{ Exposed }

The Erosion of Privacy in the Internet Era

Imagine if you waved to someone and, without your knowledge, a high-resolution camera took a photograph of your hand, capturing your fingerprints. You might be upset. Or—if you were visiting Disneyland, where they already make an image of your fingerprint to save you from waiting in a long line—you might find the novelty of the technology, and the immediate benefits...gratifying. The ambivalence we sometimes feel about new technologies that reveal identifiable personal information balances threats to privacy against incremental advantages. Indisputably, the trends toward miniaturization and mass-market deployment of cameras, recording devices, low-power sensors, and medical monitors of all kinds—when combined with the ability to digitally collect, store, retrieve, classify, and sort very large amounts of information—offer many benefits, but also threaten civil liberties and expectations of personal privacy. George Orwell's vision in *1984* of a future in which the government has the power to record everything seems not so farfetched. "But even Orwell did not imagine that the sensors would be things that everybody would have," says McKay professor of computer science Harry Lewis. "He foresaw the government putting the cameras on the lampposts—which we have. He didn't foresee the 14-year-old girl snapping pictures on the T. Or the fact that flash drives that are given away as party favors could carry crucial data on everybody in the country."

It's a Smaller World

Information technology changes the accessibility and presentation of information. Lewis gives talks on the subject of privacy to alumni groups in private homes, and often begins with an example that puts his hosts on the hot seat. He projects a Google Earth view of the house, then shows the website Zillow's assessment of how much it is worth, how many bedrooms and bathrooms and square feet it has. Then he goes to fundrace.huffingtonpost.com, an interface to the Federal Elections Commission's campaign-contributions database. "Such information has always been a matter of public record," says Lewis, "but it used to be that you had to go somewhere and give the exact name and address and they would give you back the one piece of data. Now you can just mouse over your neighborhood and little windows pop up and show how much money all the neighbors have given." In the 02138 zip code, you can see "all the Harvard faculty members who gave more than $1,000 to Barack Obama," for example. "This seems very invasive," says Lewis, "but in fact it is the opposite of an invasion of privacy: it is something that our elected representatives decided should be public."

Technology has forced people to rethink the public/private distinction. "Now it turns out that there is private, public, and *really, really* public," Lewis says. "We've effectively said that anyone in an Internet café in Nairobi should be able to see how much our house is worth." Lewis has been blogging about such issues on the website www.bitsbook.com, a companion to *Blown to Bits: Your Life, Liberty, and Happiness after the Digital Explosion,* the 2008 book of which he is a coauthor. "We think because we have a word for privacy that it is something we can put our arms around," he says. "But it's not."

One of the best attempts to define the full range of privacy concerns at their intersection with new technologies, "A Taxonomy of Privacy," appeared in the *University of Pennsylvania Law Review* in 2006. Its author, Daniel Solove, now a professor at George Washington University Law School, identified 16 privacy harms modulated by new technologies, including: information collection by surveillance; aggregation of information; insecurity of information; and disclosure, exposure, distortion, and increased accessibility of information.

That privacy would be a concern of the legal profession is not surprising. What *is* surprising is that computer scientists have been in the vanguard of those seeking ways to protect privacy, partly because they are often the first to recognize privacy problems engendered by new technologies and partly because the solutions themselves are sometimes technological. At Harvard, the Center for Research on Computation and Society (CRCS) has become a focal point for such inquiry. CRCS, which brings computer scientists together with colleagues from other schools and

by Jonathan Shaw

Latanya Sweeney

academic disciplines, was founded to develop new ideas and technologies for addressing some of society's most vexing problems, and prides itself on a forward-looking, integrative approach. Privacy and security have been a particular focus during the past few years.

