmiYumi, and as someone with celiac disease, I am constitutionally unable to like cookie dough. I did “like” the page of the Flint, Michigan, female boxing sensation Claressa Shields, whose nickname is “T-Rex.” And that is as close as Facebook got to matching my actual likes to the categories it says—to advertisers—that I’m keen on.

And this is odd, because if there is one incontrovertible thing that Facebook knows about me, it’s the Facebook pages that I have actively liked. But maybe I am more valuable to Facebook if I am presented as someone who likes Puffy AmiYumi, with its tens of thousands of fans, rather than a local band called Dugway, which has less than a thousand. But I will never know, since the composition of Facebook’s algorithms, like Google’s and other tech companies’, is a closely guarded secret.

While Facebook appears to be making seriously wrong and misdirected assumptions about me, and then cashing in on those mistakes, it is hardly alone in using its raw data to come to strange and wildly erroneous assumptions. Researchers at the Psychometrics Centre at Cambridge University in England have developed what they call a “predictor engine,” fueled by algorithms using a subset of a person’s Facebook “likes” that “can forecast a range of variables that includes happiness, intelligence, political orientation and more, as well as generate a big five personality profile.” (The big five are extroversion, agreeableness, openness, conscientiousness, and neuroticism, and are used by, among others, employers to assess job applicants. The acronym for these is OCEAN.) According to the Cambridge researchers, “we always think beyond the mere clicks or Likes of an individual to consider the subtle attributes that really drive their behavior.” The researchers sell their services to businesses with the promise of enabling “instant psychological assessment of your users based on their online behavior, so you can offer real-time feedback and recommendations that set your brand apart.”

So here’s what their prediction engine came up with for me: that I am probably male, though “liking” *The New York Review of Books* page makes me more “feminine”; that I am slightly more conservative than liberal—and this despite my stated affection for Bernie Sanders on Facebook; that I am much more contemplative than engaged with the outside world—and this though I have “liked” a number of political and activist groups; and that, apparently, I am more relaxed and laid back than 62 percent of the population. (Questionable.)

Here’s what else I found out about myself. Not only am I male, but “six out of ten men with [my] likes are gay,” which gives me “around an average probability” of being not just male, but a gay male. The likes that make me appear “less gay” are the product testing magazine *Consumer Reports*, the tech blog *Gizmodo*, and another website called *Lifehacker*. The ones that make me appear “more gay” are *The New York Times* and the environmental group 350.org. Meanwhile, the likes that make me “appear less interested in politics” are *The New York Times* and 350.org.

And there’s more. According to the algorithm of the Psychometrics Centre, “Your likes suggest you are single and not in a relationship.” Why? Because I’ve liked the page for 350.org, an organization founded by the man with whom I’ve been in a relationship for thirty years!

Amusing as this is, it’s also an object lesson, yet again, about how easy it is to misconstrue and misinterpret data. We live at a moment when very powerful computers can parse and sort very large and disparate data sets. This can lead us to see patterns where we couldn’t see them before, which has been useful for drug discovery, for example, and, apparently, for figuring out where IEDs were most likely to be planted in Afghanistan, but it can also lead us to the belief that data analysis will deliver to us a truth that is free of messiness, idiosyncrasy, and slant.

In fact, the datafication of everything is reductive. For a start, it leaves behind whatever can't be quantified. And as Cathy O'Neil points out in her insightful and disturbing book *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, datafication often relies on proxies—stand-ins that can be enumerated—that bear little or no relation to the things they are supposed to represent: credit scores as a proxy for the likelihood of being a good employee, for example, or “big five” personality tests like the ones used by the Cambridge Psychometrics Centre, even though, as O'Neil reports, “research suggests that personality tests are poor predictors of job performance.”

There is a tendency to assume that data is neutral, that it does not reflect inherent biases. Most people, for instance, believe that Facebook does not mediate what appears in one's “news feed,” even though Facebook's proprietary algorithm does just that. Someone—a person or a group of people—decides what information should be included in an algorithm, and how it should be weighted, just as a person or group of people decides what to include in a data set, or what data sets to include in an analysis. That person or group of people come to their task with all the biases and cultural encumbrances that make us who we are. Someone at the Cambridge Psychometrics Centre decided that people who read *The New York Review of Books* are feminine and people who read tech blogs are masculine. This is not science, it is presumption. And it is baked right into the algorithm.

