

1

The Race for Relevance

If you're not paying for something, you're not the customer; you're the product being sold.

—Andrew Lewis, under the alias *Blue_beetle*, on the Web site *MetaFilter*

In the spring of 1994, Nicholas Negroponte sat writing and thinking. At the MIT Media Lab, Negroponte's brainchild, young chip designers and virtual-reality artists and robot-wranglers were furiously at work building the toys and tools of the future. But Negroponte was mulling over a simpler problem, one that millions of people pondered every day: what to watch on TV.

By the mid-1990s, there were hundreds of channels streaming out live programming twenty-four hours a day, seven days a week. Most of the programming was horrendous and boring: infomercials for new kitchen gadgets, music videos for the latest one-hit-wonder band, cartoons, and celebrity news. For any given viewer, only a tiny percentage of it was likely to be interesting.

As the number of channels increased, the standard method of surfing through them was getting more and more hopeless. It's one thing to search through five channels. It's another to search through five hundred. And when the number hits five thousand—well, the method's useless.

But Negroponte wasn't worried. All was not lost: in fact, a solution was just around the corner. "The key to the future of television," he wrote, "is to stop thinking about television as television," and to start thinking about it as a device with embedded intelligence. What consumers needed was a remote control that controls itself, an intelligent automated helper that would learn what each viewer watches and capture the programs relevant to him or her. "Today's TV set lets you control brightness, volume, and channel," Negroponte typed. "Tomorrow's

will allow you to vary sex, violence, and political leaning.”

And why stop there? Negroponte imagined a future swarming with intelligent agents to help with problems like the TV one. Like a personal butler at a door, the agents would let in only your favorite shows and topics. “Imagine a future,” Negroponte wrote, “in which your interface agent can read every newswire and newspaper and catch every TV and radio broadcast on the planet, and then construct a personalized summary. This kind of newspaper is printed in an edition of one.... Call it the Daily Me.”

The more he thought about it, the more sense it made. The solution to the information overflow of the digital age was smart, personalized, embedded editors. In fact, these agents didn’t have to be limited to television; as he suggested to the editor of the new tech magazine *Wired*, “Intelligent agents are the unequivocal future of computing.”

In San Francisco, Jaron Lanier responded to this argument with dismay. Lanier was one of the creators of virtual reality; since the eighties, he’d been tinkering with how to bring computers and people together. But the talk of agents struck him as crazy. “What’s got into all of you?” he wrote in a missive to the “Wired-style community” on his Web site. “The idea of ‘intelligent agents’ is both wrong and evil.... The agent question looms as a deciding factor in whether [the Net] will be much better than TV, or much worse.”

Lanier was convinced that, because they’re not actually people, agents would force actual humans to interact with them in awkward and pixelated ways. “An agent’s model of what you are interested in will be a cartoon model, and you will see a cartoon version of the world through the agent’s eyes,” he wrote.

And there was another problem: The perfect agent would presumably screen out most or all advertising. But since online commerce was driven by advertising, it seemed unlikely that these companies would roll out agents who would do such violence to their bottom line. It was more likely, Lanier wrote, that these agents would have double loyalties—briable agents. “It’s not clear who they’re working for.”

It was a clear and plangent plea. But though it stirred up some chatter in online newsgroups, it didn’t persuade the software giants of this early Internet era. They were convinced by Negroponte’s logic: The company that figured out how to sift through the digital haystack for the nuggets of gold would win the future. They could see the attention crash coming, as the information options available to each person rose toward infinity. If you wanted to cash in, you needed to get people to tune in. And in an attention-scarce world, the best way to

do that was to provide content that really spoke to each person's idiosyncratic interests, desires, and needs. In the hallways and data centers of Silicon Valley, there was a new watchword: relevance.

Everyone was rushing to roll out an "intelligent" product. In Redmond, Microsoft released Bob—a whole operating system based on the agent concept, anchored by a strange cartoonish avatar with an uncanny resemblance to Bill Gates. In Cupertino, almost exactly a decade before the iPhone, Apple introduced the Newton, a "personal desktop assistant" whose core selling point was the agent lurking dutifully just under its beige surface.

