PHOTONIC CRYSTALS: SEMICONDUCTORS OF LIGHT

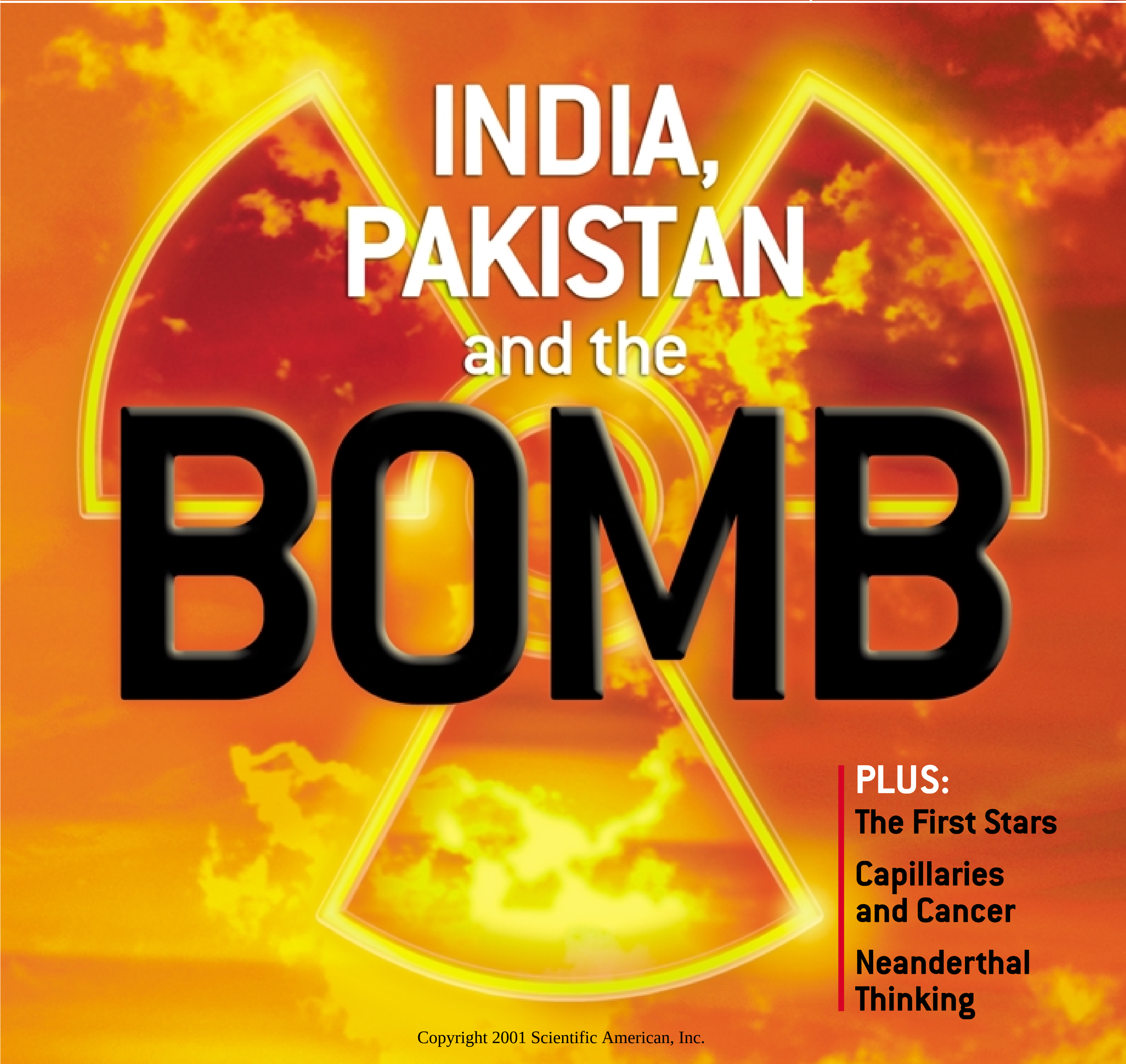
# SCIENTIFIC AMERICAN

**TECHNOLOGY AND TERROR**

*A Special News Report*

(page 20)

DECEMBER 2001 $4.95
WWW.SCIAM.COM



# INDIA, PAKISTAN and the BOMB

**PLUS:**

The First Stars

Capillaries and Cancer

Neanderthal Thinking

# december 2001 contents

SCIENTIFIC AMERICAN Volume 285 Number 6

## features

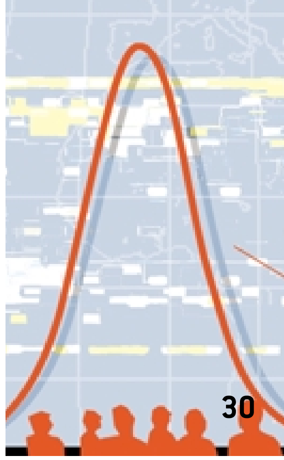


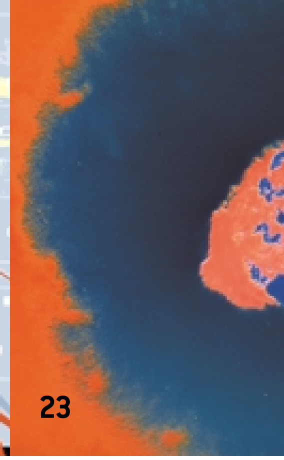
38 Controlling capillary growth

SCIENTIFIC AMERICAN Volume 285 Number 6

# departments









# columns



Cover image by Slim Films; preceding page: Hurd Studios;
this page, clockwise from top left: John McFaul; London School of Hygiene/
Photo Researchers, Inc.; Yorgos Nikas/SPL/Photo Researchers, Inc.

# SA Perspectives

# Here's Looking at You

**Before September 11,** opposition to new electronic surveillance technology in public spaces seemed to be mounting in the U.S. Security cameras were showing up everywhere: at malls, in city parks, along highways. Meanwhile concerned citizens wondered whether these ostensibly benevolent electronic eyes were developing a suspicious squint. When police in Tampa, Fla., revealed that the city's entertainment district was being "patrolled" by 36 video cameras connected to a computerized face-recognition system, a barrage of criticism descended on the city council. At one memorable event, protesters gestured obscenely at the cameras, shouting, "Digitize *this*!"

SECURITY TV monitors in London.

How the times have changed. Today the talk is of more, rather than less, surveillance. Instead of "Big Brother is watching you," we hear "Big Brother is watching out for you." Some pundits opine that the balance between privacy and security must shift in favor of the latter.

The pendulum will undoubtedly continue to swing back and forth. But as we debate the merit of these technologies, we need to keep several questions in mind.

First, how well does the technology really work? The so-called smart closed-circuit television systems are based on software that digitally matches faces with mug shots and ID photos—relying on, for instance, the relative spacing of the eyes. Developers claim an error rate of 1 percent under controlled conditions. But in the real world, people don't usually stand at arm's length from the camera with a sober facial expression and neatly combed hair. A test funded by the U.S. Defense Department last year found that even the best systems choke when the setting changes by just a tiny bit.

Second, what is the technology really being used for? People who favor greatly increased surveillance to combat terrorists may be less enthusiastic when they learn that the technology is more often used to track petty crooks or even innocent citizens. And although the robo-sentinels do not distinguish among, say, racial characteristics, the same cannot be said for the human operators. In England, where tens of thousands of security cameras monitor the streets, a recent study by criminologists at the University of Hull found that "the young, the male and the black were systematically and disproportionately targeted . . . for no obvious reason." Walking while female is another sure way to draw the camera's attention.

At present, the law offers no systematic guidelines to prevent mission creep or outright misuse. Security firms themselves recognize the need for strict rules governing whom to include in a database (or remove, in cases of false positives), how to disseminate the database and how to ensure its security.

Finally, what do we get in return for yielding up more of our privacy? Controversy rages in Britain over the effectiveness of the cameras there, and it is debatable whether new technology would have stopped the terrorists of September 11. Existing computer cross-checks picked up at least two of them; it was the humans who failed to follow through.

Perhaps people will decide to give the cameras a try. If so, we must enact time limits or sunset provisions: the cameras come down and the databases are erased after a specified period, unless we vote otherwise. That way, society can experiment with security cameras without risking a slide toward a surveillance state. The people who decide the balance between security and freedom, justice and privacy, should be the people whose faces appear on the TV monitors.

**THE EDITORS** editors@sciam.com

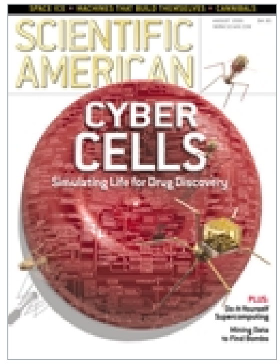
VISIONICS *AP Photo*

# Letters

EDITORS@SCIAM.COM

**"MICHAEL SHERMER'S repeated reference to John Edward as a 'former ballroom-dance instructor' ["Deconstructing the Dead," Skeptic] is argumentum ad hominem of the worst sort," writes Justin Skywatcher of Milledgeville, Ga. "Although I agree that 'psychics' of all stripes are fraudulent and that they prey on the lonely, desperate and bereaved, this tactic is unbecoming. I can just imagine those who debunk Einstein's theory of relativity referring to him as a 'former wanna-be violinist.' Obviously what Edward did before has no bearing on the issue at hand."**

**Go on and give your own reading to the rest of the letters; all are about articles in the August issue.**




## LABOR 101

**Rodger Doyle frets** that the right to strike is denied to government employees and that employees do not enjoy the right to engage in sympathy strikes ["U.S. Workers and the Law," By the Numbers, News Scan]. My understanding, though, is that employees may indeed engage in sympathy strikes in the U.S. unless they have specifically contracted that right away.

MICHAEL S. MITCHELL
Fisher & Phillips LLP
New Orleans

**Doyle asserts** that "labor rights of Americans lag behind those of other nations" just because the U.S. does not adopt "U.N. standard rights." This presupposes several facts that are not beyond dispute and only grudgingly considers that the extra labor rights might "harm the U.S. economy." The question is not just whether there would be harm to the economy but whether there would be harm to U.S. workers and consumers. Rights that drive up the cost of labor arguably cause unemployment and increase the cost of consumer goods, which erodes the standard of living.

KELLEY L. ROSS
Department of Philosophy
Los Angeles Valley College

*DOYLE REPLIES: I use the term "sympathy strike" in its commonsense meaning to denote a strike by a union for the purpose of helping another union in its strike effort. In the specialized world of labor litigation, a sympathy strike occurs when the second union has no material interest in the outcome of the primary strike. Unions engaged in a primary strike rarely ask other unions to walk out purely in sympathy, as such secondary strikes cannot bring economic pressure on the employer. Economically potent sympathy strikes—for example, strikes by the Teamsters in support of the United Auto Workers—are banned.*

*Ross has a valid point in suggesting that the word "rights" has unexamined moral overtones. A more neutral term, such as "legal protections" or "legal powers" or "legal right," might be more appropriate. I cannot, however, agree with him regarding his point on the effect of more rights (or legal powers) on the well-being of consumers and workers in general. Bringing the protections of U.S. workers up to International Labor Organization recommendations would have economic consequences, but given that economic forecasting is less than an exact science, no one can be certain of those consequences. I believe that improvements in legal protections are justified in the interest of fair play.*

## CAFE SUBSTITUTE

**U.S. automakers** didn't change because of CAFE standards ["Another Cup of CAFE, Please," SA Perspectives]; market forces compelled them to improve fuel economy to reacquire market share lost to the Japanese, who were importing much higher efficiency vehicles. If you want to see Detroit improve fuel economy, don't suggest raising the cost of gas to $5 a gallon or jacking up CAFE. Instead implement a tax-discount strategy or credit and offer it to all businesses that use alternative-fuel vehicles or vehicles with high fuel efficiency. Make the incentive lucrative, make it a graduated-scale credit, and the business owner will go looking for the higher-efficiency vehicle.

WILL STANTON
Kissimmee, Fla.

**The trouble with** maintaining different CAFE standards for cars and light trucks is that it encourages automakers to continue making big SUVs instead of big station wagons. This is bad policy, because pound for pound, SUVs are more dangerous to people in cars than other passenger vehicles are. Furthermore, SUVs probably make the roads more hazardous by blocking car drivers' view of the road. All noncommercial passenger vehicles should be required to meet the same CAFE standards.

DAVID HOLZMAN
Lexington, Mass.

## THE RELATIVE MORALITY OF CANNIBALISM

**Anyone who lived** in the 20th century must be aware that about 100 million people were murdered in Europe, Asia and Africa for no other reason than that the ruling group took a dislike to them ["Once Were Cannibals," by Tim D. White]. At least cannibals could claim to derive some physical benefit from the deaths of their victims. Considering the differences between the "civilized world" and our ancestors, the notion of moral progress is at least unclear.

CHARLES KELBER
Rockville, Md.


The Stone SouperComputer

KAY CHERNUSH

## NO SUCH THING AS A FREE COMPUTER

**In "The Do-It-Yourself Supercomputer,"** William W. Hargrove, Forrest M. Hoffman and Thomas Sterling state that as late as May 2001 the Stone SouperComputer still "contained 75 PCs with Intel 486 microprocessors." A high-performance AMD Athlon 1.4-gigahertz system with CPU performance somewhere between 30 and 60 times that of the 66-megahertz 486 systems described in the article can be purchased at today's prices for less than $500. A handful of such systems could easily replace the 75 existing ones, significantly lowering overall cost while improving system reliability.

When you consider that these 75 486 systems consume about 150 watts of power each, in total they use about 270 kilowatt-hours of electricity per day, or about $810 worth of electricity per month at an average cost of 10 cents per kilowatt-hour. If the authors purchased new systems to replace these "free" 486 systems, they could recover their investment in 30 to 60 days in power costs alone.

JOHN H. BAUN
Derwood, Md.

*THE AUTHORS REPLY: The aim of our article was how to minimize construction costs for people who have quantities of surplus PCs and infrastructural access to electricity. There may be an institutional willingness to pay energy costs but a reluctance to purchase equipment using capital monies. Full-cost accounting for supercomputers is a slippery slope. To avoid endless complexities, cost accounting typically includes only hardware and software and excludes operating costs.*

*For problems such as ours, consisting of simple calculations repeated over large data sets, raw CPU speed is not the most significant factor for performance. Using a processor that is twice as fast is unlikely to halve the time it takes to achieve a solution; multiplying bus speeds may be more important. Our measurements indicate that a complete 486-66 machine without a monitor draws 50 watts at full load. The CPU alone from a 1.5-GHz Pentium 4 requires 55 watts. At residential rates, the bill for our 128 nodes is a manageable $300 per month, less at institutional rates.*

## HOW SAFE IS THE CONCORDE?

**"Concorde's Comeback,"** by Steven Ashley [News Scan], masks the inherent reduced safety permitted by the Concorde's government certifiers. Any other four-engine transport aircraft could have sustained the Concorde's damages and made it back for a safe landing. In order to permit the Concorde to operate on existing runways, its certifiers redefined its takeoff safety speed, or $V_2$, to a speed so low that the loss of two engines would not permit the aircraft to climb without first diving a few thousand feet to build up speed. Other four-engine transports have not been afforded this convenient definition of $V_2$ and can in fact lose two engines on takeoff and still climb and maneuver to a safe landing.

JON MODREY
First Officer
Gemini Air Cargo MD11
Orlando, Fla.

## ACADEMIA WITHOUT WIRES

**In "Wireless Wonder"** [News Scan], Wendy M. Grossman mentions that M.I.T.'s campus will be made wireless within the next year. The college I attend, Franklin and Marshall College in Lancaster, Pa., has already had a mostly wireless campus for more than a year, with full coverage opening this semester with a grant from Apple. M.I.T. is not the only school with its eyes set on wireless.

PHILIP Z. BROWN
Chapel Hill, N.C.

---

**ERRATUM** The graph on page 46 of "Code Red for the Web" [October] was created by the CERT©/Coordination Center at the Software Engineering Institute of Carnegie Mellon University.

# 50, 100 & 150 Years Ago
FROM SCIENTIFIC AMERICAN

## Training Babies ▪ Eating Insects ▪ Hunting Bears

### DECEMBER 1951

FUN WITH KIDS—"The human baby is an excellent subject in learning experiments. You will not need to interfere with feeding schedules or create any other state of deprivation, because the human infant can be reinforced by very trivial environmental events; it does not need such a reward as food. Almost any 'feed-back' from the environment is reinforcing if it is not too intense. One reinforcer to which babies often respond is the flashing on and off of a table lamp. Select some arbitrary response—for example, lifting the hand. Whenever the baby lifts its hand, flash the light. In a short time a well-defined response will be generated. Incidentally, the baby will enjoy the experiment. —B. F. Skinner, professor of psychology at Harvard University"

COOL STUFF—"The huge and promising new class of chemicals known as the fluorocarbons has moved from the laboratory to the factory. They are now being produced by the ton in a plant of the Minnesota Mining and Manufacturing Company in Hastings, Minn. The outstanding quality of most fluorocarbons is their tremendous stability; they resist heat, acids, alkalies, insects and fungi."

