

The Rise of the Numerati

With the explosion of data from the Internet, cell phones, and credit cards, the people who can make sense of it all are changing our world

by Stephen Baker

An excerpt from the introduction of the book The Numerati.

Imagine you're in a café, perhaps the noisy one I'm sitting in at this moment. A young woman at a table to your right is typing on her laptop. You turn your head and look at her screen. She surfs the Internet. You watch.

Hours pass. She reads an online newspaper. You notice that she reads three articles about China. She scouts movies for Friday night and watches the trailer for *Kung Fu Panda*. She clicks on an ad that promises to connect her to old high school classmates. You sit there taking notes. With each passing minute, you're learning more about her. Now imagine that you could watch 150 million people surfing at the same time. That's more or less what Dave Morgan does.

"What is it about romantic-movie lovers?" Morgan asks, as we sit in his New York office on a darkening summer afternoon. The advertising entrepreneur is flush with details about our ramblings online. He can trace the patterns of our migrations, as if we were swallows or humpback whales, while we move from site to site. Recently he's become intrigued by the people who click most often on an ad for car rentals. Among them, the largest group had paid a visit to online obituary listings. That makes sense, he says, over the patter of rain against the windows. "Someone dies, so you fly to the funeral and rent a car." But it's the second-largest group that has Morgan scratching his head. Romantic-movie lovers. For some reason Morgan can't fathom, loads of them seem drawn to a banner ad for Alamo Rent A Car.

Groundhog Day

Morgan, a cheery 43-year-old, wears his hair pushed to the side, as if when he was young his mother dipped a comb into water, drew it across, and the hair just stayed there. He grew up in Clearfield, a small town in western Pennsylvania a short drive from Punxsutawney. Every year on the second day of February, halfway between the winter solstice and the vernal equinox, a crowd in that town gathers around a large caged rodent still groggy from hibernation. They study the animal's response to its own shadow. According to ancient Celtic lore, that single bit of data tells them whether spring will come quickly or hold off until late March. Morgan has migrated as far as can be from such folk predictions. At his New York startup, Tacoda, he hires statisticians to track our wanderings on the Web and figure out our next moves. Morgan was a pioneer in Internet advertising during the dot-com boom, starting up an agency called 24/7 Real Media. During the bust that followed he founded another company, Tacoda, and moved seamlessly into what he saw as the next big thing: helping advertisers pinpoint the most promising Web surfers for their message.

Tacoda's entire business gorges on data. The company has struck deals with thousands of online publications, from *The New York Times* (NYT) to *BusinessWeek* (MHP). Their sites allow Tacoda to drop a bit of computer code called a cookie into our computers. This lets Tacoda trace our path from one site to the next.

The company focuses on our behavior and doesn't bother finding out our names or other personal details. (That might provoke a backlash concerning privacy.) But Tacoda can still learn plenty. Let's say you visit *The Boston Globe* and read a column on the Toyota (TM) Prius. Then you look at the car section on AOL (TWX). Good chance you're in the market for wheels. So

Tacoda hits you at some point in your Web wanderings with a car ad. Click on it, and Tacoda gets paid by the advertiser—and gleans one more detail about you in the process. The company harvests 20 billion of these behavioral clues every day.

Sometimes Morgan's team spots groups of Web surfers who appear to move in sync. The challenge then is to figure out what triggers their movements. Once this is clear, the advertisers can anticipate people's online journeys—and sprinkle their paths with just the right ads. This requires research. Take the curious connection between fans of romance movies and the Alamo Rent A Car ad. To come to grips with it, Morgan and his colleagues have to dig deeper into the data. Do car renters arrive in larger numbers from a certain type of romance movie, maybe ones that take place in an exotic locale? Do members of this group have other favorite sites in common? The answers lie in the strings of ones and zeros that our computers send forth. Maybe the statistics will show that the apparent link between movie fans and car renters was just a statistical quirk. Or perhaps Morgan's team will unearth a broader trend, a correlation between romance and travel, lust and wanderlust. That could lead to all kinds of advertising insights. In either case, Morgan can order up hundreds of tests. With each one he can glean a little bit more about us and target the ads with ever more precision. He's taking analysis that once ran through an advertiser's gut, and replacing it with science. We're his guinea pigs—or groundhogs—and we never stop working for him.

