CHAPTER 2

Shopper

THE CALL COMES from my wife at the supermarket. "Do we need onions?"

I check. "We have one big one," I say, turning it over gingerly. "But it's been sprouting for a while . . ."

"Okay, I'll get some. How about milk?"

You know the routine. A few minutes later, whichever one of us is shopping arrives at the checkout counter. There, if we remember, we dig into a pocket or purse for the frayed customer loyalty card on the key chain. The cashier scans it. We get a discount on the orange juice or razor blades, and the supermarket learns about everything we buy. It's a deal we shoppers have been making for years. Stores give us what amounts to a couple bucks a week in exchange for our shopping lists.

Here's the strange part. To date, retailers have stockpiled untold mountains of our personal data, but they're only now waking up to what they can do with it. Sure, managers have used the scans to keep an eye on inventory. They can see when to order more mangoes or Snickers bars. They've learned plenty about our behavior en masse but next to nothing about us as individuals. When we walk into a store, even if it's the hundredth time this year, the system doesn't recognize us. It's clueless.

This era is coming to an end. Retailers simply cannot afford to keep herding us blindly through stores and malls, flashing discounts on Pampers to widowers in wheelchairs and ham hocks to Jews who keep kosher. It's wasteful, and competitors are getting smarter. Look online. Whether it's Amazon.com or a travel service like Orbitz, Internet merchants are working every day to figure us out.

They're tracking every click on their sites. They know where we come from, what we buy, how much we spend, which advertisements we see. They even know which ones we linger over for a moment or two with our mouse. In the online world, businesses no longer look at us as herds but as vast collections of individuals—each of us represented by scores of equations. They prove every day that merchants who know their customers have a big edge. They can study our patterns of consumption, anticipate our appetites, and entice us to spend money.

Personal service is nothing new for retailers. For centuries, it's been a privilege for the rich. Shopkeepers and tailors know their names and measurements and their taste in premier cru burgundies. They also know where to send the bill. A few generations ago, the rest of us got personal service (on a far more modest scale) in our own neighborhoods. "The retail model was a shopkeeper, a millinery, a rug merchant," says Jeff Smith, a managing partner of the retail practice at Accenture, the tech consulting giant. "You didn't serve yourself," he says. "They stood behind counters and found what you were looking for." Chummy relations with customers gave these merchants an edge.

Following World War II, however, retail took a half-century detour into mass industrialization. Shoppers were handed carts and instructed to find their own stuff. Whether they were pushing those carts through Ikea or Wal-Mart, they had entire warehouses to explore. And the merchandise was cheap, in part because the stores had eliminated the middleman—the shopkeeper at the local store who knew the customers by name. They mastered a startling new efficiency, which came from manufacturing and distributing with martial precision. That's what the brainiacs and their computers were focused on: operations. The customers? As we made our way from the massive lots through the equally massive stores, we were processed like card-carrying herd animals.

Now retailers are changing. Accenture's Smith calls it "back to the future." Instead of deploying millions of shopkeepers to twenty-first-century counters, they're relying on automatic machinery, from video cameras to newfangled customer loyalty cards. The operation runs on data, our data. The goal is to follow our footsteps in much same the way that e-tailers track our clicks. In the marketplace of the Numerati, we'll define ourselves as shoppers in ever-greater detail simply by going about our business in a store. When the stores get to know us, they'll recognize us the moment we walk in the door—just the way the corner grocer used to. And just like that grocer, they'll know our week-to-week routines and our not-so-secret cravings. They may calculate that we're probably running short on cat kibbles, and they won't forget that we spike a gallon or two of eggnog every holiday season. (And wouldn't it taste better with premium Jamaican rum this year?) The automatic systems will calculate not only what we're likely to buy but also how much money we make for the store. Many of them will learn how to lavish big spenders with special attention and nudge cheapskates toward the door.

• • •

An old shopping cart is parked next to the wall at Accenture's lab, high above downtown Chicago. The offices are chock full of tech gadgetry. Blinking video cameras hang from the ceilings, staring down on the researchers. (They're guinea pigs in a new surveillance system designed to track shoppers and workers.) In one nook of the lab is a large, always-on video connection with another Accenture lab in Silicon Valley. Around lunchtime in Chicago, you can see the California contingent coming to work, steaming coffee cups in hand. You hear their phones ringing and their footsteps echoing across the lobby 2,000 miles to the west. All of this gadgetry is backed by a wraparound view of Chicago's skyscrapers, with Lake Michigan shimmering in the distance. In this technology showcase, the shopping cart looks out of place and a little forlorn. But it reminds Rayid Ghani and his small team of researchers of their key mission: to predict the behavior of people like my wife, and you, and me as we make our way through stores.