BATTLEFIELD NUKES—"Five atomic test bombs were exploded by the Atomic Energy Commission last month at its Nevada proving ground. The experiments were designed to provide information on possible tactical uses of atomic weapons. Army troops took part in some of the tests, called 'Exercise Desert Rock.' In one exercise 1,200 paratroopers set up battle positions on the test range, withdrew from the explosion and then returned for lessons in decontaminating the equipment they had left on the site."

### DECEMBER 1901

NOVA PERSEI—"Photographs of the faint nebula surrounding the new star in Perseus have just been received from Prof. G. W. Ritchey of the Yerkes Observatory. The measurement of the negative indicates that the nebula has expanded about one minute of arc in all directions in seven weeks. The rate of motion is, of course, enormous—far beyond anything known in the stellar universe before. Indeed, the motion of the strong condensation of nebulosity approximates *that of light.* —Mary Proctor"

SHELLED MEAT—"Monsieur Dagin, a French Entomologist, recommends certain insects as an article of diet. He has not only read through the whole literature of insect-eating but has himself tasted several hundreds of species raw, boiled, fried, broiled, roasted and hashed. He has even eaten spiders but does not recommend them. Cockroaches, he says, form a most delicious soup. Wilfred de Fonvielle, the French scientist, prefers cockroaches in the larval state, which may be shelled and eaten like shrimp."

WARSHIP DESIGN—"Never before has the United States Navy built a vessel of the great displacement of 14,948 tons. The 'Georgia' was among three of the 'Virginia' class authorized on March 3, 1899. The accepted design, as shown in the accompanying illustration, was only arrived at after controversy in the Naval Board of Construction, prompted by objections to the superposed turret, in which the 8-inch guns are mounted above the 12-inch guns."


U.S.S. *GEORGIA* battleship design of 1901

### DECEMBER 1851

BEAR HUNT—"A paper published at Montauban, Spain, gives an account of the capture of a huge bear by chloroform. His bearship had for a long time been the terror of the district. Early one morning a Dr. Pegot proceeded to the cave where the bear slept, accompanied by a party of peasants. Over the cave entrance they stretched iron bars and blankets, and several times the doctor discharged a large syringe of the somnolent liquid into the interior of the cave. The bear soon fell into a deep sleep, when the doctor marched in and secured his prize in triumph. This is the first instance of the capture of a wild animal by chloroform."

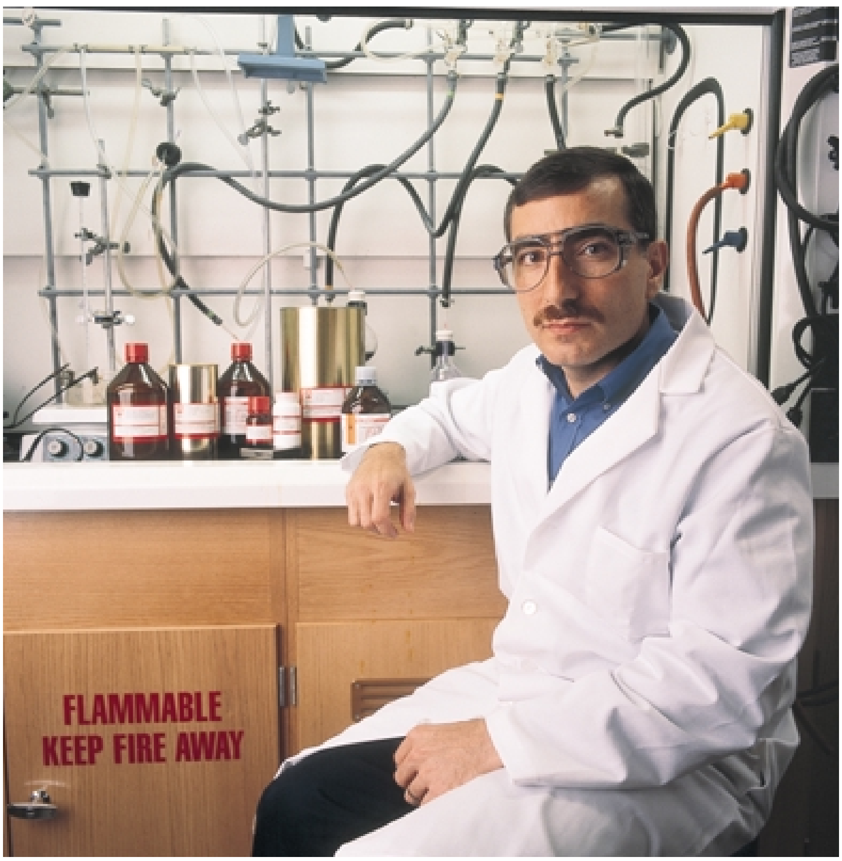
news SCAN

SPECIAL REPORT

[TECHNOLOGY AND TERROR]

# Better Killing through Chemistry

BUYING CHEMICAL WEAPONS MATERIAL THROUGH THE MAIL IS QUICK AND EASY **BY GEORGE MUSSER**




"IT'S A CINCH" to make sarin nerve gas from off-the-shelf chemicals, says chemist James M. Tour.

PAM FRANCIS

**How realistic is terrorism** using chemical weapons? The experts disagree. Some believe it is just too hard to make and disperse deadly gases; others think we shouldn't underestimate terrorists' ability and recklessness. But everyone agrees that we shouldn't make it easy for them. Which is why the experience of James M. Tour is so sobering.

While serving on a Defense Department panel to study the possibility of chemical terrorism, Tour—a Rice University organic chemist famous for co-inventing the world's smallest electronic switches—concluded that nothing stood in the way of someone trying to acquire the ingredients of a chemical weapon. In an article last year in *Chemical & Engineering News,* he argued for restricting the purchase of key chemicals. "They're too easily available," Tour says. "There are no checks and balances."

Unfortunately, the article seemed to fall into the same wastebaskets as previous such warnings. One defense analyst assured Tour that the feds already monitored "every teaspoonful" of potential weapons material.

So Tour decided to do a little test. He filled out an order form for all the chemicals needed to make sarin—the nerve agent used by the Japanese cult Aum Shinrikyo in its

1994 and 1995 attacks—and two of its relatives, soman and GF. His secretary then placed the order with Sigma-Aldrich, one of the nation's most reputable chemical suppliers. If any order should have rung the alarm bells, this one should have.

Instead Tour got a big box the next day by overnight mail. By following one of the well-known recipes for sarin—mixing dimethyl methylphosphonate, phosphorus trichloride, sodium fluoride and alcohol in the right amounts and sequence—he could have made 280 grams of the stuff or a comparable amount of soman or GF. (That's more than 100 teaspoonfuls.) All this for $130.20 plus shipping and handling.


INGREDIENTS for making sarin.

PAM FRANCIS

Nor would delivering the agent be rocket science. To avoid handling poisons, terrorists could build a binary weapon, which performs the chemical reaction in situ. An off-the-shelf pesticide sprayer could then blow the miasma into a building ventilation system. Depending on how well the sprayer worked and how crowded the building was, 280 grams of sarin could kill between a few hundred and tens of thousands of people. The Aum attack on the Tokyo subway involved about 5,000 grams and left 12 people dead, but the cult didn't use a sprayer.

To be sure, Tour is an established name and could probably order just about any chemical from Sigma-Aldrich that he wanted. Most suppliers, however, don't do any screening of their buyers. "You just go to an online distributor, you give them a credit card number, and it comes in the mail," he says. (SCIENTIFIC AMERICAN confirmed this by placing our own order from a small supply house.)

Nerve agent experts agree that something has to be done to keep tabs on such chemicals, especially since the other difficulties of mounting a gas attack seem less daunting after September 11. Says Rudy J. Richardson of the University of Michigan: "Some of the barriers that we might have thought would be there—like, Can terrorists disperse the agent and then escape?—are not there. Today's terrorists don't care if they escape."

Some worry that restrictions would put an undue burden on industry, which has legitimate uses for the chemicals, and wouldn't stop a determined terrorist anyway. But firms already manage with controls on drug-related chemicals, and some protection would be better than no protection. "Everybody points out the ways in which a monitoring system could be bypassed, and I'm the first to agree," Tour says. "But the thing is, right now there's nothing to have to bypass."



### WHAT GOOD ARE GAS MASKS?

**Contrary to some reports, chemists and military experts say that gas masks can protect against nerve gases such as sarin. Although sarin gas can seep through the skin, breathing it delivers a lethal dose about 400 times faster—so the mask could give you enough time to escape from a noxious cloud. The bad news is that you need to know whether the mask really works (surplus units are untested), how to put it on (the fit must be airtight), when to put it on (by the time you recognize the symptoms, it is probably too late) and when to take it off (the masks are too uncomfortable to keep on indefinitely). None of the experts interviewed for this article bothers to own a mask.**

### MORE ON MAIL-ORDER SARIN

**An extended version of this article appears at www.sciam.com/explorations/2001/110501sarin/**

BIOTERRORISM

# Evaluating the Threat

**DOES MASS BIOPANIC PORTEND MASS DESTRUCTION? BY ED REGIS**

**The September 11** terrorist attacks on the World Trade Center and the Pentagon produced a wave of fear that bioterrorism was next on the horizon and, along with it, an impression that the U.S. medical establishment was ill prepared to cope with what would be a vast catastrophe, with millions of Americans lying sick, dead or dying. The death of a Florida man from anthrax and the exposure or infection of others in multiples states further fueled these fears. The resulting wave of general hysteria, with civilians buying up gas masks and Cipro as if there were no tomorrow, established beyond a doubt that microorganisms are remarkably successful as instruments of mass terror. Their potential as weapons of mass destruction, however, is far less clear.

The technology of biological warfare in the modern sense of disseminating viral, bacterial or rickettsial aerosols by means of biological bombs, spray nozzles or other devices goes back at least to 1923. It was then that French scientists affiliated with the

### THE EARLY HISTORY OF CONTAGION

**The idea of using biological organisms as agents of warfare goes back to ancient times. In 400 B.C., for instance, Scythian archers dipped arrowheads in the blood of decomposing bodies, creating poisoned missiles.**

news
SCAN
[TECHNOLOGY AND TERROR]

Naval Chemical Research Laboratory detonated pathogen bombs over animals in a field at Sevran-Livry, 15 kilometers northwest of Paris, killing many of the test subjects.

Between 1943 and 1969, when President Richard M. Nixon terminated it, the U.S. pursued its own major germ warfare program, during the course of which the U.S. Army weaponized (mated with munitions and delivery systems) the causative agents of two lethal diseases, anthrax and tularemia, and three incapacitating diseases, brucellosis, Q fever and Venezuelan equine encephalitis. In addition, the army created military-grade versions of one lethal toxin, botulinum, and one incapacitating toxin, staphylococcal enterotoxin B. It also built and stockpiled more than 2.5 million biological bomb casings, ready to be filled with a biological agent when needed. During those years and afterward, several other nations, including the U.S.S.R., carried on their own germ warfare programs, amassing large amounts of hot agents, munitions and delivery systems.

The most remarkable fact about state-sponsored development of germ weapons during the 20th century, however, is that none of those nations ever used biological weapons on the battlefield, the reason being that although organisms are excellent killing machines, they make poor weapons. For one, because of the long incubation period of many pathogens, the effects of use are not immediate. Second, the resulting epidemic could be mistaken for a natural outbreak of the disease instead of one caused by the enemy. Third, the effect of biological aerosols is uncertain, dependent on chance fluctuations of wind and weather. For all these reasons, biological weapons are not as dramatic, attention-getting, reliable or visually overpowering as conventional high explosives. The possibility of retaliation in kind to a biological attack also acts as a restraint, and there is a sense of moral repugnance attached to the idea of intentionally using living organisms to cause disease, disability or death in human beings.

Nevertheless, none of those deterrents might apply to terrorists, especially to groups acting outside the bounds of traditional moral standards and whose goals are to disrupt and destabilize a society by sowing fear among the populace. Precisely because they are silent, stealthy, invisible and slow-acting, germs are capable of inducing levels of anxiety approaching hysteria. Despite the panic, the history of terrorism is not replete with successful uses of biological (or chemical) agents. Until the death of a photography editor from anthrax in Atlantis, Fla., in October, no death had ever occurred in the U.S. from a biological weapon. But even this incident—and the exposure to or infection by anthrax everywhere from media outlets to post offices to the U.S. Congress—did not amount to a full-scale attack.

The single incident of a semilarge-scale biological attack occurred in 1984, when the Oregon-based Rajneesh cult contaminated restaurant salad bars by dispersing salmonella bacteria, causing 751 cases of diarrhea. (In contrast, accidental food-borne disease incidence in the U.S. is 76 million cases a year, including 315,000 hospitalizations and 5,000 deaths.)

Even if terrorists had the motive to use biological agents and lacked the moral inhibitions that would deter them, they might not have the technological means to do so. Although popular accounts are filled with scenarios of bioterrorists growing lethal bacteria in kitchens, garages and bathtubs or with home brewing kits, the technical expertise required to culture, transport and disseminate a virulent agent in sufficient quantities to cause disease is formidable.

The successful bioterrorist must first obtain a virulent strain of the desired organism (many natural strains of infectious agents are not virulent enough for biological weapons purposes). The chosen pathogen must be cultured in quantity and then be kept alive and potent during transport from place of culture to point of dispersal. It must then withstand the heat and shock of a biological bomb explosion or the mechanical shear forces of being atomized by a nebulizer. Finally, it must be delivered to the target in the proper particle size, over a wide enough geographical area and in sufficient concentration to cause mass infection. All these activities, moreover, must escape detection by antiterrorist law-enforcement agencies. None of those feats is trivial, and it took a group of

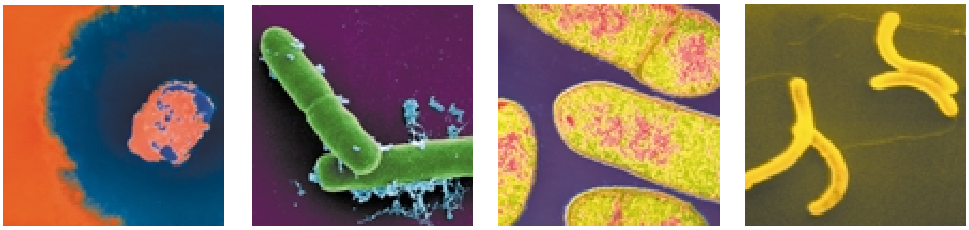
## BIOTERROR: JUST THE FACTS: I

Data from the Monterey Institute of International Studies indicate that 262 biological incidents occurred between 1900 and mid-2001.

Of the 262 incidents, 157 (60 percent) were terrorist cases, and 105 (40 percent) were criminal cases involving extortion or murder attempts not in pursuit of a political objective.

FBI AGENTS in biohazard suits investigate anthrax cases at the American Media building in Florida.

STEVE MITCHELL AP Photo



ROGUES' GALLERY of microbes that could serve as bioweapons includes (*left to right*) the pathogens that cause smallpox, anthrax, botulism and cholera.

LONDON SCHOOL OF HYGIENE/PHOTO RESEARCHERS, INC. (*smallpox, cholera*); GARY GAUGLER (*anthrax*) AND A. DOWSETT (*botulism*)/PHOTO RESEARCHERS, INC.

highly trained American germ warfare researchers more than a decade to produce the first reliable bioweapons delivery system.

In a mid-2000 study of bioterrorist threats against the U.S., Milton Leitenberg of the Center for International and Security Studies at the University of Maryland concluded (1) that hoaxes and threats were more likely than actual use of biological agents; (2) that small-scale sabotage attacks or attempts at personal murder were more likely than large-scale attempts at mass casualties; and (3) that a crude dispersal of a bioagent in a close area was the most likely mode of attack.

These predictions appeared prophetic when the October 2001 anthrax incidents all proved to be small-scale, crude dispersals of anthrax spores by means of delivered mail. It is estimated that those letters contained, in all, less than a gram of anthrax agent—a laboratory-scale amount, insignificant in comparison to what would be needed to mount a mass attack. During the heyday of the American germ weapons program, a U.S. Army production facility at Vigo, Ind., contained twelve 20,000-gallon fermentation tanks, each of them capable of turning out anthrax slurry literally by the ton. Even a small laboratory amount of a "hot" agent could cause a number of casualties if disseminated in an enclosed area such as a subway tunnel; these would not be mass casualties in the sense of millions, hundreds of thousands, or tens of thousands, but the true number is conjectural and unknown.

