<< Return to article

Surveillance Nation

Webcams, tracking devices, and interlinked databases are leading to the elimination of unmonitored public space. Are we prepared for the consequences of the intelligence-gathering network we're unintentionally building?

Illustration by Brian Stauffer.

By Dan Farmer and Charles C. Mann
April 2003

Route 9 is an old two-lane highway that cuts across Massachusetts from Boston in the east to Pittsfield in the west. Near the small city of Northampton, the highway crosses the wide Connecticut River. The Calvin Coolidge Memorial Bridge, named after the president who once served as Northampton's mayor, is a major regional traffic link. When the state began a long-delayed and still-ongoing reconstruction of the bridge in the summer of 2001, traffic jams stretched for kilometers into the bucolic New England countryside.

In a project aimed at alleviating drivers' frustration, the University of Massachusetts Transportation Center, located in nearby Amherst, installed eight shoe-size digital surveillance cameras along the roads leading to the bridge. Six are mounted on utility poles and the roofs of local businesses. Made by Axis Communications in Sweden, they are connected to dial-up modems and transmit images of the roadway before them to a Web page, which commuters can check for congestion before tackling the road. According to Dan Dulaski, the system's technical manager, running the entire webcam system—power, phone, and Internet fees—costs just $600 a month.

The other two cameras in the Coolidge Bridge project are a little less routine. Built by Computer Recognition Systems in Wokingham, England, with high-quality lenses and fast shutter speeds (1/10,000 second), they are designed to photograph every car and truck that passes by. Located eight kilometers apart, at the ends of the

zone of maximum traffic congestion, the two cameras send vehicle images to attached computers, which use special character-recognition software to decipher vehicle license plates. The license data go to a server at the company's U.S. office in Cambridge, MA, about 130 kilometers away. As each license plate passes the second camera, the server ascertains the time difference between the two readings. The average of the travel durations of all successfully matched vehicles defines the likely travel time for crossing the bridge at any given moment, and that information is posted on the traffic watch Web page.

To local residents, the traffic data are helpful, even vital: police use the information to plan emergency routes. But as the computers calculate traffic flow, they are also making a record of all cars that cross the bridge—when they do so, their average speed, and (depending on lighting and weather conditions) how many people are in each car.

Trying to avoid provoking privacy fears, Keith Fallon, a Computer Recognition Systems project engineer, says, "we're not saving any of the information we capture. Everything is deleted immediately." But the company could change its mind and start saving the data at any time. No one on the road would know.

The Coolidge Bridge is just one of thousands of locations around the planet where citizens are crossing—willingly, more often than not—into a world of networked, highly computerized surveillance. According to a January report by J.P. Freeman, a security market-research firm in Newtown, CT, 26 million surveillance cameras have already been installed worldwide, and more than 11 million of them are in the United States. In heavily monitored London, England, Hull University criminologist Clive Norris has estimated, the average person is filmed by more than 300 cameras each *day*.

The $150 million-a-year remote digital-surveillance-camera market will grow, according to Freeman, at an annual clip of 40 to 50 percent for the next 10 years. But astonishingly, other, nonvideo forms of monitoring will increase even faster. In a process that mirrors the unplanned growth of the Internet itself, thousands of personal, commercial, medical, police, and government databases and monitoring systems will intersect and entwine. Ultimately, surveillance will become so ubiquitous, networked, and searchable that unmonitored public space will effectively cease to exist.

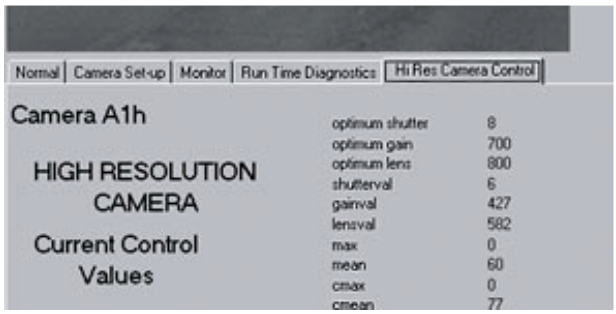
This prospect—what science fiction writer David Brin calls "the transparent society"—may sound too distant to be

Road Tools

Web-accessible video cameras installed near Northampton, MA, by the University of Massachusetts Transportation Center overlook the Calvin Coolidge Memorial Bridge on Route 9.

worth thinking about. But even the farsighted Brin underestimated how quickly technological advances—more powerful microprocessors, faster network transmissions, larger hard drives, cheaper electronics, and more sophisticated and powerful software—would make universal surveillance possible.