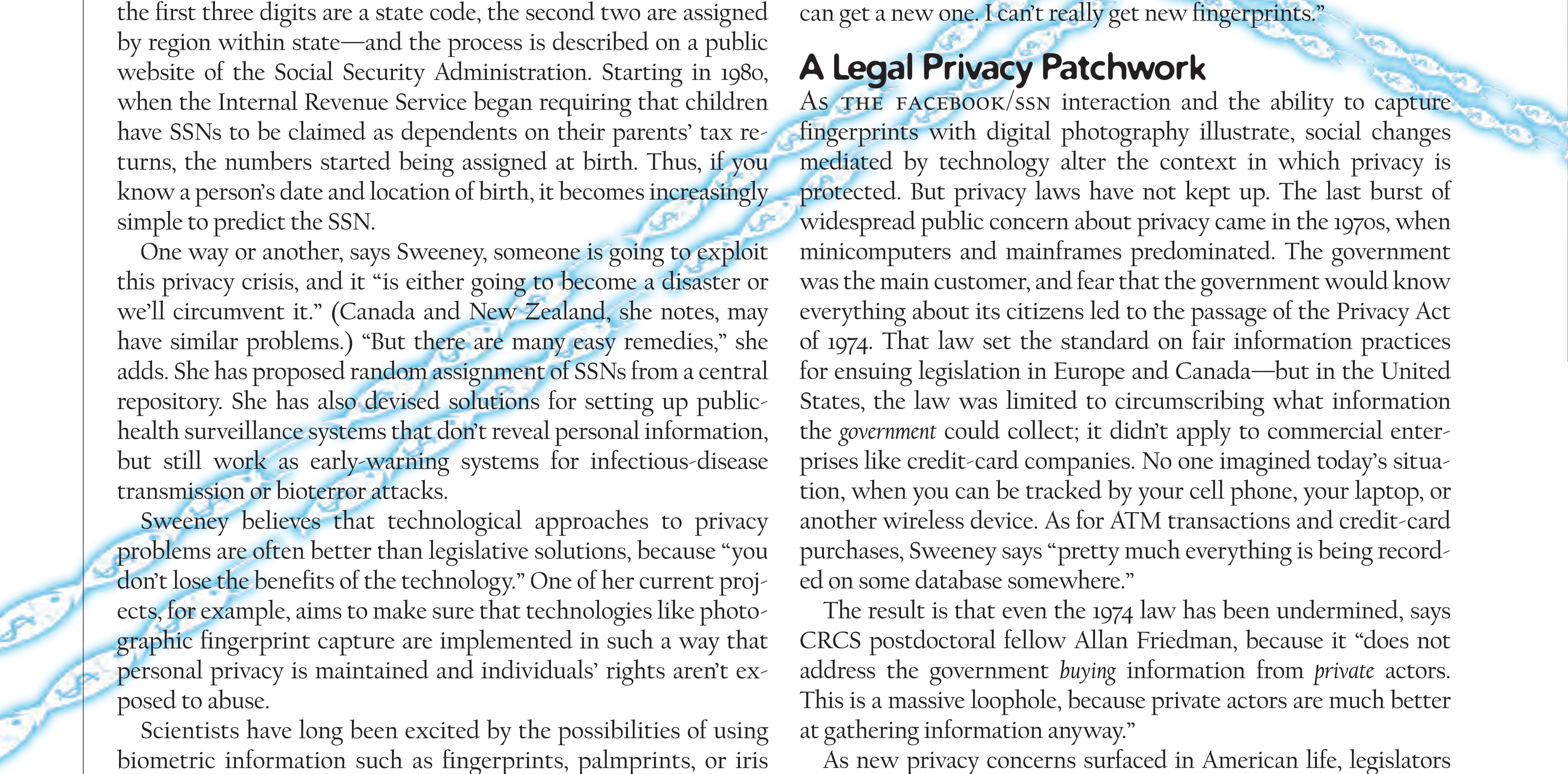
Database linking offers one such area of concern. If you tell Latanya Sweeney, A.L.B. '95, nothing about yourself except your birth date and five-digit zip code, she'll tell you your name. If you are under the age of 30 and tell her where you were born, she can correctly predict eight or nine digits of your nine-digit Social Security number. "The main reason privacy is a growing problem is that disk storage is so cheap," says the visiting professor of computer science, technology, and policy at CRCS. "People can collect data and never throw anything away. Policies on data sharing are not very good, and the result is that data tend to flow around and get linked to other data."

Sweeney became interested in privacy issues while earning her doctorate at MIT in the mid 1990s. Massachusetts had recently made "anonymized" medical information available. Such data are invaluable for research, for setting up early infectious-disease detection systems, and other public-health uses. "There was a belief at the time that if you removed explicit identifiers—name, address, and Social Security number—you could just give the data away," she recalls. That dogma was shattered when Sweeney produced a dramatic proof to the contrary.