Even a dispersal of so-called professional, military or weapons-grade anthrax (a loosely defined measure of a hot agent's potential for causing large-scale disease) does not guarantee mass destruction. In 1979 an accident inside a biological weapons production factory in Sverdlovsk, U.S.S.R., caused, by one estimate, 10 kilograms of military-grade anthrax to waft out in a plume over a city of 1.2 million, resulting in a total of 66 fatalities. A mass release of weapons-grade anthrax, therefore, does not necessarily mean mass deaths.

*Ed Regis is author of* The Biology of Doom: The History of America's Secret Germ Warfare Project *(Holt, 1999).*

### BIOTHREATS: JUST THE FACTS: II

Of all bioterror cases from 1900 to mid-2001, 66 percent were outright hoaxes or pranks; 21 percent were threatened attacks that did not materialize by those possessing a bioweapon or else attempted or successful efforts to obtain bioagents; and only 13 percent were actual uses of a bioagent.

Of the actual terrorist attacks using bioagents, 24 percent occurred within the U.S.; of these, no deaths occurred through mid-2001, but several fatalities were registered in October. During the period studied, there were 77 fatalities overseas from both terrorism and criminal incidents.

AIR SECURITY

# Reseizing the Controls

REMOTELY PILOTED HIJACK RESCUES MAY BE A BAD IDEA BY STEVEN ASHLEY

**We've all heard** breathless press reports on what some airline passengers plan to do if suicidal hijackers manage once again to board a flight. But what can aerospace engineers do to foil future attempts to turn airliners into kamikaze guided missiles?

Locking the cockpit door might be all that's needed. The flight deck bulkhead should probably also be reinforced. But the September 11 hijackings have elicited various high-technology solutions as well. One idea that has received much attention would allow a remote operator on the ground to take charge of an airliner should terrorists with flight training get into the cockpit.

It is already possible to control and land an aircraft automatically without the pilot, although such a step is typically taken only in zero-visibility conditions. Most modern aircraft have an autopilot—a computerized system that maintains altitude, speed and direction—that could be reprogrammed to ignore commands from a hijacker and instead

### TELEOPERATION: GROUND PILOT

- Military unmanned aerial vehicles regularly land under remote or autonomous control.
- Remote control of an airplane might cause an accident if it is deployed accidentally.

TOWERING RISK: Auditors have questioned the security of the nation's air traffic control computer system.

RON WATTS Corbis

take direction from the ground to make a safe, automated landing at a nearby airport.

Pilots and the aviation industry in general have reacted coolly to suggestions that direction of an aircraft be wrested from those in the cockpit, however, because of their innate misgivings about handing the controls to a computer. Further, industry experts warn that technology that could override the commands of unauthorized pilots might create greater risks than it eliminates. The system itself could be a terrorist target. Anyone capable of commandeering the ground-to-air communications links necessary to perform remote piloting could produce a disaster without having to risk their life.

A somewhat more feasible approach might be to reprogram the plane's flight computers to make it impossible for an aircraft to fly into buildings because the system would direct it to automatically turn away or climb to avoid them (using altitude measurements or digitized topological maps).

Still, any thought of using the Federal Aviation Administration's existing data communications links to pilot aircraft from afar brings up the troubling vulnerability of the nation's air traffic control (ATC) computers to terrorist takeover. In testimony before the Senate Committee on Commerce, Science and Transportation after the September 11 attacks, Gerald L. Dillingham of the General Accounting Office listed continuing security concerns about the ATC system even before mentioning the much discussed inadequacies of airport security. The FAA "had not ensured that ATC buildings were secure, that the systems themselves were protected and [that] the contractors who access these systems had undergone background checks," he said. "As a result, the ATC system was susceptible to intrusion and malicious attacks. FAA is making some progress in addressing the 22 recommendations we made to improve computer security, but most have yet to be completed."

Some weaknesses identified in GAO reports issued since 1998 could have been serious, Dillingham observed: "For example, as part of its Year 2000 readiness efforts, FAA allowed 36 mainland Chinese nationals who had not undergone required background checks to review the computer source code for eight mission-critical systems."

Only weeks before this testimony was presented, the Department of Transportation's Office of Inspector General (OIG) cautioned the FAA about recent proposals to integrate the air traffic system into the Internet, a change from the current use of dedicated networks. The OIG said that this action could make ATC "more vulnerable to unauthorized intrusion," calling the planned integration a "major risk factor."

In short, much more must be done to avoid a chilling scenario such as the one faced by aviation authorities in *Die Hard 2: Die Harder*, the 1990 Bruce Willis action flick in which terrorists seize control of airport operations by electronically bypassing the control tower and cause one plane to crash. Unfortunately, in a real-life incident John McClane (the movie's unstoppable hero) won't be there to save the day.

## SAFE PASSAGE

**Existing technologies might be adapted to bolster in-flight security, but not without trade-offs.**

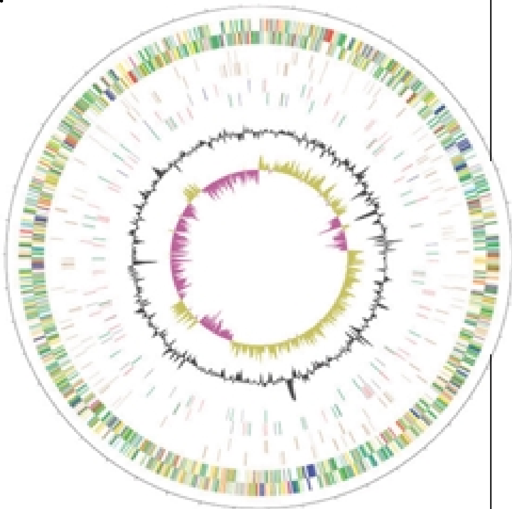
- **Airliners' cockpits could be fitted with biometric scanners that would automatically monitor the face or fingerprints of pilots to ensure that an authorized person is guiding the aircraft. Experts say, however, that it would take a lot of work to install these systems and to make sure that they would not distract pilots.**

- **Sadly, locked cockpit doors must be accompanied by clear rules that would prevent the flight crew from opening the portal even if the passengers and cabin crew were being threatened or killed.**

BIOWEAPONS

# Inside Attacks

**The outbreak of anthrax incidents** has focused new awareness on potential misuses of biotechnology. Moreover, a number of recent research findings point toward methods of fighting the emerging scourge of bioterrorism.



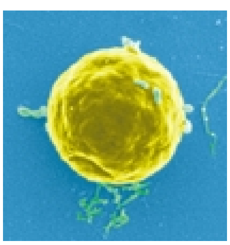
GENOME of *Yersinia pestis,* the plague bacterium, contains 4,012 genes.

## Cure or Poison?

**The economic crisis** that followed the fall of the Berlin Wall in 1989 caused Cuba to depend more on tourism as a way of attracting revenues for the country's faltering economy. It also began to neglect the nurturing of its nascent biotechnology industry, writes José de la Fuente, a former director of research and development at the Center for Genetic Engineering and Biotechnology, in the October *Nature Biotechnology*. Recently Cuba sold to Iran the "prized fruits" of its development efforts, de la Fuente notes: production technologies for several pharmaceuticals, including a recombinant hepatitis B vaccine. "There is no one who truly believes that Iran is interested in these technologies for the purpose of protecting all of the children in the Middle East from hepatitis, or treating their people with cheap streptokinase when they suffer sudden cardiac arrest," he observes in the article. An official from the Cuban Interests Section in Washington, D.C., denied that Cuba would export technology to produce biological weapons, adding that Cuba had itself been a victim of biological attacks, perhaps by Florida-based foes. *—Gary Stix*

## Plague Redux

**Scientists have now** fully sequenced the genome of the bacterium that causes bubonic plague, which killed a third of the population of Europe in the 14th century—and that is feared anew today as a biowarfare agent. Researchers at the Wellcome Trust Sanger Institute near Cambridge, England, and others published their findings in the October 4 *Nature,* giving biologists insight into how *Yersinia pestis* picked up and discarded genetic segments from other bacteria. These events provide a new understanding of how its virulence evolved and may help in the development of new vaccines and drugs against bioweapons or simply aid the roughly 3,000 people worldwide who are diagnosed every year with this endemic disease. *—Gary Stix*



ANTITOXINS may one day combat anthrax (*above*).

## Defusing Anthrax

**The anthrax bacterium** produces a potentially lethal trinary bomb: three proteins combine to form a toxin that can lead to coma and then death. R. John Collier, George M. Whitesides and their colleagues at Harvard University reported in the October *Nature Biotechnology* that they found a peptide that blocks the assembly of the toxin on the surface of the immune cells called macrophages, which are attacked by the bacterium. Rats were protected from 10 times the lethal dose of anthrax toxin. In addition, Collier has collaborated with other groups on a paper to be published in the November 8 *Nature* that identifies the receptor on cells to which the anthrax toxin binds and another paper in the same issue that elucidates the three-dimensional structure of lethal factor, one of the three proteins that make up the toxin. All these findings suggest possible routes to human antitoxins. Antibiotics kill the anthrax bacterium but have no effect on the action of the deadly toxin secreted by the bacterium. *—Gary Stix*

### WWW.SCIAM.COM ON TERRORISM

**Following the attacks on the World Trade Center, a colloquy of structural engineers highlighted the vulnerabilities of ultratall buildings to fire and pointed out steps that could be taken to lessen them.**
www.sciam.com/explorations/2001/100901wtc/

- **Remote-controlled, roach-size tanks could seek out chemical weapons, mines and bombs in hard-to-reach places.**
  www.sciam.com/news/020501/1.html
- **Putting risk-management plans for industrial sites on the Internet could help would-be terrorists attack the facilities.**
  www.sciam.com/1999/0999issue/0999cyber.html
- **A selection of links to articles from *Scientific American* and its Web site and elsewhere appears under the heading of "The Science and Technology of Terror."**
  www.sciam.com/page.cfm?section=terrorism

PHYSICS

# Trillions Entwined

CLOUDS OF ATOMS ARE LINKED BY A WEIRD QUANTUM YOKE BY GRAHAM P. COLLINS

**Schrödinger considered it** the most profound feature of quantum mechanics, and Einstein disbelievingly called it "spooky action at a distance." Entanglement, long just a controversial plaything for theorists, is the weird phenomenon whereby the quantum states of two or more objects become intrinsically entwined in a partnership that in theory would remain unbroken across a distance of light-years. Previously achieved with only a few particles at a time, this marvel has now been demonstrated with two golfball-size clouds of cesium containing trillions of atoms. Eugene S. Polzik and his coworkers at the University of Århus in Denmark entangled the cesium clouds by shooting laser pulses through them. The process will enable robust new ways to teleport quantum states and store information in quantum memories, an essential element of the emerging technology of quantum computation.

An entangled pair of atoms behaves like two magically linked coins. When the coins are flipped, each coin on its own produces heads or tails at random, but when the results for each coin are compared, they are always found to be in cahoots. The coins always match—both heads or both tails. Somehow the coins conspire to achieve this feat even if they are flipped too far apart for any signal or force to travel from one to the other in time to affect the outcome.

In place of coins, physicists use photons or atoms, with polarization states standing in for heads and tails. Cesium atoms, for example, have a magnetic moment that acts like a tiny compass needle that can orient in specific directions in a magnetic field. Alignment with the field corresponds to "heads" and antialignment to "tails." Quantum mechanics also allows superpositions of these states, meaning that the atom is in a combination of both states at once, like a spinning coin ablur in the air. A superposition state specifies the probability of heads or tails. An entangled state specifies joint probabilities—for instance, 50 percent that both coins are heads, and 50 percent that both coins are tails.

Such states generally must be kept extremely well isolated from their surroundings—for example, two atoms might be suspended in a high vacuum by magnetic traps. The slightest interaction with other atoms or even a single photon of light can disrupt the entangled state. Last year Polzik, working with physicists at the University of Innsbruck in Austria, proposed a way to entangle two quantum states that are encoded not on individual atoms but spread across a large ensemble of atoms. The experiment by the Århus team realized that proposal in practice.

Two closed cells of cesium atoms are placed in a magnetic field and prepared in highly ordered initial states. A laser pulse travels through both clouds in succession, producing the entanglement. The beauty of the system is threefold. First, relatively ordinary laser pulses suffice, unlike other schemes. Second, when individual atoms are disturbed, the other trillion or so atoms continue to carry the entanglement, albeit with a little degradation. Those two features lead to the third: the atoms are at room temperature and confined in simple glass cells instead of, say, suspended in exquisite isolation in a very high quality optical cavity. Also, the cells can be far apart.

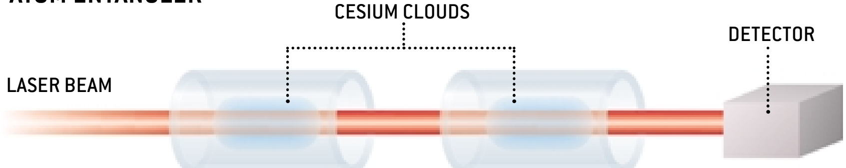
One drawback is that the entanglement is collective—in the coin analogy, it involves an average over a trillion tosses. But for many purposes, such as quantum cryptography, collective entanglement is enough. Polzik expects that his group and others will proceed with relative ease to experiments such as quantum teleportation from one cloud to another and the entanglement of more than two states, a key requirement for the ultimate application that may result from these experiments: general-purpose quantum computing.

## GHOSTLY PARTNERSHIPS

**Other recent entanglement milestones include:**

- **In 2000 a group in Colorado entangled a line of four beryllium ions in a radio-frequency trap by sending a laser pulse through them. In principle, any number of ions could be entangled this way.**
- **A group in France entangled rubidium atoms by passing them one at a time through a superconducting optical cavity and applying microwave pulses to entangle them with the light in the cavity.**
- **This year a group in England used a process analogous to laser amplification to increase the production of entangled quartets of photons by a factor of 16—a first step toward producing a laser of entangled photons.**

**ATOM ENTANGLER**




CESIUM CLOUDS ARE ENTANGLED by the quantum imprint of a laser pulse (*red*) passing through each of them in succession. Detection of a subsequent pulse verifies the effect.

SARA CHEN

MEDICINE

# Stem Cell Showstopper?

WITHOUT CLONING, THEY AREN'T LIKELY TO WORK BY CAROL EZZELL

**New pancreatic cells** for people with diabetes. Regenerated hearts for those who have suffered heart attacks. Repaired spinal cords for paraplegics. These were the hopes in everyone's mind following President George W. Bush's announcement this past August that the federal government would begin providing funds for scientists to study human embryonic stem cells—or at least the 64 colonies of stem cells that have already been isolated in laboratories worldwide.

But immediately after Bush's proclamation, scientists began to question whether all of the 64 existing colonies, or cell lines, were sufficiently established and viable for research. Indeed, U.S. Secretary of Health and Human Services Tommy G. Thompson subsequently admitted before Congress that only 24 or 25 of the lines were ready for use in experiments.

Now some researchers are expressing doubts that any of the stem cell lines will be useful for human therapies. The promise of stem cell research, they say, will be fulfilled only if they are allowed to isolate stem cells from cloned embryos created for individual patients. Under such a scenario, a patient's skin cell would be injected into a donated egg that had been stripped of its genetic material. The fused cell would then be prompted to divide into a clump of cells from which stem cells could be isolated.

Although the current stem cell lines were derived from very early embryos that had not developed beyond hollow balls of cells that fit on the tip of a needle, the cells nonetheless bear proteins on their surfaces that could cause them to be rejected as foreign by the immune system. "We've been saying all along [that stem cells] have to match the patient 100 percent" to be useful therapeutically, says Jose Cibelli, vice president of Advanced Cell Technology in Worcester, Mass., which is pursuing human therapeutic cloning. Even if scientists could generate 1,000 off-the-shelf stem cell lines for use in transplantation, he claims, they would not be able to match the cells to patients closely enough. Recipients would still face rejection risks and would need to take immune-suppressing drugs of the kind given to people with organ transplants. (The problem would not exist for adult stem cells isolated from patients, but these have been hard to find.)

Other investigators point out that even cloned or adult stem cells would not be adequate unless they had their genetic defects fixed before they were given back to a patient. Pancreatic cells derived from stem cells cloned from someone who has diabetes would still contain the genes that contributed to the person's disease in the first place, the researchers maintain. "It's one thing to re-create a pancreas, but if you have to regenerate from diseased tissue, the gene is still defective," says Inder M. Verma of the Salk Institute for Biological Studies in San Diego, Calif. "You have to correct the defect; otherwise cloning will get you what you started out with."