We need to recognize that the fallibility of human beings is written into the algorithms that humans write. While this may be obvious when we're looking at something like the Cambridge Psychometrics analysis, it is less obvious when we're dealing with algorithms that “predict” who will commit a crime in the future, for example—which in some jurisdictions is now factored into sentencing and parole decisions—or the algorithms that deem a prospective employee too inquisitive and thus less likely to be a loyal employee, or the algorithms that determine credit ratings, which, as we've seen, are used for much more than determining creditworthiness. (Facebook is developing its own credit-rating algorithm based on whom one associates with on Facebook. This might benefit poor people whose friends work in finance yet penalize those whose friends are struggling artists—or just struggling.)

Recently, some programmers decided to hold an online global beauty pageant, judged by a computer outfitted with artificial intelligence. The idea was that the computer would be able to look at the photographs uploaded by thousands of people across the globe and, in an unbiased way, find those women who represented ideal beauty. Should we have been surprised when, almost to a person, the women judged most beautiful were white? The algorithm used by the computer was developed by programmers who “trained” the computer using datasets of photos of primarily white women. In choosing those photos, the programmers had determined a standard of beauty that the computer then executed. “Although the group did not build the algorithm to treat light skin as a sign of beauty,” Sam Levin wrote in *The Guardian*, “the input data effectively led the robot judges to reach that conclusion.”

When Harvard professor Latanya Sweeney looked at 120,000 Google AdWords buys by companies that provide criminal background checks, she found that when someone did a Google search for individuals whose names were typically considered to be black, the search came back with an ad suggesting he or she had a criminal background. And then there was the case of the historically black fraternity Omega Psi Phi, which created a website to celebrate its hundredth anniversary. As Ariel Ezrachi and Maurice Stucke report in *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*, “Among the algorithm-generated ads on the website were ads for low-quality, highly criticized credit cards and ads that suggested the audience member had an arrest record.”

Advertisements show up on our Internet browser or Facebook page or Gmail and we tend to think they are there because some company is trying to sell us something it believes we want based on our browsing history or what we've said in an e-mail or what we were searching for on Google. We probably don't think they are there because we live in a particular neighborhood, or hang out with certain kinds of people, or that we have been scored a particular and obscure way by a pointillist rendering of our lives. And most likely, we don't imagine we are seeing those ads because an algorithm has determined that we are losers or easy marks or members of a particular ethnic or racial group.

As O'Neil points out, preferences and habits and zip codes and status updates are also used to create predatory ads, “ads that pinpoint people in great need and sell them false or overpriced promises.” People with poor credit may be offered payday loans; people with dead-end jobs may be offered expensive courses at for-profit colleges. The idea, O'Neil writes,

“is to locate the most vulnerable people and then use their private information against them. This involves finding where they suffer the most, which is known as the ‘pain point.’”

We have known for years that Internet commerce sites like Amazon and travel companies like Orbitz and Expedia price items according to who they say we are—where we live, our incomes, our previous purchases. And often, paradoxically, the rich pay less. Or in the case of Asian high school students signing up for Princeton Review college testing courses, or Orbitz patrons logging in on Mac computers, they pay more. Such dynamic pricing is getting more sophisticated and even more opaque. A British retailer, for example, is testing electronic price tags that display an item’s price based on who is looking at it, which it knows from the customer’s mobile phone, just as it knows that customer’s previous spending habits, also from the phone. Facebook may have ninety-eight data points on each user, but the data brokerage Acxiom has 1,500, and they are all for sale to be aggregated and diced and tossed into formulas beyond our reach.

We give our data away. We give it away in drips and drops, not thinking that data brokers will collect it and sell it, let alone that it will be used against us. There are now private, unregulated DNA databases culled, in part, from DNA samples people supply to genealogical websites in pursuit of their ancestry. These samples are available online to be compared with crime scene DNA without a warrant or court order. (Police are also amassing their own DNA databases by swabbing cheeks during routine stops.) In the estimation of the Electronic Frontier Foundation, this will make it more likely that people will be implicated in crimes they did not commit.