Fat Digital Dossiers

When it comes to producing data, we're prolific. Those of us wielding cell phones, laptops, and credit cards fatten our digital dossiers every day, simply by living. Take me. As I write on this spring morning, Verizon, my cell-phone company, can pin me down within several yards of this café in New Jersey. Visa can testify that I'm well caffeinated, probably to overcome the effects of the Portuguese wine I bought last night at 8:19. This was just in time for watching a college basketball game, which, as TiVo (TIVO) might know, I turned off after the first half. Security cameras capture time-stamped images of me near every bank and convenience store.

And don't get me started on my Web wanderings. Those are already a matter of record for dozens of Internet publishers and advertisers around the world. Dave Morgan is just one in a large and curious crowd. Late in the past century, to come up with this level of reporting, the East German government had to enlist tens of thousands of its citizens as spies. Today we spy on ourselves and send electronic updates minute by minute.

This all started with computer chips. Until the 1980s, these bits of silicon, bristling with millions of microscopic transistors, were still a novelty. But they've grown cheaper and more powerful year by year, and now manufacturers throw them into virtually anything that can benefit from a dab of smarts. They power our cell phones, the controls in our cars, our digital cameras, and, of course, our computers. Every holiday season, the packages we open bring more chips into our lives. These chips can record every instruction they receive and every job they do. They're fastidious note takers. They record the minutiae of our lives. Taken alone, each bit of information is nearly meaningless. But put the bits together, and the patterns describe our tastes and symptoms, our routines at work, the paths we tread through the mall and the supermarket. And these streams of data circle the globe. Send a friend a smiley face from your cell phone. That bit of your behavior, that tiny gesture, is instantly rushing, with billions of others, through fiber-optic cables. It's soaring up to a satellite and back down again and checking in at a server farm in Singapore before you've put the phone back in your pocket. With so many bits flying around, the very air we breathe is teeming with motes of information.

If someone could gather and organize these far-flung electronic gestures, our lives would pop into focus. This would create an ever-changing, up-to-the-minute mosaic of human behavior. The prospect is enough to make marketers quiver with excitement. Once they have a bead on our

data, they can decode our desires, our fears, and our needs. Then they can sell us precisely what we're hankering for.

Filtering Out the Noise

It sounds a lot simpler than it is. Sloshing oceans of data, from e-mails and porn downloads to sales receipts, create immense chaotic waves. In a single month, Yahoo! (YHOO) alone gathers 110 billion pieces of data about its customers, according to a 2008 study by the research firm comScore. Each person visiting sites in Yahoo's network of advertisers leaves behind, on average, a trail of 2,520 clues. Piece together these details, you might think, and our portraits as shoppers, travelers, and workers would jell in an instant. Summoning such clarity, however, is a slog. When I visit Yahoo's head of research, Prabhakar Raghavan, he tells me that most of the data trove is digital garbage. He calls it "noise" and says that it can easily overwhelm Yahoo's computers. If one of Raghavan's scientists gives an imprecise computer command while trawling through Yahoo's data, he can send the company's servers whirring madly through the noise for days on end. But a timely tweak in these instructions can speed up the hunt by a factor of 30,000. That reduces a 24-hour process to about three seconds. His point is that people with the right smarts can summon meaning from the nearly bottomless sea of data. It's not easy, but they can find us there.

The only folks who can make sense of the data we create are crack mathematicians, computer scientists, and engineers. They know how to turn the bits of our lives into symbols. Why is this necessary? Imagine that you wanted to keep track of everything you ate for a year. If you're like I was in the fourth grade, you go to the stationery store and buy a fat stack of index cards. Then, at every meal you write the different foods on fresh cards. Meat loaf. Spinach. Tapioca pudding. Cheerios. After a few days, you have a growing pile of cards. The problem is, there's no way to count or analyze them. They're just a bunch of words. These are symbols too, of course, each one representing a thing or a concept. But they are near impossible to add or subtract, or to drop into a graph illustrating a trend. Put these words in a pile, and they add up to what the specialists call "unstructured data." That's computer talk for "a big mess." A better approach would be to label all the meats with M, all the green vegetables with G, all the candies with C, and so on. Once the words are reduced to symbols, you can put them on a spreadsheet and calculate, say, how many times you ate meat or candy in a given week. Then you can make a graph linking your diet to changes in your weight or the pimple count on your face.