Verma predicts there will be "a hue and cry" for the federal government to fund studies of newly generated stem cells if animal studies using the currently available stem cell lines show promise. Cibelli hopes that one day people will have cloned embryos of themselves created and used to derive stem cells that can be frozen until needed. "It's like buying insurance," he says. Such cells could be the "perfect vehicle" for gene therapy as well, he foresees.

But therapeutic human cloning is a political hot potato right now, with bills forbidding it pending in the House and Senate. Votes on those bills may be postponed until next year because of the terrorist attacks of September 11. In the meantime, a lot of sick people who have read the headlines are pinning their hopes on this potentially revolutionary course of treatment.

EARLY EMBRYOS, such as this one shown on the tip of a needle, may become a source of stem cells. But without cloning, these cells could be useless.

YORGOS NIKAS/SPL/PHOTO RESEARCHERS, INC.

## THE NATIONAL ACADEMY WEIGHS IN

**The specter of immune rejection is "a substantial obstacle" to the use of stem cells for therapies, declared a panel of experts convened by the National Academy of Sciences in a report issued on September 11. The researchers and ethicists raised concerns about the potential health risks of using stem cell lines because such cells could contain mutations and have been grown in the presence of mouse cells, which could harbor viruses. Cloned stem cells "should be actively pursued," the report concluded.**

BY THE NUMBERS

# Why Do Prisons Grow?

FOR THE ANSWERS, ASK THE GOVERNORS **BY RODGER DOYLE**

**The U.S. has gone through** a historically unparalleled expansion in its prison population—from fewer than 400,000 in 1970 to almost 2.1 million in 2000. The expansion continued vigorously even as crime rates fell sharply in recent years. And it has happened at all levels—federal, state and local. For explanations of the causes of the increase, it is helpful to examine the state prisons, which account for 63 percent of all adult prisoners, and the local jails, which account for another 32 percent. (The remainder are held mostly in federal prisons.) Because state laws and policies affect the number of prisoners in local jails, it is proper to consider the two types of institutions together.

The map, which shows prisoners per 100,000 population, points up the uneven distribution of prison populations, such as the fivefold disparity between Texas and Vermont. You would expect that states with high prison populations would have high crime rates, and indeed there is some correlation between the two. But crime rates alone do not explain all the differences among states. Louisiana, for instance, had an incarceration rate 54 percent greater than Mississippi's in 1999, yet Mississippi's crime rate was about the same as or only moderately lower than Louisiana's.

Joseph Dillon Davey of Rowan University has attempted to explain such differences in terms of gubernatorial policy. In an analysis of 14 states, he finds that those in which governors pursue "law and order" policies have higher incarceration rates. An example is South Carolina, where Governor Carroll Campbell, a Republican, presided over a 63 percent expansion of the state prison population in his eight years in office (1987–1995). Governor James G. Martin of North Carolina, also a Republican, did not pursue a tough-on-crime policy. During his administration (1985–1993), there was an increase in the state prison population of only 25 percent, although North Carolina's crime rate was much the same as South Carolina's.

Because Davey's study covers a limited period (the 1980s and early 1990s) and a limited number of states, it cannot be taken as the last word on the subject. Nevertheless, it adds weight to the notion that tough-on-crime policies were the most important factor behind the big increase in prison population since 1970. This increase, which some say did little to deter crime, profoundly disrupted minority communities. Based on current incarceration rates, the Bureau of Justice Statistics estimates that 28 percent of black and 16 percent of Hispanic men will enter a state or federal prison during their lifetime. (The comparable figure for whites is 4 percent.)

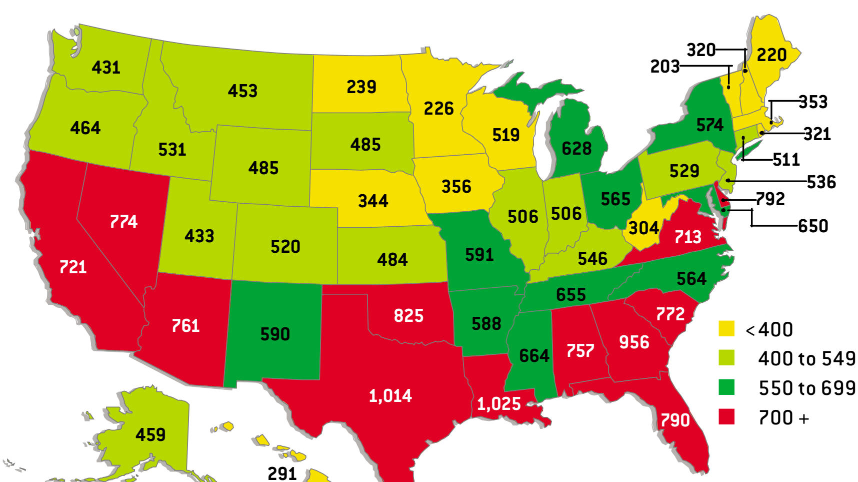
Any effort to understand what happened over the past three decades would benefit from an analysis of state policies and prison trends, the role of local media and other factors that could influence imprisonment rates. This type of study is needed if we are to find answers to such questions as: How significant were tough-on-crime policies in causing the increase in the prison population? To what extent were such policies promoted by those states with a record of racial discrimination? And could the expansion have been avoided without harm to the public?

*Rodger Doyle can be reached at rdoyle2@adelphia.net*

## WHO IS IN THE STATE PRISONS?

| | |
|---|---|
| Men | 94% |
| Women | 6% |
| Whites | 33% |
| Blacks | 46% |
| Hispanics | 17% |
| Violent offenders | 48% |
| Robbery | 14% |
| Murder/manslaughter | 13% |
| Assault | 10% |
| Rape/other sexual assault | 9% |
| Other violent offenses | 2% |
| Property offenders | 21% |
| Burglary | 10% |
| Larceny | 4% |
| Fraud | 3% |
| Motor vehicle theft | 2% |
| Other property offenses | 3% |
| Drug offenders | 21% |
| Public-order offenders | 10% |

STATE AND LOCAL PRISONERS PER 100,000 POPULATION, 1999




SOURCE: Bureau of Justice Statistics

RODGER DOYLE

# The Nobel Prizes for 2001

**In October** the Royal Swedish Academy marked the centennial of the Nobel Prizes. The laureates in each field received a portion of 10 million Swedish kronor, or about $957,000.

### PHYSIOLOGY OR MEDICINE

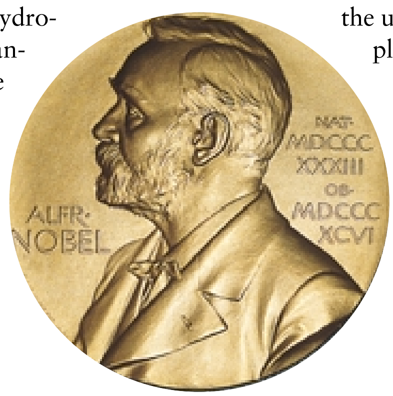
The cell cycle governs how a cell grows and makes copies of itself—and the understanding of this process achieved by this year's laureates is likely to be a major boon to cancer researchers. All the prizewinners uncovered molecules that help to control the cell cycle. In the early 1970s, working with yeast, **Leland H. Hartwell** of the Fred Hutchison Cancer Research Center in Seattle pinpointed more than 100 so-called CDC genes, or cell division cycle genes, including "start," which kicks off the cycle itself. In 1987 **Paul M. Nurse** of the Imperial Cancer Research Fund in London found the start gene in humans, now called CDK 1, or cyclin-dependent kinase 1. His work complemented the efforts of **R. Timothy Hunt**, also of the Imperial Cancer Research Fund, who discovered the first cyclin, a protein that binds to and in turn regulates the activity of CDK molecules.

### PHYSICS

In 1995 **Eric A. Cornell** and **Carl E. Wieman** of the University of Colorado at Boulder, and independently **Wolfgang Ketterle** of the Massachusetts Institute of Technology, produced one of the most sought-after substances in physics: the Bose-Einstein condensate. Named after the two men who postulated its existence, the BEC is a new state of matter in which very slow moving atoms condense into a "superatom" that moves and behaves like one particle. Working with rubidium and sodium gases, the researchers slowed down individual particles by cooling the gases to a tenth of a millionth of a degree above absolute zero. The BEC promises to provide valuable insights into quantum-mechanical processes and may one day be applied to lithography, nanotechnology and ultraprecise measurements.


### CHEMISTRY

Many molecules come in two forms, or enantiomers. Although they are mirror images of each other, the two enantiomers of one molecule can behave quite differently. **William S. Knowles, Ryoji Noyori** and **K. Barry Sharpless** developed catalysts that speed up the production of one enantiomer without its mirror image. These findings have aided in making a wide range of drugs and other products. In 1968 Knowles, working at Monsanto, produced the first catalyst to trigger a reaction that made more of one enantiomer than the other. Some years later Noyori of Nagoya University in Japan created more effective versions of these catalysts, which transfer hydrogen atoms to make an enantiomer. Sharpless of the Scripps Research Institute in La Jolla, Calif., developed catalysts that produce an excess of one enantiomer during oxidation reactions, which transfer an oxygen atom to make an enantiomer.




### ECONOMICS

Why do people distrust used-car dealers? The economics Nobel went to **George A. Akerlof** of the University of California at Berkeley, **A. Michael Spence** of Stanford University and **Joseph E. Stiglitz** of Columbia University for helping to answer this question. Their groundbreaking work explores the theory of markets with asymmetric, or imperfect, information. For example, when purchasing a car, the buyer usually has less information than the seller. Akerlof showed that this type of situation can lead to "adverse selection" when buyers are more likely to choose a "lemon," thereby undercutting confidence in the used-car market. Spence explored how people can avoid adverse selection by having the more knowledgeable side communicate the needed information. Stiglitz, meanwhile, examined what the less informed side can do to learn more.

—*Alison McCook*

AP/WIDE WORLD PHOTO

**WWW.SCIAM.COM/NEWS**
BRIEF BITS

- **Paul M. Nurse, a 2001 Nobelist in medicine, does not expect that science will soon discover a cure for cancer. Find the link to an interview that Nurse gave to *Scientific American* last year at /100901/3.html**
- **More information about the other science-related Nobels is at**
  **PHYSICS: /101001/2.html**
  **CHEMISTRY: /101101/1.html**
  **ECONOMICS: /101201/3.html**

# The Undying Pulse

**Fiber-optic technology nurtured at Bell Labs from before divestiture is ready to go commercial. But will the patience of its creators yield any competitive advantage? By GARY STIX**

**In 1834 John Scott Russell,** a Scottish civil engineer, was riding alongside a canal near Edinburgh when he noticed a curious occurrence. When a horse-drawn barge suddenly stopped, it generated a single wave that continued to move along the canal for kilometers without any change in form or speed. Since Russell's observation, solitary waves, or solitons, have gained a solid mathematical underpinning and remain objects of fascinated study in fields from physics to biology.

The most important practical use for solitons has been in fiber-optic communications; the waves, or pulses, carry digital bits to be transmitted ultralong distances without reconditioning. Much of the groundbreaking research for optical solitons came from Bell Laboratories, the institution that has served as an incubator of technologies ranging from the transistor to the laser. In the next few months the first products of Bell Labs's decades of labors on solitons may finally reach the marketplace. "I'm at long last realizing the dream I've had for the past 15 years," says Linn Mollenauer, who has headed Bell Labs's research effort on solitons.

JOHN McFAUL

Once the current communications industry slump reverses, solitons could become a technological linchpin for a new generation of optical-transmission systems intended to help stem the financial decline of Lucent Technologies, the parent of Bell Labs. But it is unclear whether more than 25 years of nurturing this research will give Lucent any advantage in commercializing solitons. In fact, several companies have already announced soliton-based products.

Solitons in optical communications date back to when Richard Nixon was in the White House and U.S. troops were withdrawing from Vietnam. In 1972 Bell Labs theoretician Akira Hasegawa suggested that nonlinear effects could counteract the dispersion of an optical pulse: light of a certain intensity could interact with optical fiber to offset the tendency of the pulse to broaden over time and eventually overlap with adjacent pulses. A soliton pulse could retain its bell-like shape indefinitely, as long as power is restored to it periodically by processing it through an amplifier.

In 1980 Mollenauer, along with his colleague Roger Stolen, demonstrated the first transmission of a soliton pulse in an optical fiber. Mollenauer became so taken with solitons that he dropped other research he was doing on tunable lasers. A child of the Bell System, he assumed that he could continue his work unimpeded as long as he kept publishing in journals such as *Applied Physics Letters*. "We really weren't required to justify what we were doing," he says.

But in the mid-1980s labors on everlasting pulses nearly came to an end. Arno Penzias, then Bell Labs's vice president of research, launched an effort to bring market relevance to some of the renowned research institution's endeavors. Solitons were on a hit list that also included superconductivity, and Mollenauer was directed to seek out some other line of research within the laboratory. "There were other ways to do the same thing," says Penzias of solitons. "It was too complicated and specialized and not flexible enough."

But Mollenauer was not about to give up so easily. "I was stunned by the news, but I decided to go ahead with an experiment we had been planning anyway," he remarks. In 1988 Mollenauer, along with postdoctoral fellow Kevin Smith, showed how a soliton could retain its original form over the span of 4,000 kilometers. Mollenauer's defiance saved his life's work. "Shortly thereafter Penzias visited the laboratory and apologized. He left me alone after that," he says. But some of Penzias's reservations about soliton transmission were not unfounded. Mollenauer considered solitons ideal for undersea transmission, but engineers found the design of soliton transmitters to be unduly complex. Solitons also turned out to be incompatible with the new generation of optical networks that emerged in the mid-1990s. Such dense wavelength division multiplexing (DWDM) networks can carry billions of bits of digital information on each of the multiple wavelengths in the same fiber. The networks also use equipment that amplified all of these wavelengths simultaneously without the expensive step of converting them first into an electrical signal.

In theory, solitons could have provided another important advance in the push toward the all-optical network—eliminating the costly signal regenerators needed every 500 or 600 kilometers to preserve perfectly shaped pulses. But the dispersion characteristics for DWDM systems made them incompatible with ordinary solitons. In addition, solitons suffer from jitter—random fluctuations in the time of arrival of a pulse at a receiver. So as DWDM and optical amplifiers were deployed commercially, solitons remained in the laboratory.

**Ultimately, other research groups, not Bell Labs, overcame the key technical hurdles that made solitons practical.**

Mollenauer was undaunted, and in 1994 his team demonstrated a record transmission of 40,000 kilometers, equivalent to the circumference of the earth. But it was ultimately other research groups that surmounted key technical hurdles and made solitons practical. In 1995 a team at KDDI in Japan—and later investigators at Aston University in Birmingham, England—reported on a phenomenon in which solitons appear to "breathe." A soliton that alternately broadens and compresses along a stretch of fiber—the optical equivalent of inhaling and exhaling—overcame many of the difficulties encountered with dispersion and jitter. These "dispersion-managed" solitons, as they were more formally called, were so bizarre that many people didn't quite believe they were real. "Every last one of us was steeped in the lore of ordinary solitons, and this seemed against the rules," Mollenauer says.



Lucent has gone on to develop soliton-transmission systems using dispersion-managed solitons and expects to announce new products that use the technology in coming months as replacements for the multimillion-dollar investment in regenerators needed to restore the shape of pulses every 500 or 600 kilometers. Still, it is uncertain whether its decades-long program will give it a clear competitive edge. Lucent promises big things to come. But it is now one of a pack. Several networking companies, including Nortel, Marconi Solstis and Corvis, have already announced their own dispersion-managed soliton products—and others will most likely follow. "The evidence is that companies don't necessarily make the most of their own long-term research," says Nick Doran, chief technology officer for Marconi Solstis. "The opportunity was there, and [Bell Labs] may have missed that opportunity."

More broadly, the soliton story encapsulates how research has evolved over the past quarter of a century. Quasi-academic endeavors in huge industrial laboratories have given way to legions of upstarts, big and small, that plunge ahead on focused development. "Will large-scale, general-purpose research laboratories continue to do this kind of work?" Penzias asks. "It's likely that they won't do much." Mollenauer's undying pulses may be among a dying breed. SA

# Patent Pamphleteer

## Gregory Aharonian's e-mail newsletter decries the issuance of a flood of bad patents while dishing dirt about the goings-on inside the patent office

**He has been called the Matt Drudge** of the patent community—and Greg Erroneous. Since 1993 Gregory Aharonian has distributed a freewheeling e-newsletter several times a week that both irks and tantalizes with its mix of information, invective and gossip. Aharonian makes his living by doing literature searches on the originality of patent applications. But he has made his reputation from his newsletter. Paying the publication costs himself, he attracts 4,500 subscribers, among them patent attorneys, inventors and even some patent examiners. Aharonian talked to SCIENTIFIC AMERICAN's Gary Stix about how he has taken on the stodgy world of patenting.