Ghani made a splash in 2002 with a study of how a clothing retailer like The Gap or Eddie Bauer could automatically build profiles of us from the things we buy. This sounds simple, but it adds a thick layer of complexity to data mining. If you unearth an old receipt gathering dust in your bedroom, you'll see that one afternoon a few months ago you bought, say, one pair of gray pants, two cotton shirts, and some socks. What can the retailer possibly learn about you from this data? That you're a human being with a body and, presumably, two feet? They take that much for granted. That you spend an average of $863 per year in the store? That's a tad more interesting. But if each one of the items you bought carried a bit more contextual information, what computer scientists call a layer of "semantic" detail, much more of you would pop into focus.

Let's say the pants are tagged as "urban youth." With this bit of knowledge, the system can move beyond your spending habits and start to delve into your personal tastes—much the way Amazon.com calculates the kind of reader you are from the books you buy. A clothing system with semantic smarts can send you coupons for garments that appeal to urban youth. It can track the proclivities of this "tribe" (that's a word marketers adore). And depending on the store's privacy policy, it might decide to sell that data to other companies eager to market songs or cars to the same group. Some, as we'll see later, might even use tribal data to push members toward one political candidate or another. Complications? No doubt. Maybe you're a 55-year-old woman who bought that pair of pants for your 16-year-old son. Maybe he hated them. That's not really you in the receipt, and it's not him either. Faced with such complexity and contradictions, machines need smart and patient teachers to guide them in making sense of us.

That's how Rayid Ghani views himself—as a personal tutor for the idiot savants we know as computers. Ghani is short, a bit round, and quick to smile. He's one of the friendliest tutors his students could hope for (not that they'd notice). A Pakistani who studied at the computer science powerhouse Carnegie Mellon, Ghani would seem to fit right in with the Numerati. But in their rarefied ranks, he's missing a standard ingredient: a doctorate. Having "only a master's" in his circle is viewed as a handicap. But the 29-year-old outsider has grown accustomed to clawing his way upward. The son of two college professors in Karachi, Pakistan, he applied to American colleges fully aware that he could afford only those offering a full scholarship. He landed at the University of the South, in Sewanee, Tennessee. Ghani calls it "a liberal arts college in the middle of nowhere." Hardly the ideal spot for a budding computer scientist, it is better known for its theology school. But one summer, Ghani won an internship at Carnegie Mellon, in Pittsburgh. He plunged into a world where classmates were teaching cars to drive by themselves and training computers to speak and read. He developed a passion for machine learning. Upon graduation from Sewanee, he proceeded to a master's program at CMU. Ghani was in a hurry. He started publishing papers nearly as soon as he arrived. And when he got his master's, he decided to look for a job "at places where they hire Ph.D.'s." He landed at Accenture, and now, at an age at which many of his classmates are just finishing their doctorate, he runs the analytics division from his perch in Chicago.

Ghani leads me out of his office and toward the shopping cart. For statistical modeling, he explains, grocery shopping is one of the first retail industries to conquer. This is because we buy food constantly. For many of us, the supermarket functions as a chilly, Muzak-blaring annex to our pantries. (I would bet that millions of suburban Americans spend more time in supermarkets than in their formal living room.) Our grocery shopping is so prodigious that just by studying one year of our receipts, researchers can detect all sorts of patterns—far more than they can learn from a year of records detailing our other, more sporadic purchases. (Most of us, for example, buy zero cars and zero TV sets in any given year.)

Three years ago, Ghani's team at Accenture began to work with a grocery chain. (They're not allowed to name it.) This project came with a windfall: two years of detailed customer records. The stores left out names, ages, and other demographic details, but none of that mattered. The 20,000 shoppers Ghani and his colleagues studied were simply numbers. But by their behavior in the stores, each number produced a detailed portrait of a shopper.

Let's assume you're one of those nameless shoppers. What can researchers learn about you? As it turns out, plenty. By the patterns of your purchases, and the amount you spend week after week, they can see if you're on a budget. They can calculate your spending limit. If they add some semantic tags to the data, they can draw other conclusions. When they see you starting to buy skim milk, or perhaps those miracle milk shakes, they can infer that you're on a diet. And they have no trouble seeing when you lapse. That carton of Ben & Jerry's in your cart, or the big wheel of Roquefort, is a giveaway. But wait! Maybe it's the holiday season, or your birthday. A few more weeks of receipts will spell out whether you're just cheating a little or in free fall. All of this they can do with the kind of statistical analysis an eighth grader could understand.