The medical data that had been made available included minimal demographic information: zip code, birth date, and gender, in addition to the diagnosis. So Sweeney went to the Cambridge City Hall and for $25 purchased a voter list on two diskettes: 54,000 names. By linking the demographic information in the voter database to the demographic information in the publicly available medical records, Sweeney found that in most cases she could narrow the demographic data down to a single person, and so restore the patient's name to the record. She tried this data-linking technique for then-governor William F. Weld '66, J.D. '70. Only six people in Cambridge shared his birthday. Just three of them were men. And he was the only one who lived in the right zip code. Sweeney had reidentified someone in a putatively anonymous database of private medical information. The system had worked, yet data had leaked. Newspaper coverage of her testimony to the state legislature about what she had discovered ultimately brought a visit from the State Police. "That was my introduction to policy," she says with a laugh. (She was recently named to the privacy and security seat of the Health Information Technology policy committee in the Obama administration.)

Later, she proved that her results were not unique to Cambridge. Fully 87 percent of the United States population is uniquely identified by date of birth, five-digit zip code, and gender, she says: "So if I know only those three things about you, I can identify you by name 87 percent of the time. Pretty cool." In fact, Sweeney's ability to identify *anyone* is close to 100 percent for most U.S. zip codes—but there are some interesting exceptions. On the west side of Chicago, in the most populated zip code in the United States, there are more than 100,000 residents. Surely that should provide some anonymity. "For younger people, that's true," she says, "but if you are older, you really stand out." Another zip code skews the opposite way: it is on the Stony Brook campus of the State University of New York and includes only dormitories.

Photographs by Jim Harrison



"Here is a *tiny* population," she says, pulling up a graphic on her computer. "Only 5,000 people." But because they are all college students of about the same age, "they are so homogenous…that I still can't figure out who is who."

A potentially even more serious privacy crisis looms in the way Social Security numbers (SSNs) are assigned, Sweeney says. "We are entering a situation where a huge number of people could tell me just their date of birth and hometown, and I can predict their SSN. Why is this a problem? Because in order to apply for a credit card, the key things I need are your name, your date of birth, your address, and your SSN. Who is the population at risk? Young people on Facebook."

Facebook asks for your date of birth and hometown, two pieces of information that most young people include on their pages simply because they want their friends to wish them a happy birthday. The problem is that SSNs have never been issued randomly—the first three digits are a state code, the second two are assigned by region within state—and the process is described on a public website of the Social Security Administration. Starting in 1980, when the Internal Revenue Service began requiring that children have SSNs to be claimed as dependents on their parents' tax returns, the numbers started being assigned at birth. Thus, if you know a person's date and location of birth, it becomes increasingly simple to predict the SSN.

One way or another, says Sweeney, someone is going to exploit this privacy crisis, and it "is either going to become a disaster or we'll circumvent it." (Canada and New Zealand, she notes, may have similar problems.) "But there are many easy remedies," she adds. She has proposed random assignment of SSNs from a central repository. She has also devised solutions for setting up public-health surveillance systems that don't reveal personal information, but still work as early-warning systems for infectious-disease transmission or bioterror attacks.

Sweeney believes that technological approaches to privacy problems are often better than legislative solutions, because "you don't lose the benefits of the technology." One of her current projects, for example, aims to make sure that technologies like photographic fingerprint capture are implemented in such a way that personal privacy is maintained and individuals' rights aren't exposed to abuse.

Scientists have long been excited by the possibilities of using biometric information such as fingerprints, palmprints, or iris scans for positive identification: people could use them to open their cars or their homes. But just how private are fingerprints? With a grant from the National Institutes of Justice, Sweeney and her students have shown that inexpensive digital cameras are already good enough to capture fingertip friction-ridge information at a range of two to three feet, and image resolution and capture speed are improving all the time, even as the cost of the technology keeps dropping. As a result, because it is contactless and very cheap, photographic fingerprint capture could become "the dominant way that prints are captured in a lot of public spaces," Sweeney explains. That means fingerprint databases are everywhere, and "you don't have any control over the use of those prints, if somebody wanted to make a false print, or track you. It is like walking around with your Social Security number on your forehead, to an extent. It *is* a little different because it isn't linked to your credit report or your credit card"—but it does not require a tremendous leap of imagination to picture a world where credit cards require fingerprint verification.