**How did you develop a reputation as a gadfly?**
Before the mid-1990s, the PTO [Patent and Trademark Office] was really an obscure bureau that nobody paid much attention to other than patent lawyers. Then along comes the Internet, and gadflies like me are talking publicly about the patent world's dirty laundry. The patent office never had to deal with the public. Then here's this Greg Aharonian who was saying that patents were issued without looking at the literature for prior art [previous inventions]. Well, no one ever publicized these things before.

**Why has there been a decline in patent quality?**
The growth rate in applications received by the PTO is higher than its ability to ramp up, so the office is at best treading water and at worst starting to drown. Also, I think the quality of examiners it is hiring has probably diminished. If you're smart enough to examine these patents, I could place you almost anywhere in a high-tech company or a law firm, at least until the recent dot-com crash. On top of that, Congress has been outright stealing PTO fees. The patent office is self-funding; all of the operational money comes from fees from applicants. It goes through the U.S. Treasury, and in theory it should come back to the patent office. Recently Congress has been skimming off the top. If anything, Congress should give it a few extra bucks. You combine these things—the increasing workload, the difficulty in hiring examiners, abusive applicants and less money. I don't care if you're a genius, your quality is going to suffer.

**How does this decline in quality manifest itself?**
A patent claim is a claim on some subset of technology that you want control over. It's a fence. The fence should be no bigger than the thing you've invented. In particular, the fence shouldn't be extended to existing inventions that are quite close or the same. Too many new patents aren't being examined in light of a lot of this relevant prior art.

AHARONIAN shreds bad patents.

ERIC MILLETTE

This is happening because examiners don't have the time and resources to seek prior art, and applicants are refusing to do much searching on their own. So instead of getting an algorithm on data compression that's very narrowly focused, you can get a patent on all data compression, which is nonsense. Then everyone has to go to court and try to figure it all out, which is a waste of time and money.

**Some people are critical of you because they say you sometimes publish unsubstantiated rumors.**
I tell people beforehand that it's gossip. It's up to people to check it out on their own. There's never been a really good mechanism to bring out publicly what's going on behind the scenes, and it belongs in the open. As long as someone passing me the gossip is someone I think is credible, I'll pass it on. Sometimes it's going to be wrong; sometimes it's going to be right. SA

BRAD HINES

# More Baloney Detection

**How to draw boundaries between science and pseudoscience, Part II   By MICHAEL SHERMER**

**When exploring** the borderlands of science, we often face a "boundary problem" of where to draw the line between science and pseudoscience. The boundary is the line of demarcation between geographies of knowledge, the border defining countries of claims. Knowledge sets are fuzzier entities than countries, however, and their edges are blurry. It is not always clear where to draw the line. Last month I suggested five questions to ask about a claim to determine whether it is legitimate or baloney. Continuing with the baloney-detection questions, we see that in the process we are also helping to solve the boundary problem of where to place a claim.

**6. Does the preponderance of evidence point to the claimant's conclusion or to a different one?**

The theory of evolution, for example, is proved through a convergence of evidence from a number of independent lines of inquiry. No one fossil, no one piece of biological or paleontological evidence has "evolution" written on it; instead tens of thousands of evidentiary bits add up to a story of the evolution of life. Creationists conveniently ignore this confluence, focusing instead on trivial anomalies or currently unexplained phenomena in the history of life.

**7. Is the claimant employing the accepted rules of reason and tools of research, or have these been abandoned in favor of others that lead to the desired conclusion?**

A clear distinction can be made between SETI (Search for Extraterrestrial Intelligence) scientists and UFOlogists. SETI scientists begin with the null hypothesis that ETIs do not exist and that they must provide concrete evidence before making the extraordinary claim that we are not alone in the universe. UFOlogists begin with the positive hypothesis that ETIs exist and have visited us, then employ questionable research techniques to support that belief, such as hypnotic regression (revelations of abduction experiences), anecdotal reasoning (countless stories of UFO sightings), conspiratorial thinking (governmental cover-ups of alien encounters), low-quality visual evidence (blurry photographs and grainy videos), and anomalistic thinking (atmospheric anomalies and visual misperceptions by eyewitnesses).

**8. Is the claimant providing an explanation for the observed phenomena or merely denying the existing explanation?**

This is a classic debate strategy—criticize your opponent and never affirm what you believe to avoid criticism. It is next to impossible to get creationists to offer an explanation for life (other than "God did it"). Intelligent Design (ID) creationists have done no better, picking away at weaknesses in scientific explanations for difficult problems and offering in their stead "ID did it." This stratagem is unacceptable in science.

**9. If the claimant proffers a new explanation, does it account for as many phenomena as the old explanation did?**

Many HIV/AIDS skeptics argue that lifestyle causes AIDS. Yet their alternative theory does not explain nearly as much of the data as the HIV theory does. To make their argument, they must ignore the diverse evidence in support of HIV as the causal vector in AIDS while ignoring the significant correlation between the rise in AIDS among hemophiliacs shortly after HIV was inadvertently introduced into the blood supply.

**10. Do the claimant's personal beliefs and biases drive the conclusions, or vice versa?**

All scientists hold social, political and ideological beliefs that could potentially slant their interpretations of the data, but how do those biases and beliefs affect their research in practice? Usually during the peer-review system, such biases and beliefs are rooted out, or the paper or book is rejected.

Clearly, there are no foolproof methods of detecting baloney or drawing the boundary between science and pseudoscience. Yet there is a solution: science deals in fuzzy fractions of certainties and uncertainties, where evolution and big bang cosmology may be assigned a 0.9 probability of being true, and creationism and UFOs a 0.1 probability of being true. In between are borderland claims: we might assign superstring theory a 0.7 and cryonics a 0.2. In all cases, we remain open-minded and flexible, willing to reconsider our assessments as new evidence arises. This is, undeniably, what makes science so fleeting and frustrating to many people; it is, at the same time, what makes science the most glorious product of the human mind. SA

*Michael Shermer is founding publisher of* Skeptic *magazine (www.skeptic.com) and author of* The Borderlands of Science.

# Thawing Scott's Legacy

**A pioneer in atmospheric ozone studies, Susan Solomon rewrites the history of a fatal polar expedition By SARAH SIMPSON**

**Halfway along** her chilly walk from the cafeteria to the laboratory, the young woman's pace slows to a crawl. Since her arrival at Antarctica's McMurdo Station 10 days ago, she has acclimatized surprisingly well. She has come to relish the two-mile stroll, even in temperatures as low as –20 degrees Fahrenheit. Yet today the air feels much more intensely frigid. Her legs start to feel numb, and her jeans turn strangely stiff. Ice crystallizes in the corner of her right eye, and the cold tears at her lungs. She suddenly realizes how lucky she is to be so near the warmth of civilization.

JOANNA PINNEO

## SUSAN SOLOMON: COOL INSIGHTS

- Claim to fame: Led the research team that provided solid evidence tying man-made chlorofluorocarbons to the emergence of the ozone hole over Antarctica.
- Current research: Studying how clouds absorb sunlight to better understand the earth's energy budget.
- Childhood inspiration: Jacques Cousteau. "That's when the 10-year-old kid in me first thought science looked fun."

That day in 1986 atmospheric chemist Susan Solomon truly understood the unremitting hostility of the earth's southernmost continent. The temperature had dipped to a dangerous –50 degrees F; the windchill was below –100 degrees F. Solomon was visiting Antarctica to study trace gases in the atmosphere, but the experience also inaugurated a 15-year investigation into the tragic expedition of Robert Falcon Scott, the English explorer who perished on the ice in 1912 after narrowly losing a race to the South Pole.

Solomon's historical conclusions culminated in *The Coldest March: Scott's Fatal Antarctic Expedition,* published this past September by Yale University Press. The book offers a compelling new explanation for what doomed Scott and four of his men. It was not the explorer's incompetence, as several popular accounts have suggested. It was lethal cold, more severe than what Solomon had experienced at McMurdo. Her analysis of meteorological records—and a careful reading of the expedition diaries—shows that the descriptions of Scott as a poor and unprepared leader were off the mark. "This is a case where science informs history," Solomon asserts. The polar party died during the coldest March on record, when temperatures plunged as low as –77 degrees F.

As the leader of the research team that confirmed the existence of the Antarctic ozone hole, Solomon, now 45, has long been accustomed to looking at the world in a different way. Examining Scott's expedition became a hobby for Solomon as she pursued the studies that definitively linked the man-made chemicals chlorofluorocarbons (CFCs) to ozone destruction in the stratosphere and made the ozone hole one of the most-talked-about environmental issues of the 20th century.

The year before her 1986 walk in the cold, Solomon was already thinking about ozone. While a researcher at the National Oceanic and Atmospheric Administration's Aeronomy Laboratory in Boulder, Colo., she hypothesized that icy clouds in the heart of the stratospheric ozone layer (about 12 miles above the planet's surface) provide the unusual conditions that activate chlorine from CFCs. The stray chlorine atoms then steal oxygen atoms from ozone (a three-oxygen molecule). As the ozone is destroyed, the earth loses much of its protection against harmful ultraviolet radiation, which can promote skin cancer and damage crops.

Multiple measurements from Solomon's Antarctic ozone expedition in 1986 and another in 1987 proved the theory right—and led many scientists to predict correctly that ozone depletion over the midlatitudes was only a matter of time. Her work led to her election to the National Academy of Sciences in 1993 and to the National Medal of Science last year.

ROBERT FALCON SCOTT and four comrades succumbed to extreme Antarctic weather on their return trip from the South Pole in 1912.

At the same time Solomon was implicating CFCs and exploring other aspects of the earth's atmosphere, Scott's expedition began capturing more of her interest. After about 12 years of casually perusing the diaries of Scott and several of his companions, she decided it would be "kind of fun to see what their meteorological data were like," she explains. "That was really when I gained a new level of respect for them." That's also when she first started to find evidence that bad weather, not poor planning, was the greatest factor in Scott's death.

Indeed, Solomon discovered that Scott's team suffered a triple-decker weather disaster while crossing the Ross Ice Shelf, the last leg of their return journey from the pole. That 400-mile crossing should have been the easiest part of their trip. Based on earlier forays and weather measurements, they expected the wind to be at their backs. Expedition meteorologist George C. Simpson also predicted relatively mild temperatures of –10 to –20 degrees F on the shelf. Instead the group encountered average daily minimum temperatures of –34 degrees F, and on only one day of their three weeks on the ice shelf did the temperature rise above –20 degrees F.

"Simpson thought their chances of having weather like that were one in 10," Solomon says. Her analysis of 15 years of meteorological measurements from modern, automated weather stations near Scott's historic path corroborates Simpson's expectations. Just one of those years, 1988, experienced March temperatures persistently that frigid.

Beyond the cold snap, the wind was unexpectedly calm, rendering useless the sails Scott hoped to employ to help move the supply sledges. Each of the men was left to haul a 200-pound sledge through snow that had the texture of gritty desert sand. Again using modern science, Solomon explains why the snow took such a bizarre form: at temperatures below about –20 degrees F, friction no longer melts snow into a slippery layer beneath sledge runners. This trio of conditions was compounded by an unusually long-lived blizzard and a frostbitten foot that eventually halted Scott's ability to walk. He and his last two surviving companions died in a tent only 11 miles from a stash of food and fuel.

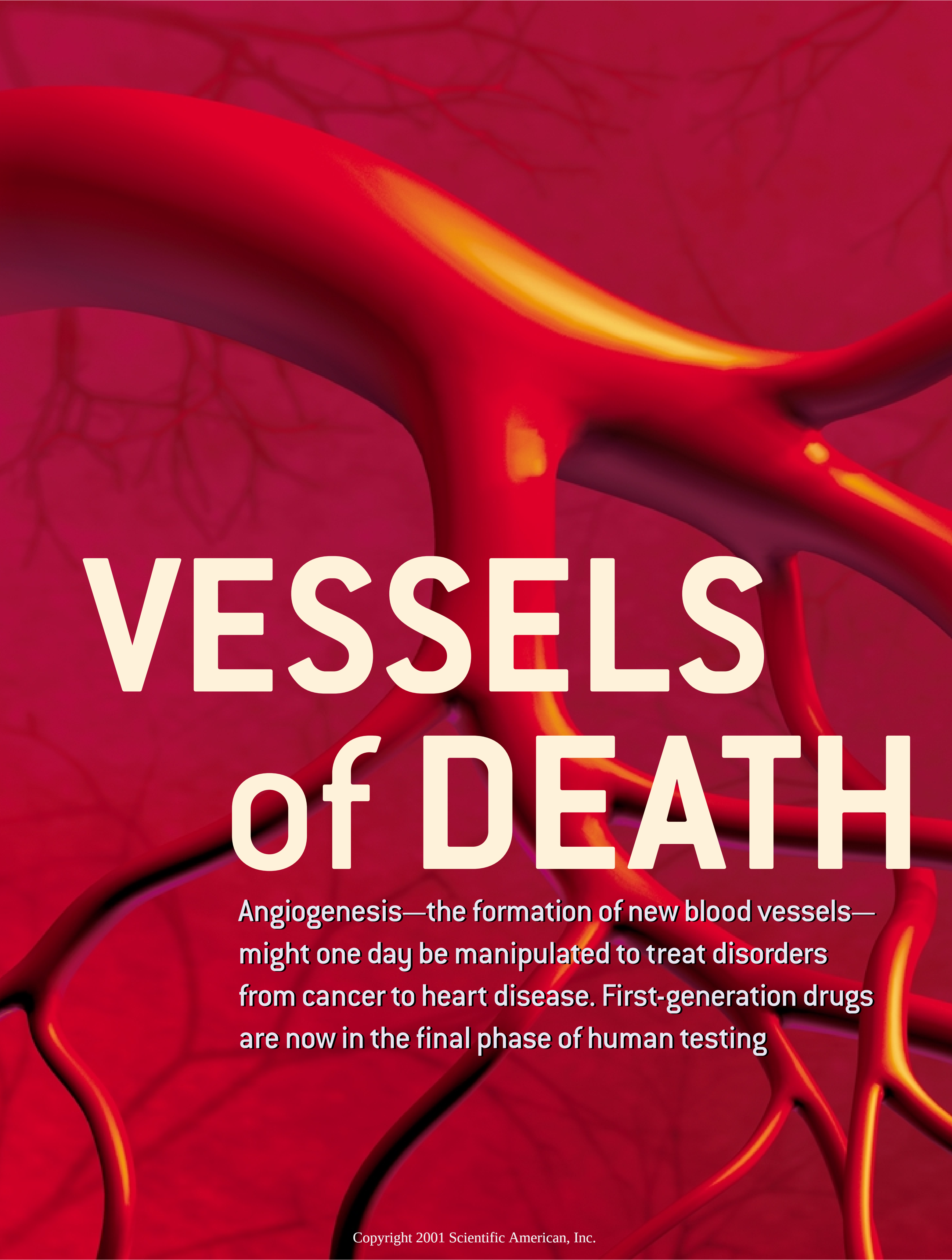
Solomon worked nights and weekends for more than three years to weave these and other findings into *The Coldest March*. "It literally poured out because it was with me for so long," she says. She credits her fiction-writing group—which has met every Tuesday for the past 12 years and includes a rancher, a liquor-store office manager and a homemaker—for helping her make the science understandable to a popular audience.

Still happily obligated to her day job as senior scientist at the Aeronomy Laboratory while writing the book, Solomon was also authoring a 41-page review article on the history of ozone research and flying on research planes to study how clouds absorb sunlight, a critical influence on the earth's energy budget. The crushing loss of a dear friend and fellow ozone researcher in a private plane crash in 1999 pushed her through the last months of writing.

"In some ways, it's a matter of principle for her to soldier on in the face of adversity," says Barry Sidwell, Solomon's husband of 12 years. "She can definitely be determined when she sets her mind to it."

As both scientist and historian, Solomon is driven by her desire to carry her message to a broad audience. "One of our shortcomings as scientists is that we don't always communicate well outside scientific circles," she observes. "When you encounter something new or interesting, I think it's a duty to convey that to the public." SA

BETTMANN/CORBIS



# VESSELS of DEATH

Angiogenesis—the formation of new blood vessels—might one day be manipulated to treat disorders from cancer to heart disease. First-generation drugs are now in the final phase of human testing

# or Life

By Rakesh K. Jain and Peter F. Carmeliet

**They snake through our bodies, literally conveying our life's blood, their courses visible through our skin only as faint bluish tracks or ropy cords. We hardly give them a thought until we cut ourselves or visit a clinic to donate blood. But blood vessels play surprisingly central roles in many serious chronic disorders.**

New growth of the body's smallest vessels, for instance, enables cancers to enlarge and spread and contributes to the blindness that can accompany diabetes. Conversely, lack of small vessel, or capillary, production can contribute to other ills, such as tissue death in cardiac muscle after a heart attack. Accordingly, we and other scientists are working to understand the mechanisms that underlie abnormal vessel growth. This effort will help us develop and optimize drugs that block vessel growth—or improve vessel function.