It gets a bit more complicated when they calculate your brand loyalty. Let's say you like Cherry Coke. You lug home a 12-pack every week. How much would Pepsi have to slash the price of its Wild Cherry Cola to entice you to switch? Ghani and two colleagues, Katharina Probst and Chad Cumby, watch how the shoppers respond to sales and promotional giveaways. They score each shopper on brand loyalty, and even loyalty to certain products within a brand. Some people, they've found, are loyal to certain foods, such as Kraft's macaroni and cheese. But does that loyalty extend to other Kraft products? For a certain group of shoppers, it does. The Accenture team takes note.

What they have on their hands is an enormous catalog of the eating habits of a small group of urban Americans in the first years of this century. Anthropologists of a certain bent would feast on it. But what good does it do a supermarket to know that you, for example, have a $95 weekly budget, are fiercely loyal to Cheetos, and flirted with the Atkins diet last barbecue season? What can they do with all that intelligence when they don't do business with you until you show up, loyalty card in hand, at the checkout counter? At that point, you've done your shopping. The chance to offer you promotions based on your profile has passed. Sure, they can throw a few coupons in your bag. Maybe you'll remember them on your next visit, but probably not. This is why, until now, supermarkets have virtually ignored the records of individual shoppers. They had little opportunity to put them to use.

The real breakthrough will come when retailers can spot you grabbing an empty cart and pushing it into the store. This has been a grocers' dream for decades. In a previous life in the 1990s, that sad little shopping cart at Accenture was a proud prototype of a "smart cart," one that allowed shoppers to swipe their loyalty cards through a computer attached to the cart, which would then lead them to bargains. "Everyone tried to do it," Ghani says. The attempts fell flat. The computers were too pricey, the analytics primitive. But computers are far cheaper now. Companies like Accenture are betting they can make systems so smart that shoppers will view the new smart cart as a personal assistant.

The first of such smart carts are just starting to roll. Stop & Shop is testing them in grocery stores in Massachusetts. Carts powered by a Microsoft program are taking their first turns in ShopRite supermarkets along the East Coast. The German chain Metro is launching them in Düsseldorf. And Samsung-Tesco, a Korean-British venture, has them operating in Seoul. A few things we know even at this early stage. For one, a computer on a shopping cart can ill afford to make dumb mistakes. This sounds axiomatic, but the fact is, we've long given grocery stores the benefit of the doubt when they offer us fliers and coupons that don't match our needs or wants, since they don't pretend to know them. But if a shopper has been buying skim milk for a year and the personalized cart insists on promoting half-and-half, the shopper may well view the smart cart as idiotic (and revert to the traditional dumb cart that specializes in rolling).

The other extreme? If these carts get too smart, we'll likely view them as creepy. I can just imagine rolling through my neighborhood Kings, when the cart starts flashing a message: STEVE: Hurry to aisle three for bargains on two of your favorite FUNGAL MEDICATIONS, plus this bonus SELECTION for the fungus you're most likely to contract NEXT! At that point, I'd be inclined to push it out to the street and under the wheels of an oncoming truck.

Setting aside such troubling scenarios, here's what shopping with one of these carts might feel like. You grab a cart on the way in and swipe your loyalty card. The welcome screen pops up with a shopping list. It's based on the patterns of your past purchases. Milk, eggs, zucchini, whatever. Smart systems might provide you with the quickest route to each item. Or perhaps they'll allow you to edit the list, to tell it, for example, never to promote cauliflower or salted peanuts again. This is simple stuff. But according to Accenture's studies, shoppers forget an average of 11 percent of the items they intend to buy. If stores can effectively remind us of what we want, it means fewer midnight runs to the convenience store for us and more sales for them.

Things get more interesting when store managers begin to manipulate our behavior. Rayid Ghani opens his laptop and shows me the supermarket control panel that he and his team have built. "Let's say you want four hundred shoppers to switch to a certain brand of frozen fish," he says. With a couple of clicks, the manager can see how many shoppers at the store buy this item. They sit in groupings known in marketing lingo as "buckets"—in this case, the frozen-fish bucket. Let's say it includes 5,000 shoppers. Among that group are those who buy rival brands of frozen fish. They're the target audience, and they sit in three smaller buckets, say, 1,000 shoppers per rival brand. Of those shoppers, one-third appear to be brand loyalists. It would likely take big discounts to pry them from the fish they usually buy. But the others, some 2,000, are more flexible when it comes to brands. They switch easily and often.

These buckets, as you can see, are getting increasingly refined. Now we're down to the brand-fickle buyers of certain types of frozen fish. Ghani plays at the controls. If he cuts the price by just 50 cents a pound—and sends word of the discount to their smart carts—he can entice a projected 150 of them to jump to the target brand. Ghani lowers the price by another 75 cents. At that level, an additional 300 bargain-hunters would line up to buy the fish. The manager can play with endless variables. He can adjust the formula to raise profits, to goose sales, to promote brands, to slash inventory. It's a virtual puppet show, all of it based on probability. The puppets, needless to say, are mathematical representations of us.