Sweeney began working with fingerprints because of concerns that, given the huge numbers of fingerprints in linked databases, there would be false positive matches to the FBI's crime database. "To the extent that fingerprint matching has been successful, it might be because only criminals are fingerprinted and criminals tend to repeat crimes," she says. But she was "ridiculed a lot by law enforcement for making those statements," until the Madrid train bombings in 2004. When a print at the scene was falsely matched by the FBI to a lawyer in California, it became clear that the science of fingerprint matching needed to be studied more deeply. (Palmprints ultimately may have a better chance at providing a unique match.) Furthermore, Sweeney points out, "What if someone advocated replacing Social Security numbers with fingerprints? If something goes horribly wrong with my number, I can get a new one. I can't really get new fingerprints."

A Legal Privacy Patchwork

As the Facebook/SSN interaction and the ability to capture fingerprints with digital photography illustrate, social changes mediated by technology alter the context in which privacy is protected. But privacy laws have not kept up. The last burst of widespread public concern about privacy came in the 1970s, when minicomputers and mainframes predominated. The government was the main customer, and fear that the government would know everything about its citizens led to the passage of the Privacy Act of 1974. That law set the standard on fair information practices for ensuing legislation in Europe and Canada—but in the United States, the law was limited to circumscribing what information the *government* could collect; it didn't apply to commercial enterprises like credit-card companies. No one imagined today's situation, when you can be tracked by your cell phone, your laptop, or another wireless device. As for ATM transactions and credit-card purchases, Sweeney says "pretty much everything is being recorded on some database somewhere."

The result is that even the 1974 law has been undermined, says CRCS postdoctoral fellow Allan Friedman, because it "does not address the government *buying* information from *private* actors. This is a massive loophole, because private actors are much better at gathering information anyway."

As new privacy concerns surfaced in American life, legislators responded with a finger-in-the-dike mentality, a "patchwork" response, Friedman continues. "The great example of this is that for almost 10 years, your video-rental records had stronger privacy protection than either your financial or your medical records." The video-rental records law—passed in 1988 after a newspaper revealed Supreme Court nominee Robert Bork's rentals—was so narrowly crafted that most people think it doesn't even apply to Netflix. "Bork didn't have much to hide," Friedman says, "but clearly enough people in Congress did." Medical records were protected under the Health Insurance Portability and Accountability Act in 1996, but financial records weren't protected until the Gramm-Leach-Bliley Act of 1999. (Student records are protected by the Family Educational Rights and Privacy Act, passed in 1974, while the Children's Online Privacy Protection Act, passed 1998, prohibits the online collection of personal information from children under the age of 13.) "Legally," Friedman concludes, "pri-

Tyler Moore, left, and Allan Friedman

vacy in this country is a mishmash based on the common-law tradition. We don't have a blanket regulation to grant us protection," as Europe does.

Online "Trust Crimes"

There is a pitched battle going on in cyberspace that pits an organized criminal ecosystem of "phishers," "money-mules," and "cashiers" against a jumbled array of private "takedown" firms, official domain-name registrars, and Internet service providers. As Tyler Moore, a postdoctoral fellow at Harvard's Center for Research on Computation and Society explains in an exclusive *Harvard Magazine* Web Extra, the bad guys take over personal computers not for their information, but for their processing power, using "botnets" to stage "fast-flux" attacks that conceal their identity even as they steal the keys to their victims' bank accounts.

Visit harvardmag.com/extras to read more about phishing.

The End of Anonymity

Friedman co-taught a new undergraduate course on the subject of privacy last year; it covered topics ranging from public policy and research ethics to wiretapping and database anonymity. "If there is a unified way to think about what digital systems have done to privacy," he says, it is that they collapse contexts: social, spatial, temporal, and financial. "If I pay my credit-card bill late, I understand the idea that it will affect a future credit-card decision," he explains. "But I don't want to live in a society where I have to think, 'Well, if I use my card in this establishment, that will change my creditworthiness in the future'"—a reference to a recent *New York Times Magazine* story, "What Does Your Credit-Card Company Know about You?" It reported that a Canadian credit-card issuer had discovered that people who used their card in a particular pool hall in Montreal, for example, had a 47 percent chance of missing four payments during the subsequent 12 months, whereas people who bought birdseed or anti-scuff felt pads for the legs of their furniture almost never missed payments. These disaggregated bits of information turn out to be better predictors of creditworthiness than traditional measures, but their use raises concerns, Friedman points out: "We don't know how our information is being used to make decisions about us."