The study of small vessel growth—a phenomenon referred to generally as angiogenesis—has such potential for providing new therapies that it has been the subject of countless news stories and has received enthusiastic interest from the pharmaceutical and biotechnology industries. Indeed, dozens of companies are now pursuing angiogenesis-related therapies, and approximately 20 compounds that either induce or block vessel formation are being tested in humans. Although such drugs can potentially treat a broad range of disorders [*see boxes on opposite page and on page 43*], many of the compounds now under investigation inhibit angiogenesis and target cancer. We will therefore focus the bulk of our discussion on those agents. Intriguingly, animal tests show that inhibitors of vessel growth can boost the effectiveness of traditional cancer treatments (chemotherapy and radiation). Preliminary studies also hint that the agents might one day be delivered as a preventive measure to block malignancies from arising in the first place in people at risk for cancer.

Results from the first human tests of several compounds that block blood vessel growth were announced earlier this year. Some observers were disappointed because few of the patients, who had cancer, showed improvement. But those tests were designed solely to assess whether the compounds are safe and nontoxic, which they appear to be. Human tests of efficacy are under way and will be a much better judge of whether angiogenesis inhibitors can live up to their very great promise.

## Overview/*Angiogenesis*

- More than 20 compounds that manipulate angiogenesis—either by stimulating new blood vessel growth or by blocking it—are now in human tests against a range of disorders, from cancer to heart disease.
- Angiogenesis inhibitors are generally safe and less toxic than chemotherapeutic drugs, but they are unlikely to treat cancer effectively on their own. Instead physicians will probably use angiogenesis inhibitors in conjunction with standard cancer treatments such as surgery, chemotherapy and radiation.
- The blood vessels of tumors are abnormal. Surprisingly, angiogenesis inhibitors appear to "normalize" tumor vessels before they kill them. This normalization can help anticancer agents reach tumors more effectively.

## The Genesis of Angiogenesis

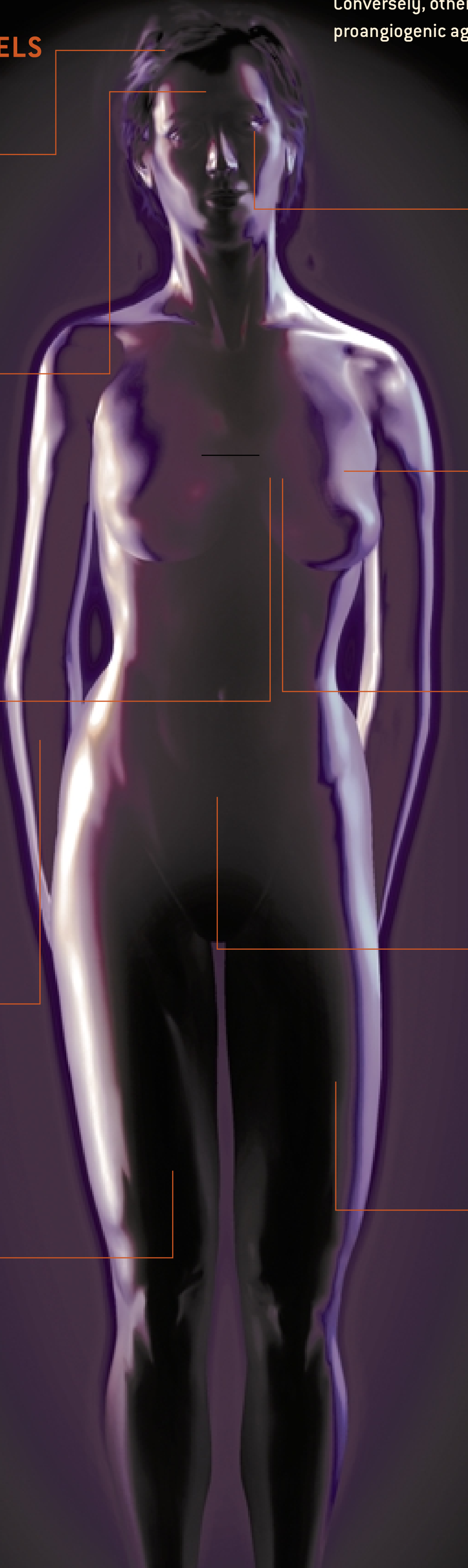
THE TERM "angiogenesis" technically refers to the branching and extension of existing capillaries, whose walls consist of just one layer of so-called endothelial cells. In its normal guise, angiogenesis helps to repair injured tissues. In females it also builds the lining of the uterus each month before menstruation and forms the placenta after fertilization. The development of blood vessels is governed by a balance of naturally occurring proangiogenic and antiangiogenic factors. Angiogenesis is switched on by growth factors such as vascular endothelial growth factor (VEGF) and is turned off by inhibitors such as thrombospondin. When the regulation of this balance is disturbed, as occurs during tumor growth, vessels form at inappropriate times and places.

Cancer researchers became interested in angiogenesis factors in 1968, when the first hints emerged that tumors might release

HURD STUDIOS [*preceding pages*]

# The Two Faces of Angiogenesis

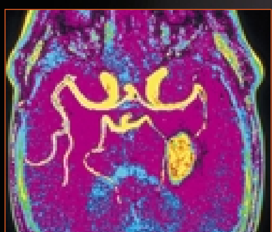
Vessel overgrowth can contribute to a variety of diseases (*right panel*) that could be treatable with angiogenesis inhibitors. Conversely, other disorders (*left panel*) could benefit from proangiogenic agents able to stimulate vessel development.



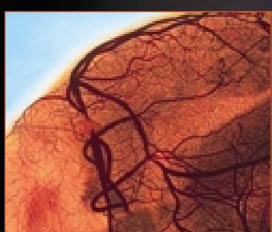
## WHEN EXTRA BLOOD VESSELS COULD HELP ...

**BALDNESS**
Hair follicles depend on a good blood supply

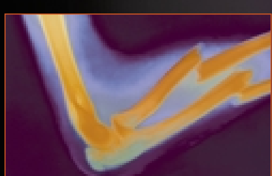
**NEURODEGENERATIVE ILLS**
An increased blood supply could minimize neuronal damage in the brain



**HEART ATTACK** *
New coronary vessels could help repair a damaged heart



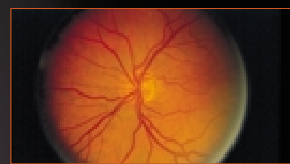
**LIMB FRACTURES**
New blood vessels could help repair broken bones



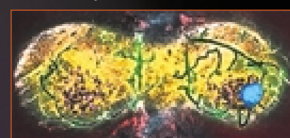
**BLOOD CLOTS IN LEGS** *
Angiogenesis could bypass clots and improve circulation

## WHEN BLOOD VESSELS ARE PART OF THE PROBLEM ...

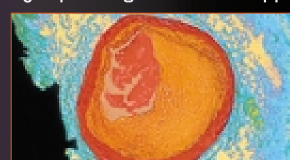
**RETINAL DISEASE** *
Angiogenesis inhibitors could help clear abnormal blood vessels from the eye



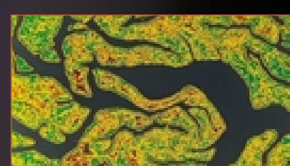
**BREAST (AND OTHER) CANCER** *
Starving cancers of a blood supply could help eradicate them



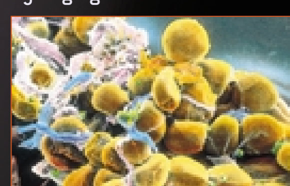
**ATHEROSCLEROSIS**
The plaques that clog vessels may support their own growth by expanding their blood supply



**ENDOMETRIOSIS**
Agents that block angiogenesis could prevent the growth of uterine tissue outside the uterus



**OBESITY**
Fat requires miles of blood vessels, which could be trimmed by angiogenesis inhibitors



* *Human tests are ongoing for these conditions.*

ILLUSTRATION BY KEITH KASNOT; PHOTOGRAPHS BY MICHAEL ABBEY (*hair follicle*); SIMON FRASER (*brain*); SALISBURY DISTRICT HOSPITAL (*broken bones*); SIMON FRASER (*blood clots in legs*); P. MOTTA *La Sapienza* (*adipose tissue*); CECIL H. FOX (*endometriosis*); SIMON FRASER *Royal Victoria Infirmary* (*breast cancer*); STANFORD EYE CLINIC (*retina*). ALL PHOTOGRAPHS FROM SPL/PHOTO RESEARCHERS, INC.

such substances to foster their own progression. Two independent research teams—Melvin Greenblatt of the University of Southern California, working with Phillipe Shubik of the University of Chicago, and Robert L. Ehrmann and Mogens Knoth of Harvard Medical School—showed that burgeoning tumors release a then unidentified substance that induces existing blood vessels to grow into them. Such proliferation promotes tumor growth because it ensures a rich supply of blood loaded with oxygen and nutrients. In 1971 Judah Folkman of Harvard proposed that interfering with this factor might be a way to kill tumors, by starving them of a blood supply. What is more, Folkman later posited that blocking the factor could slow cancer's spread, a process called metastasis, because cancer cells must enter blood vessels to travel to other parts of the body.

## Nipping New Blood Vessels in the Bud

CURRENT TESTS of angiogenesis inhibitors against cancer employ several different strategies. Chief among these is interfering with the action of VEGF. This molecule, which was initially named vascular permeability factor when it was discovered in 1983 by Harold F. Dvorak and his colleagues at Harvard, appears to be the most prevalent proangiogenic factor identified to date. Scientists gained a tool for better understanding the function of VEGF in 1989, when Napoleone Ferrara of Genentech and his co-workers isolated the gene encoding the molecule. In 1996 groups led by Ferrara and one of us (Carmeliet) independently demonstrated the critical role of VEGF in vessel formation by generating mice that lacked one of the normal two copies of the VEGF gene. The mice, which made half the usual amount of VEGF, died in the womb from insufficient and abnormally organized blood vessels.

Researchers are exploring a number of ways to neutralize VEGF's angiogenic activity in patients. These include immune system proteins called antibodies that can bind specifically to and disable VEGF; soluble forms of the cellular receptors for VEGF, to act as decoys that sop up the growth factor before it can bind to cells; and small molecules that can enter cells and block the growth messages that VEGF sends into an endothelial cell's interior after binding to receptors at the surface. The compounds under study also include factors, such as interferons, that decrease the production of VEGF and substances, such as so-called metalloproteinase inhibitors, that block the release of VEGF from storage depots in the extracellular matrix, the "glue" that binds cells together to create tissues.

Although halving the amount of VEGF is lethal to mouse embryos, wiping out cancers in humans with such therapies will probably require the complete neutralization of all the VEGF protein present in a tumor, and that might be difficult to do. VEGF is a potent agent, and trace amounts could protect the endothelial cells from death. But even after all the VEGF is neutralized, a tumor could rely on other proangiogenic factors, such as basic fibroblast growth factor or interleukin-8.

Another widely studied approach for inhibiting angiogenesis in cancer patients is administering or increasing the natural production of antiangiogenic factors. The idea for this therapy emerged when Folkman learned that Noel Bouck of Northwestern University had identified a naturally occurring inhibitor—thrombospondin—in 1989. Surgeons already knew that removing a patient's primary tumor in some cases accelerated the growth of other, smaller tumors—almost as if the primary tumor had secreted something that kept the smaller tumors in check. They have never questioned the necessity of removing the primary tumor in most cases, because such tumors often obstruct the normal functions of organs and tissues, and leaving them in place would provide a source of cancerous cells for yet more metastases. But discovery of a natural angiogenesis inhibitor suggested to Folkman that the primary tumor's secretions might be harnessed as cancer drugs to suppress the growth of both primary and small metastases.

With this concept in mind, Folkman and his colleagues discovered two more of these naturally occurring antiangiogenic substances—angiostatin and endostatin—in 1994 and 1997, respectively. These inhibitors have received a great deal of attention. This is in part because of studies by Folkman's group showing that they can eradicate tumors in mice. A front-page story heralding such successes in 1998 in the *New York Times* increased the visibility of the entire field of angiogenesis.

Clinical trials of angiostatin and endostatin are currently in early stages (experiments involving small numbers of patients to evaluate a potential drug's safety). Preliminary results reported at this year's American Society of Clinical Oncology conference, which were alluded to earlier, indicate that endostatin is safe and causes no side effects. We await the outcome of the various clinical trials of these and other angiogenesis inhibitors in the coming years.

## Going after Established Blood Vessels

THE TWO APPROACHES described thus far interfere with the formation of new blood vessels. But what about preexisting vessels in a tumor? Is it possible to target those without disrupting the established vessels in healthy tissues and organs (an approach termed antivascular therapy)?

Luckily, it turns out that the blood vessels of tumors are abnormal. Not only are they structurally disorganized, tortuous, dilated and leaky, but the cells that compose them display certain molecules on their surfaces from a class known as integrins that are absent or barely detectable in mature vessels. Biologists

THE AUTHORS

*RAKESH K. JAIN* and *PETER F. CARMELIET* bring complementary backgrounds to the study of angiogenesis. Jain, who is now the Andrew Werk Cook Professor of Tumor Biology at Harvard Medical School and director of the Edwin L. Steele Laboratory at Massachusetts General Hospital, started his career as a chemical engineer. He held posts at Columbia University and at Carnegie Mellon University before joining Harvard in 1991. Carmeliet is a professor of medicine at the University of Leuven in Belgium, where he also serves as adjunct director of the Center for Transgene Technology and Gene Therapy at the Flanders Interuniversity Institute of Biotechnology. He received his M.D. from Leuven in 1984 and his Ph.D. from the same institution in 1989.

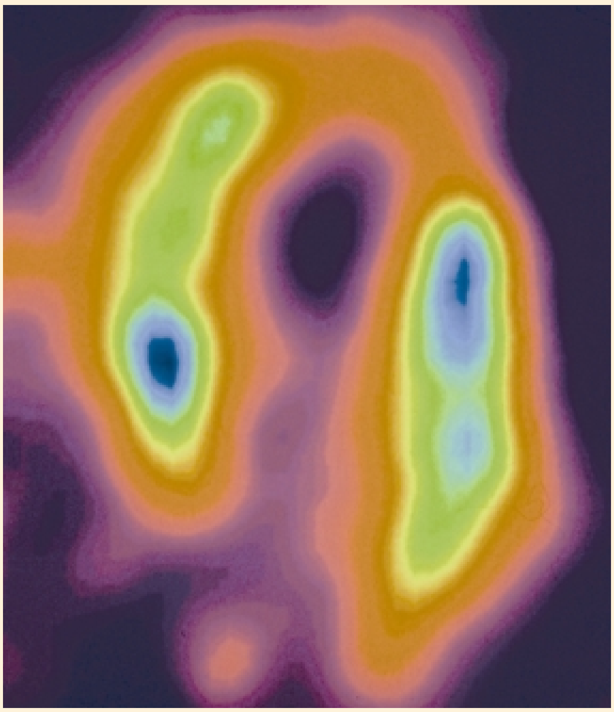
# Therapeutic Angiogenesis

## When making more blood vessels is good for the body

It's easy to understand how restricting the growth of new blood vessels could help kill tumors, but fostering vessel growth—a strategy termed therapeutic angiogenesis—could be useful against other disorders.

Researchers around the world are now evaluating whether the angiogenic substances they are trying to block to treat cancer might help heart attack patients—or those at risk for heart attack—grow new blood vessels in the heart. Those factors might also be used to treat people with vascular disorders in their feet and legs.

A heart attack, properly called a myocardial infarction,



HEART with ischemia (*blue and green areas*)—the oxygen starvation that accompanies heart attacks—could be helped by so-called proangiogenic drugs that stimulate new blood vessel growth.

CNRI *Phototake*

occurs when a blood clot forms in one of the arteries that feeds the heart muscle, preventing part of the heart from receiving oxygen and nutrients, a condition known as ischemia. Unless the clot is dissolved or dislodged rapidly, the patch of heart muscle can die. In addition, many diabetics suffer from a lack of circulation in their extremities caused by occluded blood vessels; some require amputations.