It's not all about Big Brother or Big Business, either. Widespread electronic scrutiny is usually denounced as a creation of political tyranny or corporate greed. But the rise of omnipresent surveillance will be driven as much by ordinary citizens' understandable—even laudatory—desires for security, control, and comfort as by the imperatives of business and government. "Nanny cams," global-positioning locators, police and home security networks, traffic jam monitors, medical-device radio-frequency tags, small-business webcams: the list of monitoring devices employed by and for average Americans is already long, and it will only become longer. Extensive surveillance, in short, is coming into being because people like and want it.



Two additional cameras photograph individual cars crossing the bridge...

...and send the images to computers that isolate license plates and use machine vision algorithms to read the plate numbers. Once a plate has passed both cameras, the car's travel time is computed.

"Almost all of the pieces for a surveillance society are already here," says Gene Spafford, director of Purdue University's Center for Education and Research in Information Assurance and Security. "It's just a matter of assembling them." Unfortunately, he says, ubiquitous surveillance faces intractable social and technological problems that could well reduce its usefulness or even make it dangerous. As a result, each type of monitoring may be beneficial in itself, at least for the people who put it in place, but the collective result could be calamitous.

To begin with, surveillance data from multiple sources are being combined into large databases. For example, businesses track employees' car, computer, and telephone use to evaluate their job performance; similarly, the U.S. Defense Department's experimental Total Information Awareness project has announced plans to sift through information about millions of people to find data that identify criminals and terrorists.

But many of these merged pools of data are less reliable than small-scale, localized monitoring efforts; big databases are harder to comb for bad entries, and their conclusions are far more difficult to verify. In addition, the inescapable nature of surveillance can itself create alarm, even among its beneficiaries. "Your little camera network may seem like a good idea to you," Spafford says. "Living with everyone else's could be a nightmare."

The Surveillance Ad-Hocracy

Cheap pix. The hardware behind the University of Massachusetts, Amherst's traffic cam network costs just $600 a month to run, says technical manager Dan Dulaski. (Photograph by John Soares)

Last October deadly snipers terrorized Washington, DC, and the surrounding suburbs, killing 10 people. For three long weeks, law enforcement agents seemed helpless to stop the murderers, who struck at random and then vanished into the area's snarl of highways. Ultimately, two alleged killers were arrested, but only because their taunting messages to the authorities had inadvertently provided clues to their identification.

In the not-too-distant future, according to advocates of policing technologies, such unstoppable rampages may become next to impossible, at least in populous areas. By combining police cameras with private camera networks like that on Route 9, video coverage will become so complete that any snipers who waged an attack—and all the people near the crime scene—would be trackable from camera to camera until they could be stopped and interrogated.

The unquestionable usefulness and sheer affordability of these extensive video-surveillance systems suggest that they will propagate rapidly. But despite the relentlessly increasing capabilities of such systems, video monitoring is still but a tiny part—less than 1 percent—of surveillance overall, says Carl Botan, a Purdue center researcher who has studied this technology for 15 years.

Examples are legion. By 2006, for instance, law will require that every U.S. cell phone be designed to report its precise location during a 911 call; wireless carriers plan to use the same technology to offer 24-hour location-based services, including tracking of people and vehicles. To prevent children from wittingly or unwittingly calling up porn sites, the Seattle company N2H2 provides Web filtering and monitoring services for 2,500 schools serving 16 million students. More than a third of all large corporations electronically review the computer files used by their employees, according to a recent American Management Association survey. Seven of the 10 biggest supermarket chains use discount cards to monitor customers' shopping habits: tailoring product offerings to customers' wishes is key to survival in that brutally competitive business. And as part of a new, federally mandated tracking system, the three major U.S. automobile manufacturers plan to put special radio transponders known as radio frequency identification tags in every tire sold in the nation. Far exceeding congressional requirements, according to a leader of the Automotive Industry Action Group, an industry think tank, the tags can be read on vehicles going as fast as 160 kilometers per hour from a distance of 4.5 meters.