Let's say you're notoriously fickle when it comes to brands. Even the smallest fluctuations will push you from Cheerios to Wheaties and back again. If the manager is interested in slashing inventory, you're likely to be in the first bucket he picks up. You're an easy sell. But if the goal is to switch your allegiance from one brand to another, you're a lousy bet. No offense, but you're disloyal, at least in this context. You'll pocket the discount and abandon the brand the very next time you can save a dime. The manager might fare better promoting the discount to those who stick to brands a bit longer than you do. Naturally, they're in another bucket.

You may also lose out on discounts if you hew to a weekly budget. Let's say you spend about $120 a week on groceries. The system calculates that you're on a budget because, say, 87 percent of the time you spend between $113 and $125 a week. If you're not restricted to a formal limit, you might as well be. Assume that the manager is eager to get rid of a mountain of detergent moldering in the warehouse. He's offering jumbo boxes at two for the price of one. Should he send the word to your screen? Maybe not, Ghani says. The reason is simple. For every dollar you spend on discounted products, that's one less dollar you have in your budget to spend at full price. That hurts profits. To get rid of that detergent, it's smarter to target people in freer-spending buckets.

Among the most unpleasant buckets a manager must confront are those loaded with "barnacle" shoppers. That term comes from V. Kumar, a consultant and marketing professor at the University of Connecticut. Barnacles, from a retailer's perspective, are detestable creatures. We all know a few of them. They're the folks who drive from store to store, clipped coupons in hand, buying discounted goods—and practically nothing else. Kumar calls them barnacles because, like the mollusks clinging to a ship, they hitch free rides and contribute nothing of value. In fact, they cost the retailer money. With all the consumer data pouring in, Kumar says, it's becoming a snap to calculate a projected profit (or loss) for each customer. Kumar, who sells his advice to Ralph Lauren and Procter & Gamble, says that retailers should "fire" customers who look likely to drag down profits.

This doesn't mean hiring musclebound bouncers to block these shoppers at the door. But retailers can take steps in that direction. They can start by removing barnacles from their mailing lists. Increasingly, they'll also have the means to make adjustments inside the store. If bona fide barnacles are pushing smart carts through a supermarket, for example, it might make sense to fill their screens with off-putting promotions for full-priced caviar and truffles. (Discouraging unwanted shoppers is far easier on the Internet. Already, online merchants are assailing their barnacles with advertisements. And if these bargain hunters click to browse the pages of a book or gawk at the free photos on a paid-porn site, they get shunted to the slowest servers, so that they wait and wait.)

If you think about it, barnacles thrive in markets where we're all treated alike. They feast on opportunities that the rest of us, for one reason or another, miss. But now retailers are gaining tools not only to spot barnacles but also to discriminate against them. Barnacles, of course, are the first to notice when this happens. It's their nature to keep their eyes wide open. And you can bet that they'll challenge this type of discrimination in court. In a class-action suit in 2005, lawyers representing some 6 million subscribers to Netflix, the film-by-mail rental service, charged that the service was taking longer to send movies to its most active customers. Those were the film buffs who paid a flat monthly fee of $17.99 for limitless rentals and tried to see as many movies as they could for their money. This involved watching a movie or two the very day they arrived in the mail and rushing to mail them back the next morning. (I know the routine; for my first few months on Netflix, I was an eager barnacle.) Netflix officials admitted that they favored less active (and more profitable) customers with prompt mailings. And in a settlement, they gave millions of subscribers a free month of service. But, significantly, they did not vow to change their barnacle-punishing ways. They simply adjusted the wording in their rental contracts.

Barnacles aren't the only creatures in Kumar's menagerie. He also warns retailers about "butterflies," customers who drop in at the store on occasion, spend good money, and then flit away, sometimes for months or years on end. They're unreliable, and retailers are warned to avoid lavishing attention on them. "You shouldn't chase the butterflies," the professor says. However, by studying their patterns of behavior, smart retailers may learn which butterflies they can turn into reliable customers—a bucket that Kumar calls "true friends."

As merchants learn more about us, it's going to be easier for them to figure out which customers to reward and which ones to punish. This won't make much difference to butterfly shoppers. They're oblivious. But in the age of the retailing Numerati, life for barnacles might get grim.

WITH ALL THIS TALK of butterflies and buckets, I ask Ghani, where is the individual? I expected to see myself modeled as a shopper, and here I am, sitting in buckets with other frozen-fish buyers and brand traitors. What's become of customization? Where's the fully formed mathematical model of the cheapskate who never pays the extra buck for yellow or red bell peppers? I'm talking about the reluctant clothes shopper, the one rushing through the mall with a tightening back who always takes two laps around the garage before finding his car? In short, where am I in all this data?