Take the case of someone with a venereal disease who doesn't want the people in his social network to know. "If I go to the hospital and the nurse who sees me happens to live down the street," says Friedman, "maybe I don't want her peeking at my medical records." That particular threat has always been there in charts, he notes, but problems like this scale up dramatically with online systems. Now the nurse could check the records of everyone on her street during a coffee break. He cites a related example: "Massachusetts has a single State Police records system and there have been tens of thousands of lookups for Tom Brady and other local sports stars." Unlike celebrities, ordinary people have not had to worry about such invasions of privacy in the past, but now computers can be used to find needles in haystacks—virtually every time. There are nearly seven billion people on the planet: a big number for a human brain, but a small number for a computer to scan. "John Smith is fairly safe," says Friedman, "unless you know something critical about John Smith, and then all of a sudden, it is easy to find him."

Digital systems have virtually eliminated a simple privacy that many people take for granted in daily life: the idea that there can be anonymity in a crowd. Computer scientists often refer to a corollary of this idea: security through obscurity. "If you live in a house, you might leave your door unlocked," Friedman says. "The chances that someone is going to try your front door are fairly small. But I think you have to lock your door if you live in an apartment building. What digital systems do is allow someone to pry and test things very cheaply. And they can test a lot of doors."

He notes that computers running the first version of Windows XP will be discovered *and* hacked, on average, in less than four minutes, enabling the criminal to take control of the system without the owner's consent or knowledge (see online Extra at www.harvardmagazine.com). Botnets—networks of machines that have been taken over—find vulnerable systems through brute force, by testing every address on the Internet, a sobering measure of the scale of such attacks. (Another measure: the CEO of AT&T recently testified before Congress that Internet crime costs an estimated $1 trillion annually. That is clearly an overestimate, says Friedman, but nobody knows how much Internet crime actually *does* cost, because there are no disclosure requirements for online losses, even in the banking industry.)

The durability of data represents another kind of contextual collapse. "Knowing whether something is harmful now versus whether it will be harmful in the future is tricky," Friedman notes. "A canonical example occurred in the 1930s, when intellectuals in some circles might have been expected to attend socialist gatherings. Twenty years later," during the McCarthy era, "this was a bad piece of information to have floating around." Friedman wonders what will happen when young bloggers with outspoken opinions today start running for political office. How will their earlier words be used against them? Will they be allowed to change their minds?

Because personal information is everywhere, inevitably it leaks. Friedman cites the research of former CRCS fellow Simson Garfinkel, now an associate of the School of Engineering and Applied Sciences and associate professor at the Naval Postgraduate School, who reported in 2003 that fully one-third of 1,000 used hard drives he had purchased on eBay and at swap meets still contained sensitive financial information. One that had been part of an ATM machine was loaded with thousands of credit-card numbers, as was another that a supermarket had used to transmit credit-card payments to its bank. Neither had been properly "wiped" of its data.

Data insecurity is not just accidental, however. *Most* Web-based data transmitted over wireless networks is sent "in the clear," unencrypted. Anyone using the same network can intercept and read it. (Google is the only major Web-based e-mail provider that offers encryption, but as of this writing, users must hunt for the option to turn it on.) Harry Lewis smiled at the naiveté of the question when asked what software the laptop used to write this article would need to intercept e-mails or other information at a Starbucks, for example. "Your computer is all set up to do it, and there are a million free "packet sniffers" you can download to make it easy," he said. And the risk that somebody might detect this illegal surveillance? "Zero, unless somebody looks at your screen and sees what you are doing," because the packet sniffers passively record airborne data, giving out no signals of their presence.

Civil libertarians are more concerned that the government can easily access electronic communications because the data are centralized, passing through a relatively few servers owned by companies that can legally be forced to allow surveillance without public disclosure. Noting that the conversation tends to end whenever privacy is pitted against national-security interests, Friedman nevertheless asks, "Do we want to live in a society where the government can—regardless of whether they use the power or not—have access to all of our communications? So that they can, if they feel the need, drill down and find us?"

Social Changes

Paralleling changes in the way digital systems compromise our security are the evolving social changes in attitudes toward privacy. How much do we really value it? As Lewis points out, "We'll give away data on our purchasing habits for a 10-cent discount on a bag of potato chips." But mostly, he says, "people don't really know what they want. They'll say one thing and then do something else."