As it turned out, the new intelligent products bombed. In chat groups and on e-mail lists, there was practically an industry of snark about Bob. Users couldn't stand it. *PC World* named it one of the twenty-five worst tech products of all time. And the Apple Newton didn't do much better: Though the company had invested over \$100 million in developing the product, it sold poorly in the first six months of its existence. When you interacted with the intelligent agents of the midnineties, the problem quickly became evident: They just weren't that smart.

Now, a decade and change later, intelligent agents are still nowhere to be seen. It looks as though Negroponte's intelligent-agent revolution failed. We don't wake up and brief an e-butler on our plans and desires for the day.

But that doesn't mean they don't exist. They're just hidden. Personal intelligent agents lie under the surface of every Web site we go to. Every day, they're getting smarter and more powerful, accumulating more information about who we are and what we're interested in. As Lanier predicted, the agents don't work only for us: They also work for software giants like Google, dispatching ads as well as content. Though they may lack Bob's cartoon face, they steer an increasing proportion of our online activity.

In 1995 the race to provide personal relevance was just beginning. More than perhaps any other factor, it's this quest that has shaped the Internet we know today.

The John Irving Problem

Jeff Bezos, the CEO of Amazon.com, was one of the first people to realize that you could harness the power of relevance to make a few billion dollars. Starting in 1994, his vision was to transport online bookselling "back to the days of the

small bookseller who got to know you very well and would say things like, ‘I know you like John Irving, and guess what, here’s this new author, I think he’s a lot like John Irving,’” he told a biographer. But how to do that on a mass scale? To Bezos, Amazon needed to be “a sort of a small Artificial Intelligence company,” powered by algorithms capable of instantly matching customers and books.

In 1994, as a young computer scientist working for Wall Street firms, Bezos had been hired by a venture capitalist to come up with business ideas for the burgeoning Web space. He worked methodically, making a list of twenty products the team could theoretically sell online—music, clothing, electronics—and then digging into the dynamics of each industry. Books started at the bottom of his list, but when he drew up his final results, he was surprised to find them at the top.

Books were ideal for a few reasons. For starters, the book industry was decentralized; the biggest publisher, Random House, controlled only 10 percent of the market. If one publisher wouldn’t sell to him, there would be plenty of others who would. And people wouldn’t need as much time to get comfortable with buying books online as they might with other products—a majority of book sales already happened outside of traditional bookstores, and unlike clothes, you didn’t need to try them on. But the main reason books seemed attractive was simply the fact that there were so many of them—3 million active titles in 1994, versus three hundred thousand active CDs. A physical bookstore would never be able to inventory all those books, but an online bookstore could.

When he reported this finding to his boss, the investor wasn’t interested. Books seemed like a kind of backward industry in an information age. But Bezos couldn’t get the idea out of his head. Without a physical limit on the number of books he could stock, he could provide hundreds of thousands more titles than industry giants like Borders or Barnes & Noble, and at the same time, he could create a more intimate and personal experience than the big chains.

Amazon’s goal, he decided, would be to enhance the process of discovery: a personalized store that would help readers find books and introduce books to readers. But how?

Bezos started thinking about machine learning. It was a tough problem, but a group of engineers and scientists had been attacking it at research institutions like MIT and the University of California at Berkeley since the 1950s. They called their field “cybernetics”—a word taken from Plato, who coined it to mean a self-regulating system, like a democracy. For the early cyberneticists, there was

nothing more thrilling than building systems that tuned themselves, based on feedback. Over the following decades, they laid the mathematical and theoretical foundations that would guide much of Amazon's growth.