Many if not most of today's surveillance networks were set up by government and big business, but in years to come individuals and small organizations will set the pace of growth. Future sales of Net-enabled surveillance cameras, in the view of Fredrik Nilsson, Axis Communications' director of business development, will be driven by organizations that buy more than eight but fewer than 30 cameras—condo associations, church groups, convenience store owners, parent-teacher associations, and anyone else who might like to check what is happening in one place while he is sitting in another. A dozen companies already help working parents monitor their children's nannies and day-care centers from the office; scores more let them watch backyards, school buses, playgrounds, and their own living rooms. Two new startups—Wherify Wireless in Redwood Shores, CA, and Peace of Mind at Light Speed in Westport, CT—are introducing bracelets and other

portable devices that continuously beam locating signals to satellites so that worried moms and dads can always find their children.

As thousands of ordinary people buy monitoring devices and services, the unplanned result will be an immense, overlapping grid of surveillance systems, created unintentionally by the same ad-hocracy that caused the Internet to explode. Meanwhile, the computer networks on which monitoring data are stored and manipulated continue to grow faster, cheaper, smarter, and able to store information in greater volume for longer times. Ubiquitous digital surveillance will marry widespread computational power—with startling results.

The factors driving the growth of computing potential are well known. Moore's law—which roughly equates to the doubling of processor speed every 18 months—seems likely to continue its famous march. Hard drive capacity is rising even faster. It has doubled every year for more than a decade, and this should go on "as far as the eye can see," according to Robert M. Wise, director of product marketing for the desktop product group at Maxtor, a hard drive manufacturer. Similarly, according to a 2001 study by a pair of AT&T Labs researchers, network transmission capacity has more than doubled annually for the last dozen years, a tendency that should continue for at least another decade and will keep those powerful processors and hard drives well fed with fresh data.

Today a company or agency with a $10 million hardware budget can buy processing power equivalent to 2,000 workstations, two petabytes of hard drive space (two million gigabytes, or 50,000 standard 40-gigabyte hard drives like those found on today's PCs), and a two-gigabit Internet connection (more than 2,000 times the capacity of a typical home broadband connection). If current trends continue, simple arithmetic predicts that in 20 years the same purchasing power will buy the processing capability of 10 million of today's workstations, 200 exabytes (200 million gigabytes) of storage capacity, and 200 exabits (200 million megabits) of bandwidth. Another way of saying this is that by 2023 large organizations will be able to devote the equivalent of a contemporary PC to monitoring every single one of the 330 million people who will then be living in the United States.

One of the first applications for this combination of surveillance and computational power, says Raghu Ramakrishnan, a database researcher at the University of Wisconsin-Madison, will be continuous intensive monitoring of buildings, offices, and stores: the spaces where middle-class people spend most of their lives. Surveillance in the workplace is common now: in 2001, according to the American Management Association survey, 77.7 percent of major U.S. corporations electronically monitored their employees, and that statistic had more than doubled since 1997 (*see "Eye on Employees," below*). But much more is on the way. Companies like Johnson Controls and Siemens, Ramakrishnan says, are already "doing simplistic kinds of 'asset tracking,' as they call it." They use radio frequency identification tags to monitor the locations of people as well as inventory. In January, Gillette began attaching such tags to 500 million of its Mach 3 Turbo razors. Special "smart shelves" at Wal-Mart stores will record the removal of razors by shoppers, thereby alerting stock clerks whenever shelves need to be refilled—and effectively transforming Gillette customers into walking radio beacons. In the future, such tags will be used by hospitals to ensure that patients and staff maintain quarantines, by law offices to keep visitors from straying into rooms containing clients' confidential papers, and in kindergartens to track toddlers.

By employing multiple, overlapping types of monitoring, Ramakrishnan says, managers will be able to "keep track of people, objects, and environmental levels throughout a whole complex." Initially, these networks will be installed for "such mundane things as trying to figure out when to replace the carpets or which areas of lawn get the most traffic so you need to spread some grass seed preventively." But

as computers and monitoring equipment become cheaper and more powerful, managers will use surveillance data to construct complex, multidimensional records of how spaces are used. The models will be analyzed to improve efficiency and security—and they will be sold to other businesses or governments. Over time, the thousands of individual monitoring schemes inevitably will merge together and feed their data into large commercial and state-owned networks. When surveillance databases can describe or depict what every individual is doing at a particular time, Ramakrishnan says, they will be providing humankind with the digital equivalent of an ancient dream: being "present, in effect, almost anywhere and anytime."