Ghani smiles as he delivers the bad news. There's no fully formed "me" in that data. There's no you, at least not yet. We exist in these databases as shards of our behavior, my hang-up with the bell peppers, your habit of casually tossing a bag of M&Ms onto the pile as you wait at the checkout. (By the way, those seemingly impulsive purchases, often accompanied by a what-the-hell shrug, are no afterthoughts, Ghani's data shows. Many shoppers buy the candy bars and breath mints more predictably than they purchase milk or toilet paper.) In any case, all of those pieces of our shopping selves reside in endless buckets with other people's slivers. Much as we might find it flattering to sit in a unique bucket all by our lonesome, for retailers there's no point. They don't have a customized marketing campaign for me or for you. They want to sell pork or crew-neck sweaters. And for this, they'd like to bring together 1,000 or 50,000 people. Just because they like to microtarget doesn't mean that they wouldn't rather reach lots of people with the same message. They still love big numbers. They just prefer to target customers more intelligently. It would be easy to mistake these new buckets for the demographic groupings marketers have worked with for decades: Hispanics, yuppies, soccer moms, the super rich who inhabit the 90210 zip code in Los Angeles. Those are buckets too. But there's a world of difference.

In the old days, marketers knew next to nothing about the individual, so they assumed that he or she shared values and urges with similar people—those who also made six-figure salaries or had a last name with a vowel at the end. This was a crude indicator. But given the information they had, it was the best they could do. And in the decades of industrialized consumption, in the 1950s and '60s, it wasn't half bad. Choices were limited. Why bother learning about a person if, chances were, he had little choice but to watch *The Honeymooners,* eat one of three different kinds of peanut butter on his sandwich, or buy a car that looked pretty much like a Chevy? We have thousands more choices now, from the supermarket shelves to the remote on the TV, not to mention the Internet. So marketers, as Tacoda's Dave Morgan demonstrates, can shift their focus from who we are to how we behave. For this, they need the new buckets.

To see how different these new groupings are, consider the demographics of these buckets we inhabit. Start by looking at the skinflints who, like me, forgo the pleasures of red and yellow bell peppers. In this green-pepper bucket I'll wager that I'm surrounded by people of all races. Both genders are represented (though I'd imagine, based on my family sample, that more of us are guys). We drive all kinds of cars. Some of us hunt; others would just as soon outlaw guns. The district attorney might be in there, sharing bucket space with the FBI's most-wanted killer. You could say we have nothing in common, and you'd be absolutely right—except for one thing: our behavior when it comes to buying bell peppers.

These bits of our behavior sit in thousands of buckets, all of them created automatically by machines. Most of them—like my green-pepper bucket—are never used. If you strung all of your buckets one after the next, you'd see your own special combination, your unique shopping genome. Spend time with microtargeting marketers these days, and you'll hear them refer to these behavior patterns as a consumer's DNA. This comparison is not fair or accurate, though it sounds temptingly simple. Unlike our genetic code, our behavior changes all the time. We learn. (Who knows? After one tasty Moroccan meal I might be inspired to spring for a basket of exorbitant red peppers imported from Holland.)

Still, forget those technicalities for a minute. Think of buckets as genes. Each base pair of a gene (which provides instructions to produce amino acids) is described by combinations of two of four chemicals known as nucleotides. They're represented by the letters A, G, T, and C. That basic code is pretty simple. But there are key variations, both in the DNA code for individual genes and in the 3.2 billion base pairs in the genome. To a large degree, those differences shape our bodies and our lives, distinguishing us not only from other plants and animals, but also from each other.

Since the 1990s, thousands of the world's leading mathematicians and computer scientists have been drawing up algorithms to comb through vast databases of DNA and other health data. They're looking for patterns in those billions of base pairs that might point to a proclivity for leukemia, creative genius, alcoholism, or perhaps a deadly allergy to peanuts. The research is still at an early stage, but scientists have built an enormous mathematical toolbox for linking symptoms to variations in the four building blocks of DNA.

Why does that matter to a grocer? For now, it doesn't. But let's say that a supermarket, a few years down the road, organizes each aspect of our shopping data into four groups. For example, we buy candy at the checkout

1. More than 90 percent of the time
2. From 25 to 89 percent
3. From 1 to 24 percent
4. Never

With modern computing, it wouldn't be that hard to organize thousands, or even millions, of our grocery-shopping habits into similar groups of four. They'll be arbitrary, much like the census or the categories on insurance forms. The point here, however, isn't to model one entire person accurately but instead to decode the patterns of human behavior. Consider the people who buy luxury chocolates. Is there anything in their purchasing behavior that appears to trigger chocolate lust? Grocers have wrestled with these questions for centuries. They make sensible correlations. Chocolate lovers might be

interested in almonds. Catch them at the holidays and before Valentine's Day. But how about the correlations that humans wouldn't think to look for, such as the romance-movie lovers who clicked on Alamo car rental ads? How do grocers unearth those hidden links?