In 1990, a team of researchers at the Xerox Palo Alto Research Center (PARC) applied cybernetic thinking to a new problem. PARC was known for coming up with ideas that were broadly adopted and commercialized by others—the graphical user interface and the mouse, to mention two. And like many cutting-edge technologists at the time, the PARC researchers were early power users of e-mail—they sent and received hundreds of them. E-mail was great, but the downside was quickly obvious. When it costs nothing to send a message to as many people as you like, you can quickly get buried in a flood of useless information.

To keep up with the flow, the PARC team started tinkering with a process they called collaborative filtering, which ran in a program called Tapestry. Tapestry tracked how people reacted to the mass e-mails they received—which items they opened, which ones they responded to, and which they deleted—and then used this information to help order the inbox. E-mails that people had engaged with a lot would move to the top of the list; e-mails that were frequently deleted or unopened would go to the bottom. In essence, collaborative filtering was a time saver: Instead of having to sift through the pile of e-mail yourself, you could rely on others to help presift the items you'd received.

And of course, you didn't have to use it just for e-mail. Tapestry, its creators wrote, "is designed to handle any incoming stream of electronic documents. Electronic mail is only one example of such a stream: others are newswire stories and Net-News articles."

Tapestry had introduced collaborative filtering to the world, but in 1990, the world wasn't very interested. With only a few million users, the Internet was still a small ecosystem, and there just wasn't much information to sort or much bandwidth to download with. So for years collaborative filtering remained the domain of software researchers and bored college students. If you e-mailed ringo@media.mit.edu in 1994 with some albums you liked, the service would send an e-mail back with other music recommendations and the reviews. "Once an hour," according to the Web site, "the server processes all incoming messages and sends replies as necessary." It was an early precursor to Pandora; it was a personalized music service for a prebroadband era.

But when Amazon launched in 1995, everything changed. From the start, Amazon was a bookstore with personalization built in. By watching which books

people bought and using the collaborative filtering methods pioneered at PARC, Amazon could make recommendations on the fly. (“Oh, you’re getting *The Complete Dummy’s Guide to Fencing*? How about adding a copy of *Waking Up Blind: Lawsuits over Eye Injury*?”) And by tracking which users bought what over time, Amazon could start to see which users’ preferences were similar. (“Other people who have similar tastes to yours bought this week’s new release, *En Garde!*”) The more people bought books from Amazon, the better the personalization got.

In 1997, Amazon had sold books to its first million customers. Six months later, it had served 2 million. And in 2001, it reported its first quarterly net profit—one of the first businesses to prove that there was serious money to be made online.

If Amazon wasn’t quite able to create the feeling of a local bookstore, its personalization code nonetheless worked quite well. Amazon executives are tight-lipped about just how much revenue it’s brought in, but they often point to the personalization engine as a key part of the company’s success.

At Amazon, the push for more user data is never-ending: When you read books on your Kindle, the data about which phrases you highlight, which pages you turn, and whether you read straight through or skip around are all fed back into Amazon’s servers and can be used to indicate what books you might like next. When you log in after a day reading Kindle e-books at the beach, Amazon is able to subtly customize its site to appeal to what you’ve read: If you’ve spent a lot of time with the latest James Patterson, but only glanced at that new diet guide, you might see more commercial thrillers and fewer health books.

Amazon users have gotten so used to personalization that the site now uses a reverse trick to make some additional cash. Publishers pay for placement in physical bookstores, but they can’t buy the opinions of the clerks. But as Lanier predicted, buying off algorithms is easy: Pay enough to Amazon, and your book can be promoted as if by an “objective” recommendation by Amazon’s software. For most customers, it’s impossible to tell which is which.

Amazon proved that relevance could lead to industry dominance. But it would take two Stanford graduate students to apply the principles of machine learning to the whole world of online information.

Click Signals

As Jeff Bezos's new company was getting off the ground, Larry Page and Sergey Brin, the founders of Google, were busy doing their doctoral research at Stanford. They were aware of Amazon's success—in 1997, the dot-com bubble was in full swing, and Amazon, on paper at least, was worth billions. Page and Brin were math whizzes; Page, especially, was obsessed with AI. But they were interested in a different problem. Instead of using algorithms to figure out how to sell products more effectively, what if you could use them to sort through sites on the Web?