Eye on Employees

Percentage of major U.S. employers that record and review their workers' activities

SURVEILLANCE ACTIVITY	1997	1998	1999	2000	2001
Recording telephone conversations	10.04	11.2	10.6	11.5	11.9
Monitoring telephone usage	34.4	40.2	38.6	44.0	43.3
Storing and reviewing voice mail	5.3	5.3	5.8	6.8	7.8
Storing and reviewing computer files	13.7	19.6	21.4	30.8	36.1
Storing and reviewing e-mail	14.9	20.2	27.0	38.1	46.5
Monitoring Internet connections	NA[1]	NA[1]	NA[1]	54.1	62.8
Clocking overall computer use	16.1	15.9	15.2	19.4	18.9
Video recording of employee performance	15.7	15.6	16.1	14.6	15.2
Video surveillance for security	33.7	32.7	32.8	35.3	37.7
Any active electronic monitoring	35.3	42.7	45.1	73.5[2]	77.7[2]

1. not available. 2. Includes Internet monitoring, which was not measured prior to 2000.

Garbage In, Gragbea Otu

In 1974 Francis Ford Coppola wrote and directed *The Conversation*, which starred Gene Hackman as Harry Caul, a socially maladroit surveillance expert. In this remarkably prescient movie, a mysterious organization hires Caul to record a quiet discussion that will take place in the middle of a crowd in San Francisco's Union Square. Caul deploys three microphones: one in a bag carried by a confederate and two directional mikes installed on buildings overlooking the area. Afterward Caul discovers that each of the three recordings is plagued by background noise and distortions, but by combining the different sources, he is able to piece together the conversation. Or, rather, he thinks he has pieced it together. Later, to his horror, Caul learns that he misinterpreted a crucial line, a discovery that leads directly to the movie's chilling denouement.

The Conversation illustrates a central dilemma for tomorrow's surveillance society. Although much of the explosive growth in monitoring is being driven by consumer demand, that growth has not yet been accompanied by solutions to the classic difficulties computer systems have integrating disparate sources of information and arriving at valid conclusions. Data quality problems that cause little inconvenience on a local scale—when Wal-Mart's smart shelves misread a razor's radio frequency identification tag—have much larger consequences when organizations assemble big databases from many sources and attempt to draw conclusions about, say, someone's capacity for criminal action. Such problems, in the long run, will play a large role in determining both the technical and social impact of surveillance.

The experimental and controversial Total Information Awareness program of the Defense Advanced Research Projects Agency exemplifies these issues. By merging records from corporate, medical, retail, educational, travel, telephone, and even veterinary sources, as well as such "biometric" data as fingerprints, iris and retina scans, DNA tests, and facial-characteristic measurements, the program is intended to create an unprecedented repository of information about both U.S. citizens and foreigners with U.S. contacts. Program director John M. Poindexter has explained

that analysts will use custom data-mining techniques to sift through the mass of information, attempting to "detect, classify, and identify foreign terrorists" in order to "preempt and defeat terrorist acts"—a virtual Eye of Sauron, in critics' view, constructed from telephone bills and shopping preference cards.

In February Congress required the Pentagon to obtain its specific approval before implementing Total Information Awareness in the United States (though certain actions are allowed on foreign soil). But President George W. Bush had already announced that he was creating an apparently similar effort, the Terrorist Threat Integration Center, to be led by the Central Intelligence Agency. Regardless of the fate of these two programs, other equally sweeping attempts to pool monitoring data are proceeding apace. Among these initiatives is Regulatory DataCorp, a for-profit consortium of 19 top financial institutions worldwide. The consortium, which was formed last July, combines members' customer data in an effort to combat "money laundering, fraud, terrorist financing, organized crime, and corruption." By constantly poring through more than 20,000 sources of public information about potential wrongdoings—from newspaper articles and Interpol warrants to disciplinary actions by the U.S. Securities and Exchange Commission—the consortium's Global Regulatory Information Database will, according to its owner, help clients "know their customers."

Equally important in the long run are the databases that will be created by the nearly spontaneous aggregation of scores or hundreds of smaller databases. "What seem to be small-scale, discrete systems end up being combined into large databases," says Marc Rotenberg, executive director of the Electronic Privacy Information Center, a nonprofit research organization in Washington, DC. He points to the recent, voluntary efforts of merchants in Washington's affluent Georgetown district. They are integrating their in-store closed-circuit television networks and making the combined results available to city police. In Rotenberg's view, the collection and consolidation of individual surveillance networks into big government and industry programs "is a strange mix of public and private, and it's not something that the legal system has encountered much before."