Therapeutic angiogenesis can involve directly administering a vessel growth–promoting substance, such as vascular endothelial growth factor (VEGF). It can also be accomplished using gene therapy, administering to a patient genetically engineered viruses, cells or pieces of DNA that carry the gene encoding VEGF or another angiogenic factor.

Therapeutic angiogenesis with VEGF or fibroblast growth factor (FGF) has been explored for the past 10 years. In 1991 scientists led by Stephen H. Epstein of the National Institutes of Health studied the effects of FGF on the heart vessels of animals. A year later Paul Friedmann and his co-workers at Baystate Medical Center in Springfield, Mass., showed that FGF injections could prompt angiogenesis in the hind limbs of rabbits. In the mid-1990s several groups—including those led by Epstein, Michael Simons of Harvard Medical School, Jeffrey M. Isner of St. Elizabeth's Medical Center in Boston and Ronald G. Crystal of Cornell University Medical School in New York City—demonstrated that therapy involving angiogenic factors or the genes that encode them could stimulate angiogenesis in the hearts and limbs of animals.

Clinical trials aimed at evaluating the safety and efficacy of angiogenic factors in patients are now under way. Carmeliet and others are also testing the therapeutic potential of other promising molecules, such as placental growth factor, a relative of VEGF. Creating functional blood vessels appears to be a formidable challenge, however. Researchers are trying to find the best combinations of such proangiogenic agents as well as the optimal dose, administration schedule and delivery route for the drugs. They are also evaluating whether transplants of endothelial stem cells—the precursors of the endothelial cells that make up blood vessels—can augment the regeneration of blood vessels. Such stem cells can be isolated from the bone marrow of adults.

But potential risks accompany the promise of proangiogenic therapy. Therapeutic angiogenesis could increase a patient's risk of cancer by allowing tiny tumors that had been dormant in the body to gain a blood supply and grow. In addition, because the atherosclerotic plaques that underlie heart disease require their own blood supply as they become larger, therapeutic angiogenesis could backfire as a treatment for cardiac disease by stimulating the growth of plaques that had caused the individual's heart attack in the first place.

Human studies to evaluate the likelihood of these dire scenarios have only recently begun. We hope one day to be able to use genetic tests to evaluate a patient's natural balance of proangiogenic and antiangiogenic factors before beginning to treat them with proangiogenic drugs. This information might also help us understand whether myocardial ischemia results from the insufficient production of angiogenic factors or from the excess production of angiogenic inhibitors. The results will undoubtedly aid in the development of more directed strategies for therapeutic angiogenesis. *—R.K.J. and P.F.C.*

have recently produced small proteins, called RGD peptides, that preferentially recognize the integrins on tumor vessels. These peptides can be linked to cell-killing drugs to target such therapeutic agents to tumors without damaging other tissues. They could also be used to clog the vessels that feed the tumor, by delivering molecules that cause blood clots to form.

But it might not be so easy for any drug to zero in on all a given tumor's blood vessels. The individual cells that make up even a single tumor vessel can vary widely. Studies in one of our labs (Jain's) have found that 15 percent of the blood vessels in human colon cancers are mosaic: some have a particular protein on their surfaces, whereas others do not. If the proteins targeted by new drugs turn out to differ from one tumor to the next or to vary within a tumor during the course of its growth or treatment, this heterogeneity will make it difficult to get therapies that target blood vessels to work on their own.

## Combine and Conquer

MOST LIKELY, surgery or radiation—or both—will continue to be used to attempt to eliminate the original tumor. Today chemotherapy is often administered before or after such therapy to shrink tumors and mop up undetectable malignant cells remaining in the body. Antiangiogenic drugs could well be combined with any of the other approaches to improve the success rate.

Following the pioneering studies of Beverly Teicher of Harvard in the 1990s, several groups have shown the benefits of such a combined approach. Recently Folkman, Robert Kerbel of the University of Toronto and Jain's group have found that combined therapy can produce long-term cures in mice.

Interestingly, antiangiogenic therapy appears to boost the effectiveness of traditional cancer treatments. This is surprising because chemotherapeutic agents depend on blood vessels to reach a tumor, and radiation kills only those cells that have an adequate supply of oxygen (it turns oxygen into toxic free radicals). Logic suggests that by compromising the blood supply of tumors, antiangiogenic therapy would interfere with the effectiveness of these standard treatments. But scientists have demonstrated that the delivery of chemotherapy—as well as nutrients and oxygen—improves during the course of some antiangiogenic therapies.

Indeed, researchers led by Jain have shown that antiangiogenic factors can "normalize" tumor vasculature before killing it by pruning excess, inefficient vessels while leaving efficient vessels temporarily intact. In studies of mice, the researchers found that angiogenesis inhibitors decreased the diameters of tumor blood vessels and made them less leaky, so they began to resemble normal vessels. If such studies pan out in humans, however, physicians will need to work out the optimal dosage and timing of administration.

As is true for many drugs, future generations of antiangiogenic agents are likely to be more effective than the first generation. To optimize future drugs, researchers will need to modify their investigation methods. Most preclinical studies, performed before a drug can be tested in people, are carried out on tumors that are artificially grown under the skin of animals such as mice. But few human tumors arise beneath the skin. To get a more realistic idea of whether a given cancer drug will work in people, researchers will need to study animals with spontaneously occurring tumors growing in more natural sites.

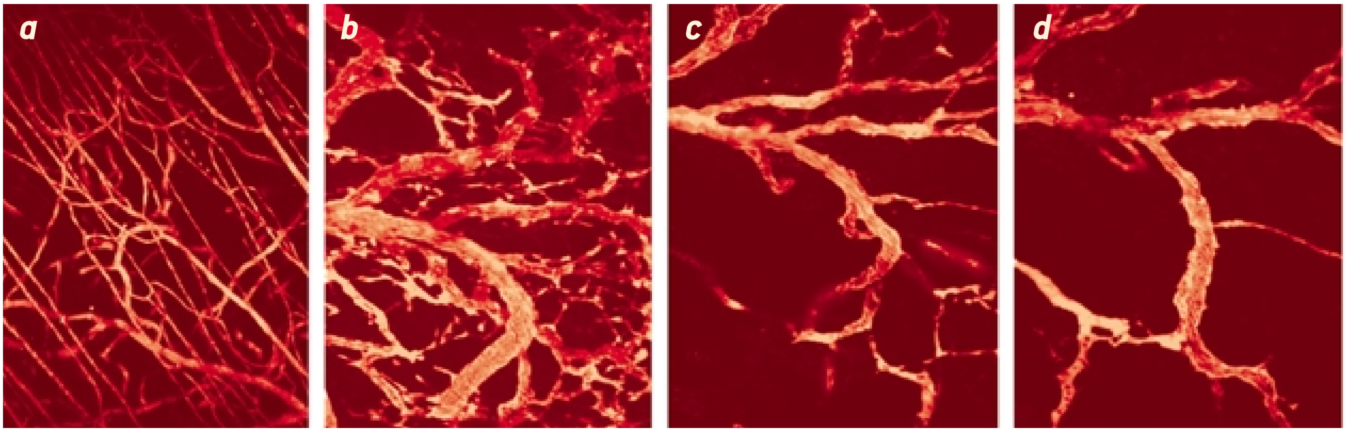
Another limitation of preclinical studies is that they are time-intensive and costly, so researchers usually halt them when tumors begin to shrink but before they can be sure a treatment being tested will actually eradicate the cancers. Because tumors can recur from even a very small number of surviving cancer cells, scientists should follow treated animals for longer periods to better determine the promise of new drug candidates. In addition, investigators tend to begin administering experimental drugs to animals before tumors are fully established, at a time when the cancers are vulnerable—possibly tilting the scales in the drug's favor. Animal tumors also tend to grow more quickly than those in people, and drugs that kill such fast-growing cancers might not be effective against slower-growing human tumors.

Researchers also need to study combinations of antiangiogenic drugs. Cancer cells are masters of evasion. Each tumor produces different combinations of angiogenic molecules that may vary or broaden as they grow. Administering an antiangiogenic drug that blocks only one molecule, such as VEGF, can simply

## ANGIOGENESIS INHIBITORS NEARING THE MARKET

These potential therapies for cancer are in phase III testing, the last stage before Food and Drug Administration approval. Angiostatin and endostatin are in earlier phases of evaluation. Similar compounds are also in trials against the eye disease macular degeneration.

| PRODUCT | DEVELOPER | DESCRIPTION | DISEASE TARGET |
|---|---|---|---|
| Avastin | Genentech | Monoclonal antibody that disables vascular endothelial growth factor (VEGF), a promoter of angiogenesis | Breast and colorectal cancer |
| BMS275291 | Bristol-Myers Squibb | Synthetic compound having multiple effects | Nonsmall cell lung cancer |
| Interferon alpha | Roche, Schering | Protein that inhibits release of growth factors such as VEGF | Various tumors |
| Marimastat | British Biotech | Synthetic compound having multiple effects | Breast and prostate cancer |
| Neovastat | Aeterna | Naturally occurring inhibitor with a range of properties | Nonsmall cell lung and renal cancer |
| SU5416 | Sugen | Synthetic compound that blocks the receptor for VEGF | Colorectal cancer |
| Thalidomide | Celgene | Organic molecule whose specific mechanism of action is unknown | Renal cancer and multiple myeloma |




BLOOD VESSELS change in two useful ways in response to antiangiogenic therapy. Normal vessels (*a*) are well organized and have even diameters, whereas those from a colon cancer (*b*) are dilated and tortuous. Angiogenesis inhibitors prune excess, inefficient vessels (*c*)—which initially "normalizes" the vasculature and helps chemotherapeutic drugs to reach tumors. Eventually, though, increasing numbers of vessels begin to die (*d*).

EDWARD BROWN *Massachusetts General Hospital*

prompt tumors to use another proangiogenic substance to attract a blood supply. In the end, optimal antiangiogenic therapy might consist of a cocktail of several angiogenesis inhibitors.

## An Ounce of Prevention

IF ANGIOGENESIS INHIBITORS fulfill their early promise against cancer, patients will probably need to take them for a long time. The drugs might also be administered as cancer preventatives to people with a high risk of particular cancers—an approach initially suggested in 1976 by Pietro M. Gullino of the National Cancer Institute. Consequently, they must be shown to be safe over the long term. (The drug interferon, an indirect antiangiogenic agent, has been given for years with no side effects to pediatric patients with hemangiomas—benign blood vessel tumors.) The existing human trials will not address this question; they are designed to evaluate safety for just a few months. Animal studies hint that some antiangiogenic compounds might not be safe enough for the long-term administration required to prevent growth or relapse of cancer. Mice that have been genetically manipulated to reduce their production of VEGF can develop neurological defects after a prolonged period, for example, as shown in experiments by Carmeliet.

Insufficient angiogenesis can also impair the heart's recovery from ischemia, tissue starvation stemming from a poor supply of blood. During a heart attack, a blood clot lodges in an artery that supplies the heart muscle, killing a part of the organ. Indeed, researchers are testing agents that spur angiogenesis as treatments for ischemic heart disease. Accordingly, antiangiogenic cancer treatments might increase a patient's risk of ischemic heart disease. As with any therapy, then, physicians and patients will have to carefully weigh the risks and benefits of using angiogenesis inhibitors.

Nevertheless, the burgeoning understanding of angiogenesis has changed our thinking about how to attack cancer. Current treatment with radiation and chemotherapy halts many cancers, but too often the existing treatments bring about only a temporary symptom-free period before the tumor shows up again, spreads throughout the body and kills. Part of the problem is that physicians and pathologists lack reliable, sensitive, cheap and easy-to-use tests that can identify characteristics about each patient's cancer that indicate the best treatment strategy. Genetic analyses of tumors and patients promise to improve the accuracy of diagnoses as well as the efficacy and safety of treatments in the future, but we suspect that within the next 10 or 20 years, better visualization of abnormal vessel structure and function will help as well.

Antiangiogenic approaches have already shown benefit in patients with hemangiomas. As knowledge of tumor angiogenesis progresses, cancers may be detected through elevated levels of angiogenic molecules in the blood—long before clinical symptoms. Physicians may begin to examine patients regularly using molecular tests and new imaging techniques to determine an individual's profile of proangiogenic and antiangiogenic factors.

Based on such tests, doctors will be able to devise treatment plans that, along with other therapies, incorporate a mix of angiogenesis inhibitors appropriate for that individual's tumor. Tests that detect the presence of abnormal vessels will allow doctors to detect possible relapses at an early, potentially treatable stage. Perhaps, as safe oral antiangiogenic drugs are developed and become available, cancer patients will be able to take "a pill a day to keep the cancer away." If so, forms of cancer that are currently untreatable will be reduced to chronic health problems similar to hypertension or diabetes, and many more people will be able to live long, satisfying lives. SA

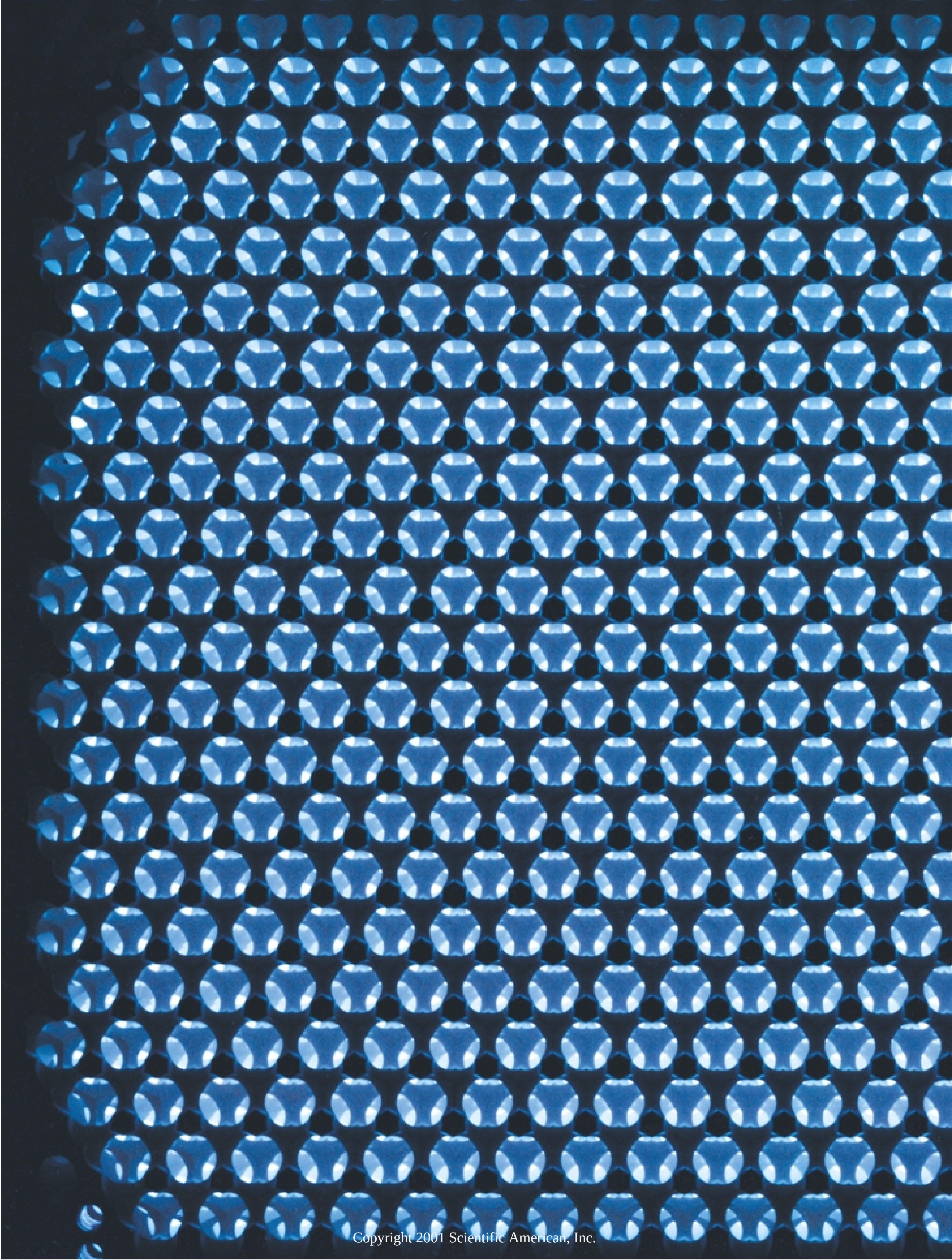
### MORE TO EXPLORE

**An Address System in the Vasculature of Normal Tissues and Tumors.** E. Ruoslahti and D. Rajotte in *Annual Review of Immunology,* Vol. 18, pages 813–827; 2000.

**Angiogenesis in Cancer and Other Diseases.** P. Carmeliet and R. K. Jain in *Nature,* Vol. 407, pages 249–257; September 14, 2000.