Page had come up with a novel approach, and with a geeky predilection for puns, he called it PageRank. Most Web search companies at the time sorted pages using keywords and were very poor at figuring out which page for a given word was the most relevant. In a 1997 paper, Brin and Page dryly pointed out that three of the four major search engines couldn't find themselves. "We want our notion of 'relevant' to only include the very best documents," they wrote, "since there may be tens of thousands of slightly relevant documents."

Page had realized that packed into the linked structure of the Web was a lot more data than most search engines made use of. The fact that a Web page linked to another page could be considered a "vote" for that page. At Stanford, Page had seen professors count how many times their papers had been cited as a rough index of how important they were. Like academic papers, he realized, the pages that a lot of other pages cite—say, the front page of Yahoo—could be assumed to be more "important," and the pages that those pages voted for would matter more. The process, Page argued, "utilized the uniquely democratic structure of the web."

In those early days, Google lived at google.stanford.edu, and Brin and Page were convinced it should be nonprofit and advertising free. "We expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers," they wrote. "The better the search engine is, the fewer advertisements will be needed for the consumer to find what they want.... We believe the issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm."

But when they released the beta site into the wild, the traffic chart went vertical. Google worked—out of the box, it was the best search site on the Internet. Soon, the temptation to spin it off as a business was too great for the twenty-something cofounders to bear.

In the Google mythology, it is PageRank that drove the company to

worldwide dominance. I suspect the company likes it that way—it's a simple, clear story that hangs the search giant's success on a single ingenious breakthrough by one of its founders. But from the beginning, PageRank was just a small part of the Google project. What Brin and Page had really figured out was this: The key to relevance, the solution to sorting through the mass of data on the Web was ... more data.

It wasn't just which pages linked to which that Brin and Page were interested in. The position of a link on the page, the size of the link, the age of the page—all of these factors mattered. Over the years, Google has come to call these clues embedded in the data *signals*.

From the beginning, Page and Brin realized that some of the most important signals would come from the search engine's users. If someone searches for "Larry Page," say, and clicks on the second link, that's another kind of vote: It suggests that the second link is more relevant to that searcher than the first one. They called this a *click signal*. "Some of the most interesting research," Page and Brin wrote, "will involve leveraging the vast amount of usage data that is available from modern web systems.... It is very difficult to get this data, mainly because it is considered commercially valuable." Soon they'd be sitting on one of the world's largest stores of it.

Where data was concerned, Google was voracious. Brin and Page were determined to keep everything: every Web page the search engine had ever landed on, every click every user ever made. Soon its servers contained a nearly real-time copy of most of the Web. By sifting through this data, they were certain they'd find more clues, more signals, that could be used to tweak results. The search-quality division at the company acquired a black-ops kind of feel: few visitors and absolute secrecy were the rule.

"The ultimate search engine," Page was fond of saying, "would understand exactly what you mean and give back exactly what you want." Google didn't want to return thousands of pages of links—it wanted to return one, the one you wanted. But the perfect answer for one person isn't perfect for another. When I search for "panthers," what I probably mean are the large wild cats, whereas a football fan searching for the phrase probably means the Carolina team. To provide perfect relevance, you'd need to know what each of us was interested in. You'd need to know that I'm pretty clueless about football; you'd need to know who I was.

The challenge was getting enough data to figure out what's personally relevant to each user. Understanding what someone means is tricky business—

and to do it well, you have to get to know a person's behavior over a sustained period of time.

But how? In 2004, Google came up with an innovative strategy. It started providing other services, services that required users to log in. Gmail, its hugely popular e-mail service, was one of the first to roll out. The press focused on the ads that ran along Gmail's sidebar, but it's unlikely that those ads were the sole motive for launching the service. By getting people to log in, Google got its hands on an enormous pile of data—the hundreds of millions of e-mails Gmail users send and receive each day. And it could cross-reference each user's e-mail and behavior on the site with the links he or she clicked in the Google search engine. Google Apps—a suite of online word-processing and spreadsheet-creation tools—served double duty: It undercut Microsoft, Google's sworn enemy, and it provided yet another hook for people to stay logged in and continue sending click signals. All this data allowed Google to accelerate the process of building a theory of identity for each user—what topics each user was interested in, what links each person clicked.