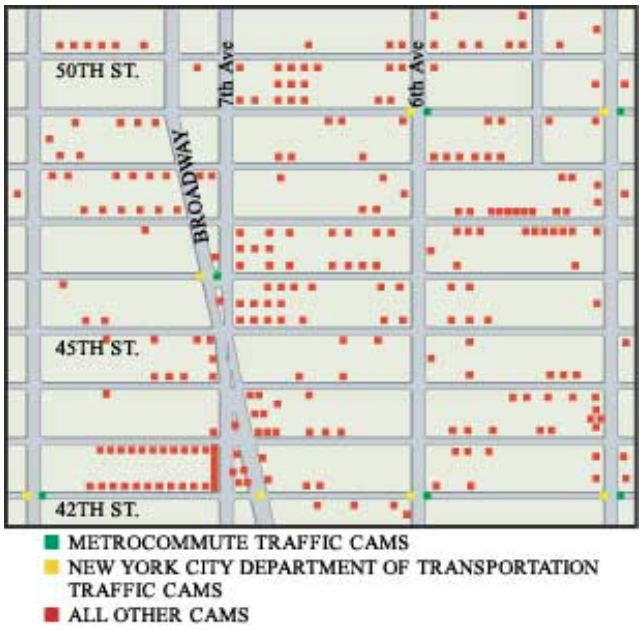
Managing the sheer size of these aggregate surveillance databases, surprisingly, will not pose insurmountable technical difficulties. Most personal data are either very compact or easily compressible. Financial, medical, and shopping records can be represented as strings of text that are easily stored and transmitted; as a general rule, the records do not grow substantially over time.

Even biometric records are no strain on computing systems. To identify people, genetic-testing firms typically need stretches of DNA that can be represented in just one kilobyte—the size of a short e-mail message. Fingerprints, iris scans, and other types of biometric data consume little more. Other forms of data can be preprocessed in much the way that the cameras on Route 9 transform multimegabyte images of cars into short strings of text with license plate numbers and times. (For investigators, having a video of suspects driving down a road usually is not as important as simply knowing that they were there at a given time.) To create a digital dossier for every individual in the United States—as programs like Total Information Awareness would require—only "a couple terabytes of well-defined information" would be needed, says Jeffrey Ullman, a former Stanford University database researcher. "I don't think that's really stressing the capacity of [even today's] databases."

Instead, argues Rajeev Motwani, another member of Stanford's database group, the real challenge for large surveillance databases will be the seemingly simple task of

gathering valid data. Computer scientists use the term GIGO—garbage in, garbage out—to describe situations in which erroneous input creates erroneous output. Whether people are building bombs or buying bagels, governments and corporations try to predict their behavior by integrating data from sources as disparate as electronic toll-collection sensors, library records, restaurant credit-card receipts, and grocery store customer cards—to say nothing of the Internet, surely the world's largest repository of personal information. Unfortunately, all these sources are full of errors, as are financial and medical records. Names are misspelled and digits transposed; address and e-mail records become outdated when people move and switch Internet service providers; and formatting differences among databases cause information loss and distortion when they are merged. "It is routine to find in large customer databases defective records—records with at least one major error or omission—at rates of at least 20 to 35 percent," says Larry English of Information Impact, a database consulting company in Brentwood, TN.

Unfortunately, says Motwani, "data cleaning is a major open problem in the research community. We are still struggling to get a formal technical definition of the problem." Even when the original data are correct, he argues, merging them can introduce errors where none had existed before. Worse, none of these worries about the garbage going into the system even begin to address the still larger problems with the garbage going out.



People passing through Manhattan's Times Square area leave a trail of images on scores of webcams and private and city-owned surveillance cameras. New York privacy activist Bill Brown compiled this map in September 2002.

The Dissolution of Privacy

Almost every computer-science student takes a course in algorithms. Algorithms are sets of specified, repeatable rules or procedures for accomplishing tasks such as sorting numbers; they are, so to speak, the engines that make programs run. Unfortunately, innovations in algorithms are not subject to Moore's law, and progress in the field is notoriously sporadic. "There are certain areas in algorithms we basically can't do better and others where creative work will have to be done," Ullman says. Sifting through large surveillance databases for information, he says, will essentially be "a problem in research in algorithms. We need to exploit some of the stuff that's been done in the data-mining community recently and do it much, much better."