Counting the Pebbles

The key to this process is to find similarities and patterns. We humans do this instinctively. It's how we figured out, long ago, which plants to eat and how to talk. But while many of us were focusing on specific challenges, others were thinking more symbolically. I picture early humans sitting around a fire. Some, naturally, are jousting for the biggest piece of meat or busy with mating rituals. But off to the side, a select few are toying with stones, thinking, "If each of these pebbles represents one mammoth, then this rock..." Later, notes Tobias Dantzig in *Number: The Language of Science*, the Romans used their word *calcula*, meaning "pebble," to give a name to this thought process. But the pebble was just the start. The essence of calculation was to advance from the physical pebbles to ever-higher realms of abstract reasoning.

That science developed over the centuries, and we now have experts who are comfortable working with ridiculously large numbers, the billions and trillions that the rest of us find either unimaginable or irrelevant. They are heirs to the science that turns our everyday realities into symbols. As the data we produce continue to explode and as computers grow relentlessly stronger, these maestros gain in power. Two of them made a big splash in the late 1990s by founding Google (GOOG). For the age we're entering, Google is the marquee company. It's built almost entirely upon math, and its very purpose is to help us hunt down data. Google's breakthrough, which transformed a simple search engine into a media giant, was the discovery

that our queries—the words we type when we hunt for Web pages—are of immense value to advertisers. The company figured out how to turn our data into money. And lots of others are looking to do the same thing. Data whizzes are pouring into biology, medicine, advertising, sports, politics. They are adding us up. We are being quantified.

When this process began, a half-century ago, the first computers were primitive boxes the size of a garbage truck. They kept their distance from us, purring away in air-conditioned rooms. At this early stage, the complexity of the human animal was too much for them. They couldn't even beat us at chess. But in certain numerical domains, they showed promise. An early test involved consumer credit.

In 1956, two Stanford graduates, a mathematician named Bill Fair and his engineer friend Earl Isaac, came up with the idea of replacing loan officers with a computer. This hulking machine knew practically nothing, not even what the applicants did for a living. It certainly hadn't learned if they'd gotten a raise or filed for divorce. Legions of human loan officers, by contrast, were swimming in data. They often knew the families of the loan applicants. They were acquainted with how much the applicant had struggled in high school and how his engagement had fallen through, probably because of a drinking problem (if he was anything like his uncle). The loan officers had enough details to write sociological monographs, if they were so inclined, about the families in their towns. But they lacked a scientific system to analyze it all. Bankers depended, for the most part, on their gut.

Scores to Quantify Risk

By contrast, the computerized approach zeroed in on only a small set of numbers, most of them concerning bank balances, debts, and payment history. Bare bones. Fair and Isaac built a company to analyze the patterns of those numbers. They developed a way to determine the odds that each customer would default on a loan. Everyone got a number. These risk scores proved to be much better predictors than the gut-trusting humans. Most borrowers with high credit scores made good on their loans. And more people qualified for them. The machine, after all, didn't discriminate on the basis of anything but numbers. It was equal-opportunity banking. Like a lot of analytical systems, it was fairer. Its narrow scope, paradoxically, returned broad-minded results. What's more, a lot of people turned out to be better bets than the loan officers suspected. The market for credit expanded.

Still, the computer knew its place. It thrived in the world of numbers, and it stayed there. Those of us who specialized in words and music and images barely noticed it. Yet over the following decades, the computer grew in power, gobbling up ever more ones and zeros per millisecond. It got cheaper and smaller, and it linked up with others around the world. It produced jaw-dropping efficiencies. And from the viewpoint of the humanities crowd (including this history major), it swallowed entire technologies. It supplanted typewriters and moved on, like an imperial force, to rout record players and film cameras. It took over the mighty telephone. Finally, in the 1990s, even those of us who had long viewed computers as aliens from the basement world of geekdom started to make room for them in our homes and offices. We learned that we could use these machines to share our words and movies and photos with the entire world.

In fact, we had little choice. The old ways were laughably slow. But there was one condition: We had to render everything we sent, the very stuff of our lives, into ones and zeros. That's how we came to deliver our riches, the key to communications on earth, to the masters of the symbolic language. Now these mathematicians and computer scientists are in a position to rule the information of our lives. I call them the Numerati.

Adapted from The Numerati by Stephen Baker, copyright ©2008. Reprinted by permission of Houghton Mifflin Harcourt Publishing Co. All rights reserved.