By November 2008, Google had several patents for personalization algorithms—code that could figure out the groups to which an individual belongs and tailor his or her result to suit that group's preference. The categories Google had in mind were pretty narrow: to illustrate its example in the patent, Google used the example of “all persons interested in collecting ancient shark teeth” and “all persons not interested in collecting ancient shark teeth.” People in the former category who searched for, say, “Great White incisors” would get different results from the latter.

Today, Google monitors every signal about us it can get its hands on. The power of this data can't be underestimated: If Google sees that I log on first from New York, then from San Francisco, then from New York again, it knows I'm a bicoastal traveler and can adjust its results accordingly. By looking at what browser I use, it can make some guesses about my age and even perhaps my politics.

How much time you take between the moment you enter your query and the moment you click on a result sheds light on your personality. And of course, the terms you search for reveal a tremendous amount about your interests.

Even if you're not logged in, Google is personalizing your search. The neighborhood—even the block—that you're logging in from is available to Google, and it says a lot about who you are and what you're interested in. A query for “Sox” coming from Wall Street is probably shorthand for the financial

legislation “Sarbanes Oxley,” while across the Upper Bay in Staten Island it’s probably about baseball.

“People always make the assumption that we’re done with search,” said founder Page in 2009. “That’s very far from the case. We’re probably only 5 percent of the way there. We want to create the ultimate search engine that can understand anything.... Some people could call that artificial intelligence.”

In 2006, at an event called Google Press Day, CEO Eric Schmidt laid out Google’s five-year plan. One day, he said, Google would be able to answer questions such as “Which college should I go to?” “It will be some years before we can at least partially answer those questions. But the eventual outcome is ... that Google can answer a more hypothetical question.”

Facebook Everywhere

Google's algorithms were unparalleled, but the challenge was to coax users into revealing their tastes and interests. In February 2004, working out of his Harvard dorm room, Mark Zuckerberg came up with an easier approach. Rather than sifting through click signals to figure out what people cared about, the plan behind his creation, Facebook, was to just flat out ask them.

Since he was a college freshman, Zuckerberg had been interested in what he called the "social graph"—the set of each person's relationships. Feed a computer that data, and it could start to do some pretty interesting and useful things—telling you what your friends were up to, where they were, and what they were interested in. It also had implications for news: In its earliest incarnation as a Harvard-only site, Facebook automatically annotated people's personal pages with links to the *Crimson* articles in which they appeared.

Facebook was hardly the first social network: As Zuckerberg was hacking together his creation in the wee hours of the morning, a hairy, music-driven site named MySpace was soaring; before MySpace, Friendster had for a brief moment captured the attention of the technorati. But the Web site Zuckerberg had in mind was different. It wouldn't be a coy dating site, like Friendster. And unlike MySpace, which encouraged people to connect whether they knew each other or not, Facebook was about taking advantage of existing real-world social connections. Compared to its predecessors, Facebook was stripped down: the emphasis was on information, not flashy graphics or a cultural vibe. "We're a utility," Zuckerberg said later. Facebook was less like a nightclub than a phone company, a neutral platform for communication and collaboration.

Even in its first incarnation, the site grew like wildfire. After Facebook expanded to a few select Ivy League campuses, Zuckerberg's inbox was flooded with requests from students on other campuses, begging him to turn on Facebook for them. By May of 2005, the site was up and running at over eight hundred colleges. But it was the development of the News Feed the following September that pushed Facebook into another league.