Working with databases requires users to have two mental models. One is a model of the data. Teasing out answers to questions from the popular search engine Google, for example, is easier if users grasp the varieties and types of data on the Internet—Web pages with words and pictures, whole documents in a multiplicity of formats, downloadable software and media files—and how they are stored. In exactly the same way, extracting information from surveillance databases will depend on a user's knowledge of the system. "It's a chess game," Ullman says. "An unusually smart analyst will get things that a not-so-smart one will not."

Second, and more important according to Spafford, effective use of big surveillance databases will depend on having a model of what one is looking for. This factor is especially crucial, he says, when trying to predict the future, a goal of many commercial and government projects. For this reason, what might be called *reactive* searches that scan recorded data for specific patterns are generally much more likely to obtain useful answers than *proactive* searches that seek to get ahead of things. If, for instance, police in the Washington sniper investigation had been able to tap into a pervasive network of surveillance cameras, they could have tracked people seen near the crime scenes until they could be stopped and questioned: a reactive process. But it is unlikely that police would have been helped by proactively asking surveillance databases for the names of people in the Washington area with the requisite characteristics (family difficulties, perhaps, or military training and a recent penchant for drinking) to become snipers.

In many cases, invalid answers are harmless. If Victoria's Secret mistakenly mails 1 percent of its spring catalogs to people with no interest in lingerie, the price paid by all parties is small. But if a national terrorist-tracking system has the same 1 percent error rate, it will produce millions of false alarms, wasting huge amounts of investigators' time and, worse, labeling many innocent U.S. citizens as suspects. "A 99 percent hit rate is great for advertising," Spafford says, "but terrible for spotting terrorism."

Because no system can have a success rate of 100 percent, analysts can try to decrease the likelihood that surveillance databases will identify blameless people as possible terrorists. By making the criteria for flagging suspects more stringent, officials can raise the bar, and fewer ordinary citizens will be wrongly fingered. Inevitably, however, that will mean also that the "borderline" terrorists—those who don't match all the search criteria but still have lethal intentions—might be overlooked as well. For both types of error, the potential consequences are alarming.

Yet none of these concerns will stop the growth of surveillance, says Ben Shneiderman, a computer scientist at the University of Maryland. Its potential benefits are simply too large. An example is what Shneiderman, in his recent book *Leonardo's Laptop: Human Needs and the New Computing Technologies*, calls the World Wide Med: a global, unified database that makes every patient's complete medical history instantly available to doctors through the Internet, replacing today's scattered sheaves of paper records (see "Paperless Medicine,"). "The idea," he says, "is that if you're brought to an ER anywhere in the world, your medical records pop up in 30 seconds." Similar programs are already coming into existence. Backed by the Centers for Disease Control and Prevention, a team based at Harvard Medical School is planning to monitor the records of 20 million walk-in hospital patients throughout the United States for clusters of symptoms associated with bioterror agents. Given the huge number of lost or confused medical records, the benefits of such plans are clear. But because doctors would be

continually adding information to medical histories, the system would be monitoring patients' most intimate personal data. The network, therefore, threatens to violate patient confidentiality on a global scale.

In Shneiderman's view, such tradeoffs are inherent to surveillance. The collective by-product of thousands of unexceptionable, even praiseworthy efforts to gather data could be something nobody wants: the demise of privacy. "These networks are growing much faster than people realize," he says. "We need to pay attention to what we're doing right now."

In *The Conversation*, surveillance expert Harry Caul is forced to confront the tradeoffs of his profession directly. The conversation in Union Square provides information that he uses to try to stop a murder. Unfortunately, his faulty interpretation of its meaning prevents him from averting tragedy. Worse still, we see in scene after scene that even the expert snoop is unable to avoid being monitored and recorded. At the movie's intense, almost wordless climax, Caul rips his home apart in a futile effort to find the electronic bugs that are hounding him.

The Conversation foreshadowed a view now taken by many experts: surveillance cannot be stopped. There is no possibility of "opting out." The question instead is how to use technology, policy, and shared societal values to guide the spread of surveillance—by the government, by corporations, and perhaps most of all by our own unwitting and enthusiastic participation—while limiting its downside.

Next month: how surveillance technology is changing our definition of privacy—and why the keys to preserving it may be in the technology itself.

Dan Farmer is a software engineer and computer security expert.
Charles C. Mann has written for *Technology Review* about the free software movement (January/February 1999) and the use of genetic engineering in agriculture (July/August 1999).