Noting young people's willingness to post all kinds of personal information on social networking sites such as Facebook—including photographs that might compromise them later—some commentators have wondered if there has been a generational shift in attitudes towards privacy. In "Say Everything," a February 2007 *New York Magazine* article, author Emily Nussbaum noted:

> Younger people....are the only ones for whom it seems to have sunk in that the idea of a truly private life is already an illusion. Every street in New York has a surveillance camera. Each time you swipe your debit card at Duane Reed or use your MetroCard, that transaction is tracked. Your employer owns your e-mails. The NSA owns your phone calls. Your life is being lived in public whether you choose to acknowledge it or not.... So it may be time to consider the possibility that young people who behave as if privacy doesn't exist are actually the sane people, not the insane ones.

Some bloggers, noting that our hunter-gatherer ancestors would have lived communally, have even suggested that privacy may be an anomalous notion, a relatively recent historical invention that might again disappear. "My response to that," says Lewis, "is that, yes, it happened during the same few years in history that are associated with the whole development of individual rights, the empowerment of individuals, and the rights of the individual against government authorities. That is a notion that is tied up, I think, with the notion of a right to privacy. So it is worrisome to me."

Nor is it the case that young people don't care about privacy, says danah boyd, a fellow at the Law School's Berkman Center for Internet and Society who studies how youth engage with social media. "Young people care deeply about privacy, but it is a question of control, not what information gets out there," she explains. "For a lot of teenagers, the home has never been a private place. They feel they have more control on a service like Facebook or MySpace than they do at home."

She calls this not a generational difference, but a life-stage difference. Adults, boyd says, understand context in terms of physical space. They may go out to a pub on Friday night with friends, but not with their boss. For young people, online contexts come just as naturally, and many, she has found, actually share their social network passwords with other friends as a token of trust

Harry Lewis

or intimacy (hence the analogy to a safe space like a pub).

Teens *do* realize that someone other than their friends may access this personal information. "They understand the collapse of social context, but may decide that status among their peers is more important," she notes. "But do they understand that things like birth dates can be used by entities beyond their visibility? No. Most of them are barely aware that they have a Social Security number. But should they be the ones trying to figure this out, or do we really need to rethink our privacy structures around our identity information and our financial information?

"My guess," boyd continues, "is that the kinds of systems we have set up—which assume a certain kind of obscurity of basic data—won't hold going into the future. We need to rethink how we do identity assessment for credit cards and bank accounts and all of that, and then to try to convince people not to give out their birth dates."

Friedman agrees that financial information needs to be handled differently. Why, he asks, is a credit record always open for a new line of credit by default, enabling fraud to happen at any time? "Is it because the company that maintains the record gets a fee for each credit check?" (Security freezes on a person's credit report are put in place only ex post facto in cases of identity theft at the request of the victim.) Friedman believes that the best way to fight widespread distribution and dissemination of personal information is with better transparency, because that affords individuals and policymakers a better understanding of the risks involved.

"You don't necessarily want to massively restrict information-sharing, because a lot of it is voluntary and beneficial," he explains. Privacy, in the simplest of terms, is about *context* of information sharing, rather than *control* of information sharing: "It is about allowing me to determine what kind of environment I am in, allowing me to feel confident in expressing myself in that domain, without having it spill over into another. That encompasses everything from giving my credit-card number to a company—and expecting them to use it securely and for the intended purpose only—to Facebook and people learning not to put drunk pictures of themselves online." Some of this will have to be done through user empowerment—giving users better tools—and some through regulation. "We do need to revisit the Privacy Act of 1974," he says. "We do need to have more information about who has information about us and who is buying that information, even if we don't have control."

There is always the possibility that we will decide as a society not to support privacy. Harry Lewis believes that would be society's loss. "I think ultimately what you lose is the development of individual identity," he says. "The more we are constantly exposed from a very young age to peer and other social pressure for our slightly aberrant behaviors, the more we tend to force ourselves, or have our parents force us, into social conformity. So the loss of privacy is kind of a regressive force. Lots of social progress has been made because a few people tried things under circumstances where they could control who knew about them, and then those communities expanded, and those new things became generally accepted, often not without a fight. With the loss of privacy, there is some threat to that spirit of human progress through social experimentation."

Jonathan Shaw '89 is managing editor of this magazine.