On Friendster and MySpace, to find out what your friends were up to, you had to visit their pages. The News Feed algorithm pulled all of these updates out of Facebook's massive database and placed them in one place, up front, right when you logged in. Overnight, Facebook had turned itself from a network of

connected Web pages into a personalized newspaper featuring (and created by) your friends. It's hard to imagine a purer source of relevance.

And it was a gusher. In 2006, Facebook users posted literally billions of updates—philosophical quotes, tidbits about who they were dating, what was for breakfast. Zuckerberg and his team egged them on: The more data users handed over to the company, the better their experience could be and the more they'd keep coming back. Early on, they'd added the ability to upload photos, and now Facebook had the largest photo collection in the world. They encouraged users to post links from other Web sites, and millions were submitted. By 2007, Zuckerberg bragged, “We're actually producing more news in a single day for our 19 million users than any other media outlet has in its entire existence.”

At first, the News Feed showed nearly everything your friends did on the site. But as the volume of posts and friends increased, the Feed became unreadable and unmanageable. Even if you had only a hundred friends, it was too much to read.

Facebook's solution was EdgeRank, the algorithm that powers the default page on the site, the Top News Feed. EdgeRank ranks every interaction on the site. The math is complicated, but the basic idea is pretty simple, and it rests on three factors. The first is affinity: The friendlier you are with someone—as determined by the amount of time you spend interacting and checking out his or her profile—the more likely it is that Facebook will show you that person's updates. The second is the relative weight of that type of content: Relationship status updates, for example, are weighted very highly; everybody likes to know who's dating whom. (Many outsiders suspect that the weight, too, is personalized: Different people care about different kinds of content.) The third is time: Recently posted items are weighted over older ones.

EdgeRank demonstrates the paradox at the core of the race for relevancy. To provide relevance, personalization algorithms need data. But the more data there is, the more sophisticated the filters must become to organize it. It's a never-ending cycle.

By 2009, Facebook had hit the 300 million user mark and was growing by 10 million people per month. Zuckerberg, at twenty-five, was a paper billionaire. But the company had bigger ambitions. What the News Feed had done for social information, Zuckerberg wanted to do for all information. Though he never said it, the goal was clear: Leveraging the social graph and the masses of information Facebook's users had provided, Zuckerberg wanted to put Facebook's news algorithm engine at the center of the web.

Even so, it was a surprise when, on April 21, 2010, readers loaded the *Washington Post* homepage and discovered that their friends were on it. In a prominent box in the upper right corner—the place where any editor will tell you the eye lands first—was a feature titled Network News. Each person who visited saw a different set of links in the box—the *Washington Post* links their friends had shared on Facebook. The *Post* was letting Facebook edit its most valuable online asset: its front page. The *New York Times* soon followed suit.

The new feature was one piece of a much bigger rollout, which Facebook called “Facebook Everywhere” and announced at its annual conference, f8 (“fate”). Ever since Steve Jobs sold the Apple by calling it “insanely great,” a measure of grandiosity has been part of the Silicon Valley tradition. But when Zuckerberg walked onto the stage on April 21, 2010, his words seemed plausible. “This is the most transformative thing we’ve ever done for the web,” he announced.

The aim of Facebook Everywhere was simple: make the whole Web “social” and bring Facebook-style personalization to millions of sites that currently lack it. Want to know what music your Facebook friends are listening to? Pandora would now tell you. Want to know what restaurants your friends like? Yelp now had the answer. News sites from the *Huffington Post* to the *Washington Post* were now personalized.

Facebook made it possible to press the Like button on any item on the Web. In the first twenty-four hours of the new service, there were 1 billion Likes—and all of that data flowed back into Facebook’s servers. Bret Taylor, Facebook’s platform lead, announced that users were sharing 25 billion items a month. Google, once the undisputed leader in the push for relevance, seemed worried about the rival a few miles down the road.

The two giants are now in hand-to-hand combat: Facebook poaches key executives from Google; Google’s hard at work constructing social software like Facebook. But it’s not totally obvious why the two new-media monoliths should be at war. Google, after all, is built around answering questions; Facebook’s core mission is to help people connect with their friends.