Or consider the data from fitness trackers, like Fitbit. As reported in *The Intercept*:

During a 2013 FTC panel on “Connected Health and Fitness,” University of Colorado law professor Scott Peppet said, “I can paint an incredibly detailed and rich picture of who you are based on your Fitbit data,” adding, “That data is so high quality that I can do things like price insurance premiums or I could probably evaluate your credit score incredibly accurately.”

Consider, too, that if you take one of the random personality quizzes that consistently show up on Facebook—“What your handwriting says about you”—there’s a good chance it will be used by a company called Cambridge Analytica to gain access not only to your OCEAN score but to your Facebook profile, including your name. (According to *The New York Times*, Cambridge Analytica was advising the Trump campaign.)

Meanwhile, every time you hail an Uber car or use Google Maps, to name two mobile applications, you are revealing your location and leaving a trail for others—certainly the police, possibly hackers and other criminals, and definitely commercial interests—to follow and exploit. Not long ago I was at a restaurant in New York when I got a message congratulating me for my choice of dining venues and informing me of the day’s specials. Though I hadn’t used Google Maps to get there, just by having location services activated on my phone I was fair game—a sitting duck.

Aside from the creepy factor, does it matter? That’s the question we need to ask ourselves and one another.

Chances are, if you query most people who use Facebook or Google products or ride in Uber cars or post selfies on Twitter if they mind that their personal information is being sold like the commodity it is, they will tell you that this is a small and largely inconsequential price to pay for the convenience of free turn-by-turn directions or e-mail or staying in touch with old friends. Chances are they will tell you that handing over bits and pieces of personal information is the cost of doing business, even when the real business is not what they are getting but what they are handing over.

If it is true, as Mark Zuckerberg has said, that privacy is no longer a social norm, at what point does it also cease to be a political norm? At what point does the primacy of the individual over the state, or civil liberties, or limited government also slip away? Because it would be naive to think that governments are not interested in our buying habits, or where we were at 4 PM yesterday, or who our friends are. Intelligence agencies and the police buy data from brokers, too. They do it to bypass laws that restrict their own ability to collect personal data; they do it because it is cheap; and they do it because commercial databases are multifaceted, powerful, and robust.

Moreover, the enormous data trail that we leave when we use Gmail, post pictures to the Internet, store our work on

Google Drive, and employ Uber is available to be subpoenaed by law enforcement. Sometimes, though, private information is simply handed over by tech companies, no questions asked, as we learned not long ago when we found out that Yahoo was monitoring all incoming e-mail on behalf of the United States government. And then there is an app called Geofeedia, which has enabled the police, among others, to triangulate the openly shared personal information from about a dozen social media sites in order to spy on activists and shut down protests in real time.


Or there is the secretive Silicon Valley data analysis firm Palantir, funded by the Central Intelligence Agency and used by the NSA, the CIA, the FBI, numerous police forces, American Express, and hundreds of other corporations, intelligence agencies, and financial institutions. Its algorithms allow for rapid analysis of enormous amounts of data from a vast array of sources like traffic cameras, online purchases, social media posts, friendships, and e-mail exchanges—the everyday activities of innocent people—to enable police officers, for example, to assess whether someone they have pulled over for a broken headlight is possibly a criminal. Or someday may be a criminal.

It would be naive to think that there is a firewall between commercial surveillance and government surveillance. There is not.


Many of us have been concerned about digital overreach by our governments, especially after the Snowden revelations. But the consumerist impulse that feeds the promiscuous divulgence of personal information similarly threatens our rights as individuals and our collective welfare. Indeed, it may be more threatening, as we mindlessly trade ninety-eight degrees of freedom for a bunch of stuff we have been mesmerized into thinking costs us nothing.

RELATED



[Are We Puppets in a Wired World?](#) 
Sue Halpern



[How Your Data Are Being Deeply Mined](#)

Alice E. Marwick



[They've Got You, Wherever You Are](#)
Jacob Weisberg