This is where the data-mining algorithms could come in and lead to randomized experiments with shoppers, Ghani says. Once the retailers have our behaviors grouped into four variables, they can retool one of these genomic algorithms and feed our shopping data to it. The computers whir through our purchases, looking at literally billions of combinations. The great majority are utterly senseless. Do people who buy both Brussels sprouts and sugared cereal also buy Swiss chocolates more than the mean? No sane person would bother looking for such a connection. That's why it's the perfect job for computers. Set them on a hunt, and they might find correlations we humans would never think to consider. Just as they've helped medical researchers find genetic markers pointing to certain types of breast cancer and Huntington's disease, they might tell grocers what kinds of fruit to promote to buyers of canned food or what types of magazines dog-food buyers tend to read. These suggestions may sound frivolous. But if a retailer can tweak promotions, bucket by bucket, and gain a boost of even 2 percent of sales, it's cause to rush down aisle seven and pop a magnum of Mumm's. They measure profit margins in this industry by the tenth of a percent.

As Ghani talks about shopping patterns and genomic researchers, I think about putting all the people we've been talking about—the grocers, the microtargeting advertisers, the mathematical geneticists—into one room. They wouldn't seem to have much in common. Yet they do. In nearly every industry, the data we produce is represented by ones and zeros.

It all travels through the same networks and vies for space in the same computers. This means that the mathematical tools used to analyze this data can cross disciplines and industries, from the barnyard to the aisles of Saks, almost effortlessly. This has a nearly miraculous multiplier effect—the brains working in one industry can power breakthroughs in many others. Researchers long isolated in different fields, different departments on campus, different industries are now solving the same problems. The analysis of networks, for example, extends from physics to sociology. In a sense, all of these scientists are working in one global networked laboratory.

All of which is to say that researchers whose tools will one day decipher the secrets of your shopping—perhaps the subconscious patterns you don't even know about yet—may not be working for Wal-Mart or Google or Ghani's team at Accenture. Today they might be studying earthworms or nanotechnology, or maybe the behavior of Democratic voters in swing states. For example, one researcher at Microsoft, David Heckerman, was hard at work building a program to comb e-mail traffic and identify spam. He knew that spammers systematically altered their mailings to break through ever more sophisticated defenses. He was dealing with a phenomenon similar in nature to biological mutations. His system had to anticipate these variations. Heckerman, a physician as well as a computer scientist, knew that if his tool could detect mutations in spam, it might also work in medicine. Sure enough, in 2003, he shifted his focus to HIV, the virus that causes AIDS. His tools, with their legacy in spam, could eventually lead to an AIDS vaccine. "It's the very same [software] code," he says. In the Numerati's world, breakthroughs can come from any direction.

• • •

Consider for a moment the clothes you put on this morning. If Rayid Ghani and his colleagues had a picture of you as you made your way down to breakfast or out the door, would they know from your clothing what tribe to put you in? Chances are, they could come pretty close. Humans have specialized in tribal recognition since we climbed down from trees. It's a survival skill.

But how does Ghani teach that skill to a machine? Computers, after all, have to figure out what kind of clothing we're buying if they're going to classify us as dweebs, business drones, hip-hoppers, earth mothers, or whatever other fashion buckets the marketers create. It is true, of course, that armies of people could flip through these garments, giving each one a tribal tag. But this procedure would cost a bundle, and the workers (who themselves come from different tribes) would surely disagree on what's sexy, fashion-forward, or retro. Humans are just too subjective. This is a job for computers. However, when it comes to classifying clothing, Ghani says, machines fare no better than the most clueless of humans—at least for now. So the Accenture team in the Chicago lab has to cheat.

Here's how. They hire a group of people to teach the computer. These trainers slog through a questionnaire from an online department store catalog. For several hundred garments, they answer a series of multiple-choice questions. Is it formal or casual? Is it business attire? On a scale of one to ten, how sporty is it? How trendy is it? What age group is it for? On and on. Several people evaluate each item. This smooths over their individual quirks and produces a consensus. As the humans answer these questions, the computer learns about each piece of clothing. If it were human, perhaps it would be able to develop an eye for what's sporty and what's hip, and then be able to classify the rest of the fashion universe by itself. But computers don't yet have such discerning eyes. Instead, the machine focuses on the promotional language that accompanies each picture. *Zesty! Hot! Spring fever!* It learns to associate those words with the values spelled out by its human trainers.