But both businesses’ bottom lines depend on the same thing: targeted, highly relevant advertising. The contextual advertisements Google places next to search results and on Web pages are its only significant source of profits. And while Facebook’s finances are private, insiders have made clear that advertising is at the core of the company’s revenue model. Google and Facebook have different starting points and different strategies—one starts with the relationships among

pieces of information, while the other starts with the relationships among people—but ultimately, they're competing for the same advertising dollars.

From the point of view of the online advertiser, the question is simple. Which company can deliver the most return on a dollar spent? And this is where relevance comes back into the equation. The masses of data Facebook and Google accumulate have two uses. For users, the data provides a key to providing personally relevant news and results. For advertisers, the data is the key to finding likely buyers. The company that has the most data and can put it to the best use gets the advertising dollars.

Which brings us to lock-in. Lock-in is the point at which users are so invested in their technology that even if competitors might offer better services, it's not worth making the switch. If you're a Facebook member, think about what it'd take to get you to switch to another social networking site—even if the site had vastly greater features. It'd probably take a lot—re-creating your whole profile, uploading all of those pictures, and laboriously entering your friends' names would be extremely tedious. You're pretty locked in. Likewise, Gmail, Gchat, Google Voice, Google Docs, and a host of other products are part of an orchestrated campaign for Google lock-in. The fight between Google and Facebook hinges on which can achieve lock-in for the most users.

The dynamics of lock-in are described by Metcalfe's law, a principle coined by Bob Metcalfe, the inventor of the Ethernet protocol that wires together computers. The law says that the usefulness of a network increases at an accelerating rate as you add each new person to it. It's not much use to be the only person you know with a fax machine, but if everyone you work with uses one, it's a huge disadvantage not to be in the loop. Lock-in is the dark side of Metcalfe's law: Facebook is useful in large part because everyone's on it. It'd take a lot of mismanagement to overcome that basic fact.

The more locked in users are, the easier it is to convince them to log in—and when you're constantly logged in, these companies can keep tracking data on you even when you're not visiting their Web sites. If you're logged into Gmail and you visit a Web site that uses Google's Doubleclick ad service, that fact can be attached to your Google account. And with tracking cookies these services place on your computer, Facebook or Google can provide ads based on your personal information on third-party sites. The whole Web can become a platform for Google or Facebook.

But Google and Facebook are hardly the only options. The daily turf warfare between Google and Facebook occupies scores of business reporters and

gigabytes of blog chatter, but there's a stealthy third front opening up in this war. And though most of the companies involved operate under the radar, they may ultimately represent the future of personalization.

The Data Market

The manhunt for accomplices of the 9/11 killers was one of the most extensive in history. In the immediate aftermath of the attacks, the scope of the plot was unclear. Were there more hijackers who hadn't yet been found? How extensive was the network that had pulled off the attacks? For three days, the CIA, FBI, and a host of other acronymed agencies worked around the clock to identify who else was involved. The country's planes were grounded, its airports closed.

When help arrived, it came from an unlikely place. On September 14, the bureau had released the names of the hijackers, and it was now asking—pleading—for anyone with information about the perpetrators to come forward. Later that day, the FBI received a call from Mack McLarty, a former White House official who sat on the board of a little-known but hugely profitable company called Acxiom.

As soon as the hijackers' names had been publicly released, Acxiom had searched its massive data banks, which take up five acres in tiny Conway, Arkansas. And it had found some very interesting data on the perpetrators of the attacks. In fact, it turned out, Acxiom knew more about eleven of the nineteen hijackers than the entire U.S. government did—including their past and current addresses and the names of their housemates.

We may never know what was in the files Acxiom gave the government (though one of the executives told a reporter that Acxiom's information had led to deportations and indictments). But here's what Acxiom knows about 96 percent of American households and half a billion people worldwide: the names of their family members, their current and past addresses, how often they pay their credit card bills whether they own a dog or a cat (and what breed it is), whether they are righthanded or left-handed, what kinds of medication they use (based on pharmacy records) ... the list of data points is about 1,500 items long.

Acxiom keeps a low profile—it may not be an accident that its name is nearly unpronounceable. But it serves most of the largest companies in America—nine of the ten major credit card companies and consumer brands from Microsoft to Blockbuster. “Think of [Acxiom] as an automated factory,” one engineer told a

reporter, “where the product we make is data.”

To get a sense of Acxiom’s vision for the future, consider a travel search site like Travelocity or Kayak. Ever wondered how they make money? Kayak makes money in two ways. One is pretty simple, a holdover from the era of travel agents: When you buy a flight using a link from Kayak, airlines pay the site a small fee for the referral.

The other is much less obvious. When you search for the flight, Kayak places a cookie on your computer—a small file that’s basically like putting a sticky note on your forehead saying “Tell me about cheap bicoastal fares.” Kayak can then sell that piece of data to a company like Acxiom or its rival BlueKai, which auctions it off to the company with the highest bid—in this case, probably a major airline like United. Once it knows what kind of trip you’re interested in, United can show you ads for relevant flights—not just on Kayak’s site, but on literally almost any Web site you visit across the Internet. This whole process—from the collection of your data to the sale to United—takes under a second.

The champions of this practice call it “behavioral retargeting.” Retailers noticed that 98 percent of visitors to online shopping sites leave without buying anything. Retargeting means businesses no longer have to take “no” for an answer.

Say you check out a pair of running sneakers online but leave the site without springing for them. If the shoe site you were looking at uses retargeting, their ads—maybe displaying a picture of the exact sneaker you were just considering—will follow you around the Internet, showing up next to the scores from last night’s game or posts on your favorite blog. And if you finally break down and buy the sneakers? Well, the shoe site can sell that piece of information to BlueKai to auction it off to, say, an athletic apparel site. Pretty soon you’ll be seeing ads all over the Internet for sweat-wicking socks.

This kind of persistent, personalized advertising isn’t just confined to your computer. Sites like Loopt and Foursquare, which broadcast a user’s location from her mobile phone, provide advertisers with opportunities to reach consumers with targeted ads even when they’re out and about. Loopt is working on an ad system whereby stores can offer special discounts and promotions to repeat customers on their phones—right as they walk through the door. And if you sit down on a Southwest Airlines flight, the ads on your seat-back TV screen may be different from your neighbors’. Southwest, after all, knows your name and who you are. And by cross-indexing that personal information with a database like Acxiom’s, it can know a whole lot more about you. Why not show

you your own ads—or, for that matter, a targeted show that makes you more likely to watch them?

TargusInfo, another of the new firms that processes this sort of information, brags that it “delivers more than 62 billion real-time attributes a year.” That’s 62 billion points of data about who customers are, what they’re doing, and what they want. Another ominously named enterprise, the Rubicon Project, claims that its database includes more than half a billion Internet users.

For now, retargeting is being used by advertisers, but there’s no reason to expect that publishers and content providers won’t get in on it. After all, if the *Los Angeles Times* knows that you’re a fan of Perez Hilton, it can front-page its interview with him in your edition, which means you’ll be more likely to stay on the site and click around.

What all of this means is that your behavior is now a commodity, a tiny piece of a market that provides a platform for the personalization of the whole Internet. We’re used to thinking of the Web as a series of one-to-one relationships: You manage your relationship with Yahoo separately from your relationship with your favorite blog. But behind the scenes, the Web is becoming increasingly integrated. Businesses are realizing that it’s profitable to share data. Thanks to Acxiom and the data market, sites can put the most relevant products up front and whisper to each other behind your back.

The push for relevance gave rise to today’s Internet giants, and it is motivating businesses to accumulate ever more data about us and to invisibly tailor our online experiences on that basis. It’s changing the fabric of the Web. But as we’ll see, the consequences of personalization for how we consume news, make political decisions, and even how we think will be even more dramatic.