In the end, the computer builds up a matrix of words, all of them defined by their statistical relationships to each category of clothing. *Bra,* to cite an obvious example, would have a near-zero probability of belonging in men's wear. In every example marked by the humans, it shows up for women. But that doesn't tell the computer whether a certain bra is sporty, casual, or Gen Y. For that, it must find clues in other words.

Ghani shows me the vocabulary his system has mastered. He calls up "conservative" words. The computer spits out *trouser, classic, blazer, Ralph,* and *Lauren.* Words that rank low on the conservative scale? Ghani calls it up and laughs. "*Leopard*! That's a good one." Others are *rose, chemise, straps, flirty, spray, silk,* and *platform.* I'd say the computer has figured out a thing or two. When Ghani asks it for "high brand appeal," *DKNY* and *imported* show up, along with that now familiar duo of *Ralph* and *Lauren.* (This system, Ghani explains, has no fancy understanding of context. Unlike other artificial intelligence programs, it is unburdened by grammar. It just plows through the English words it has encountered and pegs each one to a set of probabilities.)

Figuring out that a certain white blouse is business attire for a female baby boomer is merely step one for the computer. The more important task is to build a profile of the shopper who buys that blouse. Let's say it's my wife. She goes to Macy's and buys four or five items for herself. Underwear, pants, a couple of blouses, maybe a belt. All of the items fit that boomer profile. She's coming into focus. Then, on the way out she remembers to buy a birthday present for our 16-year-

old niece. Last time we saw her, this girl was wearing black clothing with a lot of writing on it, most of it angry. She told us she was a goth. So my wife goes into an "alternative" section and—what the hell?—picks up one of those dog collars bristling with sharp spikes.

How does Ghani's system interpret this surprising deviation? Jaime Carbonell, a professor of machine learning at Carnegie Mellon, thinks about these issues a lot. In the early days, he says, consumers were often averaged. He noticed that Amazon.com, for example, saw that he was interested in Civil War history and in computational biology. So it combined them. He got recommendations for the history of biology and the north-south divide on some scientific question. "The average modeling doesn't work well," he says. "We're not the average of our interests." The newer approach is to use clustering software. This divides his interests into different groups and gives him recommendations based on each one.

Let's say my wife's purchases were clustered. The system could look at most of her purchases and conclude that she's a female boomer. The dog collar? It's what statisticians call an outlier. In these early days, it's something that's safer to ignore. But as analysis gets more sophisticated, it will latch onto those bits of our lives that appear to be deviations. After all, which details are more likely to lay us bare, our day-to-day behavior that appears "normal" or the apparent quirks that we often work to hide? A detective will opt for the outlier in a New York second. The marketer might too. But it's tough to make sense of such data with automatic systems.

In any case, suppose that next week my wife returns to the same store and buys piercing tools and green hair dye. At that point, the software might turn the spike collar she purchased, that apparent outlier, into its own cluster. So what would that new cluster tell us about her? Hard to say. Is she a middle-aged professional who commutes Monday through Friday in sober attire and then, on weekends, straps on the spiked collar and goes goth? Could be. Or perhaps she's buying for two people. Ghani says that some systems in grocery stores look at the different clusters and try to come to conclusions about the composition of a family. Others look at the different signals as varying dimensions of one person. Sometimes, though, "mutually exclusive" purchases in the same cart—small socks and big shoes—indicate that more than one person is involved.

Accenture's automatic fashion maven isn't yet grappling with such subtle distinctions. It's still in the research phase. But once this type of technology is in the marketplace, stores will have strong signals as to what types of shoppers we are. At the same time, they'll be compiling ever more detailed and valuable customer lists. As we'll see, plenty of other marketers, such as those in dating services or political groups, would pay richly for, say, a list of 10,000 trendy Gen Y women in Seattle, Chicago, or Miami. And yes, there will be lively markets, no doubt, for assorted varieties of goths.

LET'S SAY YOU go to a department store with a shopping list. If you come back missing a couple of items, the store has failed an important test. Even if you locate and buy everything on the list, your visit, from the store's perspective, falls short of an unqualified success. No, they want you to stumble upon countless temptations as you make your way up and down the aisles. In their dreams, you teeter up to the checkout under such a pile of serendipitous finds that you have to pay a young assistant or two to help you lug it to the car.

How to make that happen? The first step is to map our

migrations through the store. In the old days, some store managers and museum curators would gauge foot traffic by the wear on the floor tiles. Then they would redeploy their offerings to draw customers off the beaten paths. But that approach is a tad slow for the Numerati.

Ghani and his team have another idea. As we walk around the Accenture office, cameras hanging from the ceiling are tracking our every move. There are about 40 of them, Ghani says matter-of-factly. From my perspective, it's insidious workplace surveillance. With this kind of spy network installed in my skyscraper offices in New York, I think I'd find myself rationing my trips to the bathroom. But Ghani and his colleagues view the cameras as just one more experiment, this one to track workers and customers. The Accenture workers are offering themselves as specimens, and they don't seem to mind a bit.

This type of monitoring system isn't that relevant to Accenture's lab setting, where the flow of information counts for more than bodily movements. But Ghani sees growing numbers of cameras tracking the movements of customers and employees in big stores, hotels, and casinos. They could also find a home in factories. Such cameras are already installed as a security measure, Ghani says. So now it's just a matter of giving the camera another job.

With this type of snooping technology, managers can start scrutinizing our movements. In these early days, they focus more on overall patterns of traffic than on individuals. That's because today's cameras have foggy vision. They see us as little more than moving blurs, Ghani says. They'd be hard-pressed to identify our faces, even if we stood perfectly still, gazed up at them, and mouthed, ever so slowly, our names. Most automatic surveillance systems, which seem to recognize faces so well in the movies, don't yet work such magic in the real world. Douglas Arnold, director at the Institute for Mathematics and Its Applications at the University of Minnesota, says that facial recognition was oversold as far back as the 1960s. Researchers are making progress, Arnold says, but "if people start relying on facial recognition systems today, they're going to be burned."

So how will Accenture's cameras pick out individual workers and shoppers? Ghani introduces me to what he calls massive redundancy. This involves getting lots of cameras to work together as a team. Each one provides a bit of detail. It's a little like a group of witnesses who see a thief dash by. One might remember his red hat, another the bandage on his hand. A third points to the alley he ran down. In Accenture's case, the system can stitch together these smidgens and come up with a guess as to who each blur is most likely to be. They can be pretty sure, for example, about the identity of a short, dark-haired figure wearing a blue shirt who emerges from Ghani's office accompanied by a taller stranger with an oddly stiff neck (me, bad hotel pillows). Stature, colors, and the patterns of movement all indicate that the person is Ghani. The system makes similar calculations about the other Accenture employees on the floor. This produces truckloads of visual data. Accenture's computers use that information to feed all sorts of analysis. They can create charts showing each person's migratory patterns, social hubs, and yes, even bathroom visits. Similar analysis could be focused on us as customers. In time, perhaps a store will recognize us by our movements in the aisles as likely butterflies or barnacles, or even potential shoplifters. And as the facial recognition systems improve, they may spot the barnacles among us the moment we enter the store.

If cameras don't pick us up, a radio technology known as

RFID just might. These are little computer chips fastened to a piece of merchandise, a shopping cart, or even a customer loyalty card. Each chip has a unique number, identifying the item or the shopper. But unlike a bar code, which has to be passed under a scanner, these chips can be read by radio signals sent by an automatic reader in the area. It's great for logistics. Open a huge cargo truck, and instead of piling through it and scanning each bar code, the chips all transmit their data at once. The detailed contents of the shipment pop up in a split second.

These same chips can track us in stores and at conventions. AllianceTech, a company in Austin, Texas, puts these radio tags into the ID badges people wear around their necks or clipped to their lapels at trade shows. The company also puts receivers into the booths they visit. Then, when IBM or Texas Instruments wishes to know who visited its booth, AllianceTech can give them the names of the people (at least those who agreed to participate), their companies and industries, and the amount of time they spent in the booth. They can even see how much time these people spent visiting a rival's booth. If you look at the flow of customer data, it's as if the whole trade show is taking place on the Internet.

Imagine what would happen if retail stores used the same technology. Some are moving in that direction. Germany's Metro, the world's fifth-largest retailer, is equipping smart carts with radio transmitters in several of its stores. The technology, says Albrecht von Truchsess, a spokesman for Metro in Düsseldorf, is meant to provide shoppers with enhanced service but not to compile their shopping-related data or to profile them. (Data privacy is a far more explosive issue in Europe than in the United States.) As shoppers push the smart cart, they scan the bar code of each item they toss in. This information is sent by wireless connection to the computer, and, much like a driver cruising through an automatic tollbooth, the shopper can roll the cart out of the store without stopping to pay. The technology has taken care of that.

By mapping the trail of those scans, Metro can trace the minute-by-minute wanderings of each shopper. Even without building personal profiles, Metro's analysts will be able to study patterns. They may discover that many of the most carefree, sky's-the-limit shoppers never encounter the display in aisle three promoting sinfully rich (and expensive) Belgian chocolate. Just like a website, the store has plenty of options to entice the consumer: It can flag the chocolates on the smart cart screens. Or it can tweak the layout of the store, placing the chocolates along the pathways most popular among spendthrifts. Pity the dieters who dare to shop in the markets of the Numerati.