**Angiogenesis.** J. Folkman in *Harrison's Principles of Internal Medicine.* Fifteenth edition. Edited by E. Braunwald, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo and J. L. Jameson. McGraw-Hill, 2001.

The National Cancer Institute Web site provides updates on cancer trials that are using angiogenesis inhibitors: **www.cancertrials.nci.nih.gov**



# PHOTONIC CRYSTALS: SEMICONDUCTORS OF LIGHT

By Eli Yablonovitch

Nanostructured materials containing ordered arrays of holes could lead to an optoelectronics revolution, doing for light what silicon did for electrons

**FIRST SUCCESSFUL PHOTONIC CRYSTAL** was formed by drilling three intersecting arrays of holes into a block of a ceramic material. Each array is angled 35 degrees from vertical (into the page), producing a structure now called yablonovite. The pattern of six-millimeter-diameter holes blocks radio waves from 13 to 16 gigahertz.

## IT WAS THE SECOND EXASPERATING PHONE CALL THAT I HAD RECEIVED.

**Yet another group of theorists was saying that my discovery did not work. That was distressing. I had spent three long years trying and discarding countless designs to arrive at what I thought was success, but if the theorists were right, I had to go back to the lab and continue searching. And maybe what I was trying to create—** an artificial crystal structure that could manipulate beams of light in the same way that silicon and other semiconductors control electric currents—was not possible at all.

Electronic semiconductors, of course, are at the heart of all the computers and other devices that pervade the global economy. Semiconductors of light could lead the information and telecommunications revolution still further by enabling higher-capacity optical fibers, nanoscopic lasers and photonic integrated circuits that might one day replace today's microchips.

Indeed, despite a rocky start in the late 1980s and much skepticism from the photonics research community early on, the field of photonic crystals has thrived. Around the world many researchers (including me) have founded companies that are developing commercial products. The key was proving the skeptics wrong by showing that it was possible to create for light the same kind of phenomenon seen in electronic semiconductors—namely, a so-called band gap.

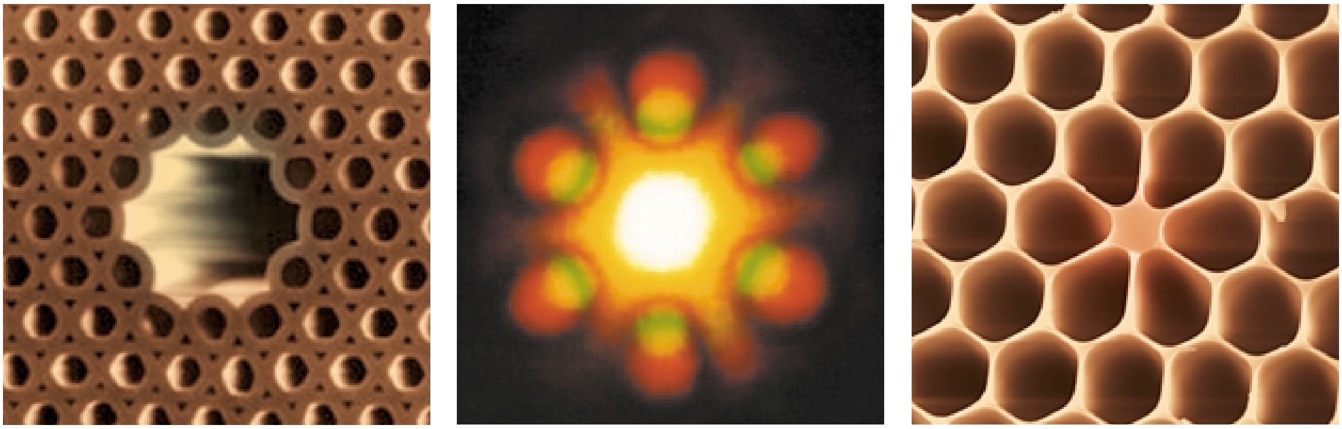
The electronic band gap is a forbidden zone, a narrow range of energies that electrons cannot occupy. When the electrons in the semiconductor fill all the states available to them below the band gap, electric current cannot flow, because each electron has nowhere to go. Boosting an electron above the gap takes a lot of energy. If there are a few excess electrons, however, they automatically must sit above the gap, where they can easily roam through the wide open spaces of empty states. Similarly, a deficit of electrons opens up some positively charged "holes" below the gap, again providing a way for current to flow readily.

All the magic of semiconductors—the switching and logic functions—comes about from controlling the availability of electrons and holes above and below the band gap. The existence and properties of an electronic band gap depend crucially

### Overview/*Photonic Crystals*

- The microelectronics and information revolution is based on the elaborate control of electric currents achieved with semiconductors such as silicon. That control depends on a phenomenon called the band gap: a range of energies in which electrons are blocked from traveling through the semiconductor.
- Scientists have produced materials with a photonic band gap—a range of wavelengths of light that is blocked by the material—by structuring the materials in carefully designed patterns at the nanoscopic-size scale. These photonic crystals function as "semiconductors for light" and promise innumerable technological applications.
- Many researchers greeted the idea of a photonic band gap with skepticism and disinterest when it was first proposed, but today photonic crystals are rapidly turning into big business. Photonic crystals have applications such as high-capacity optical fibers, color pigments and photonic integrated circuits that would manipulate light in addition to electric currents.

COURTESY OF ELI YABLONOVITCH (*preceding pages*)



OPTICAL FIBERS can use the photonic band-gap principle to guide light. The cladding of several hundred silica capillary tubes forms an optical band-gap material that confines light to the central hole, which is about 15 microns in diameter (*left*). In the design at the right, in which the light is confined to the two-micron solid core, the fiber is highly nonlinear, which can be useful for switching and shaping light pulses. In the center, a pattern of colors illustrates how the confinement property of a band-gap fiber varies for different wavelengths of light.

COURTESY OF BLAZE PHOTONICS (*left* and *center*); CRYSTAL FIBRE A/S (*right*)

on the type of atoms in the material and their crystal structure—the spacing and shape of the lattice that they form. By substituting various other atoms (called dopants) into the lattice or its interstices, engineers can dictate the number of electrons or holes in the semiconductor and thereby tailor its properties.

In silicon and other semiconductors, adjacent atoms are separated by about a quarter of a nanometer. Photonic band-gap materials involve similar structures but at larger scales. A typical example would be a block of special glass drilled through with a closely spaced array of cylindrical holes, each with a diameter of 400 nanometers. These openings are analogous to the atoms in a semiconductor. In general, but not always, the spacing of the array must be reasonably close to the wavelength of the light or the electromagnetic waves to be controlled. Visible light has wavelengths ranging from about 400 to 700 nanometers; many cell phones use waves around 35 centimeters long.

Light entering the holey material will refract through and partially reflect off the myriad internal interfaces between air and glass. The complex pattern of overlapping beams will reinforce or cancel one another out according to the light's wavelength, its direction of travel through the crystal, the refractive index of the glass, and the size and arrangement of all the holes. Perfect cancellation in all directions for a narrow band of wavelengths is like the band gap for electrons in semiconductors: that band of light cannot propagate through the crystal. Modifying the band gap structure—for instance, by filling some holes—produces other effects, similar to what can be done by doping electronic semiconductors. Often a photonic crystal is made of an electronic semiconductor material, and so the crystal has both an electronic band gap and a photonic band gap.

## 500,000 Holes

THE QUEST for a photonic band gap originated quietly enough in 1987 with two independent proposals submitted for publication just two months apart: one by me and the other by Sajeev John, then at Princeton University. We had two very different goals in mind. I was at Bell Communications Research, the telephone research consortium in New Jersey, and I was seeking to make telecommunications lasers more efficient. Most of the electric current consumed to produce lasing was wasted as spontaneous light emission, and the photonic band gap could suppress that waste: atoms cannot spontaneously emit light when they are part of a material that forbids light propagation.

John, in contrast, was pursuing a pure research goal. He proposed the photonic band gap to create what is known as light localization. The electronic analogue of this phenomenon, a quantum effect called electron localization, occurs in disordered materials such as amorphous semiconductors. The disorder traps, or localizes, electrons in fixed locations, obstructing current flow.

John and I had never met, but when we learned of each other's proposal, we were curious enough to arrange a get-acquainted lunch. We thought we were onto something, and we agreed to use the same terminology: "photonic band gap" and "photonic crystal." I returned to my lab rather overconfident. I thought that I might create the first working model within only a few months.

Although "photonic" refers to light, the principle of the band gap applies equally well to electromagnetic waves of all wavelengths. Consequently, I could make trial crystal structures with any convenient row spacing and size and then test them with electromagnetic waves of the appropriate wavelength. Indeed, I began

THE AUTHOR

*ELI YABLONOVITCH* was an inventor of the photonic band-gap concept and made the first photonic band-gap crystal while at Bell Communications Research in New Jersey. In 1992 he moved to the electrical engineering department at the University of California, Los Angeles, where he leads the optoelectronics group. He is a founder of two companies in the burgeoning field of photonic crystals: Ethertronics and Luxtera. Before he became a faculty member, Yablonovitch had enough time to sail racing sloops.

my quest for a photonic band-gap material in a machine shop, carving structures out of dielectric plates with a drill. Only human imagination limited the crystal design and structure. Therein lay a problem, however. Out of the innumerable choices available, which design would produce a photonic band gap?

In electronic semiconductor crystals, the band gap arises because electrons behave partly like a wave, and the waves scatter off the layers or rows of atoms. Part of the wave scatters back the way it came, and if the wavelength is about the same as the spacing of successive layers, all the backscattered waves add up coherently. Consequently, the electron's wave is reflected back completely, like light hitting a mirror. For a full band gap, this perfect reflection must occur over a range of wavelengths and for waves heading in any direction through the crystal.

For an electromagnetic band gap, I knew one could not simply emulate a silicon crystal. For light, the scattering is caused by changes of refractive index (for instance, between air and glass), and an interaction directly analogous to electrons and silicon atoms would require a material with an extraordinary refractive index.

Nor could one simply deduce a structure from theory: the band gap depends on how the waves interact with many hundreds of holes, a very complicated process. Theorists had developed computer models for doing the calculations for semiconductors, but these programs could not be used for photons. First, the equations of motion are different—Schrödinger's equation governs electrons, but Maxwell's equations describe the behavior of light. Second, with photons one cannot safely neglect polarization the way one can with electrons. Consequently, I had no way to determine whether a proposed structure would have a photonic band gap. And so, guided as much by physical intuition as calculations, my coworkers and I built structure after structure, searching for the right one. In the course of four years, my loyal machinist, John Gural, drilled more than 500,000 holes in dielectric (insulating) plates, admittedly assisted by a numerically controlled machine. It became unnerving as we produced failure after failure.

## The Surprise of Diamond

WE EXPECTED the face-centered cubic (fcc) structure to be particularly favorable for making electromagnetic band gaps. You can build this structure by taking a checkerboard and placing a black cube on each white square and a white one on each black square. On the second layer, continue placing black cubes on white and vice versa, and so on up. The black cubes (and separately also the white ones) form an fcc lattice.

Unlike lattices of atoms, **photonic crystals** have structural possibilities limited only by the human imagination. **Any shape** can be sculpted at the lattice sites.

That structure still leaves an infinite variety of choices because you can substitute any other geometric shape for the black cubes, which alters how the light waves will be refracted and reflected. After two years, we arrived at something that seemed to work: an fcc structure in which each black cube was replaced by a spherical void in the material. I published this result, but I was mistaken.

By now the theorists had started to catch up, and a few of them had retooled their band-structure computer programs to work with light. Several theory groups, including those led by K. Ming Leung of Polytechnic University and Kai Ming Ho of Iowa State University, began making those dreaded phone calls. My long-sought fcc structure had only a pseudogap: a forbidden "band" having zero width, meaning that just one exact wavelength of light was forbidden. After our years of effort, it appeared that nature might not permit a photonic band gap to exist at all. Perhaps it required a substance with a refractive index far beyond that of any existing transparent material.

Within weeks, however, the Iowa State group found that the diamond structure, the tetrahedral crystal geometry associated with the precious jewel, would produce a band gap. The form that gives the widest band gap consists of dielectric rods in the positions of the chemical bonds between carbon atoms, with the atoms shrunk to geometric points. Diamond itself is not a photonic band-gap material, as far as we know. Earlier in this piece I said that when we began our research, we knew we could not simply emulate the silicon crystal structure to produce a photonic band gap. How wrong we were: silicon's crystal structure is precisely that of diamond.

That the tetrahedral structure is the best for making a photonic band gap is startling and profound. Before the advent of photonic crystals, the diamond configuration was merely another mineral structure, arising out of a complex interplay of atoms, chemical bonds and energy minimization under suitable conditions of temperature and pressure. Its utility for forming a photonic band gap, which emerges entirely and solely from Maxwell's equations (the laws of electricity, magnetism and light), shows that the diamond configuration also has fundamental significance in relation to electromagnetism and the geometry of three-dimensional space.

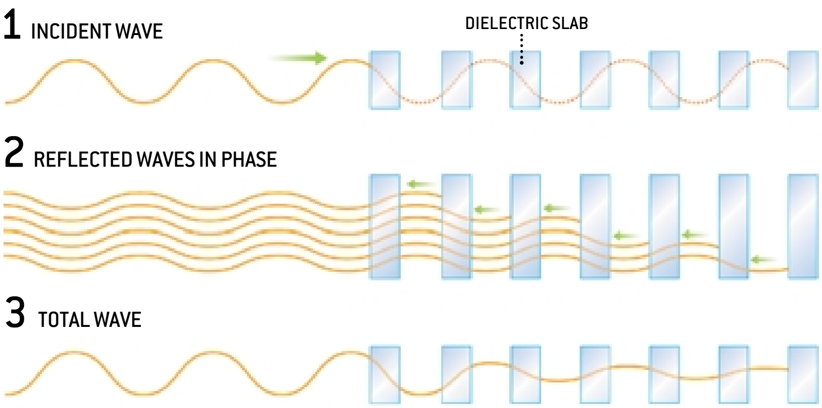
Diamond's tetrahedral structure takes on many different appearances according to what shape is placed in each lattice site and from which angle the crystal is viewed. The box on the opposite page includes two very dissimilar photonic crystals that are based on the diamond structure. My group made the first successful photonic band-gap crystal (this time for real) in 1991 using a variant of the diamond structure now called yablonovite. Nature is kind after all: a band gap occurs in the diamond structure for a refractive

# MAKING BAND GAPS IN ALL DIMENSIONS

## ONE DIMENSION

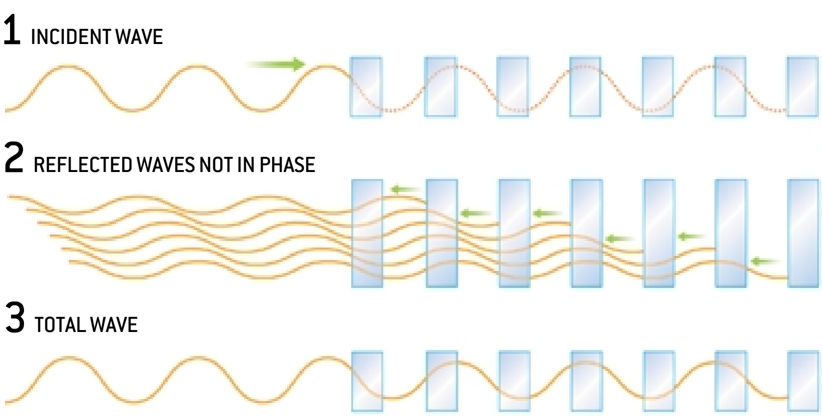
### FOR WAVELENGTH IN BAND GAP

A wave incident on a band-gap material (*1*) partially reflects off each layer of the structure (*2*). The reflected waves are in phase and reinforce one another. They combine with the incident wave to produce a standing wave (*3*) that does not travel through the material.



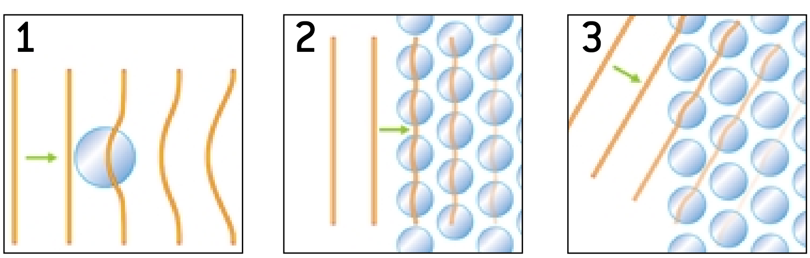

### FOR WAVELENGTH NOT IN BAND GAP

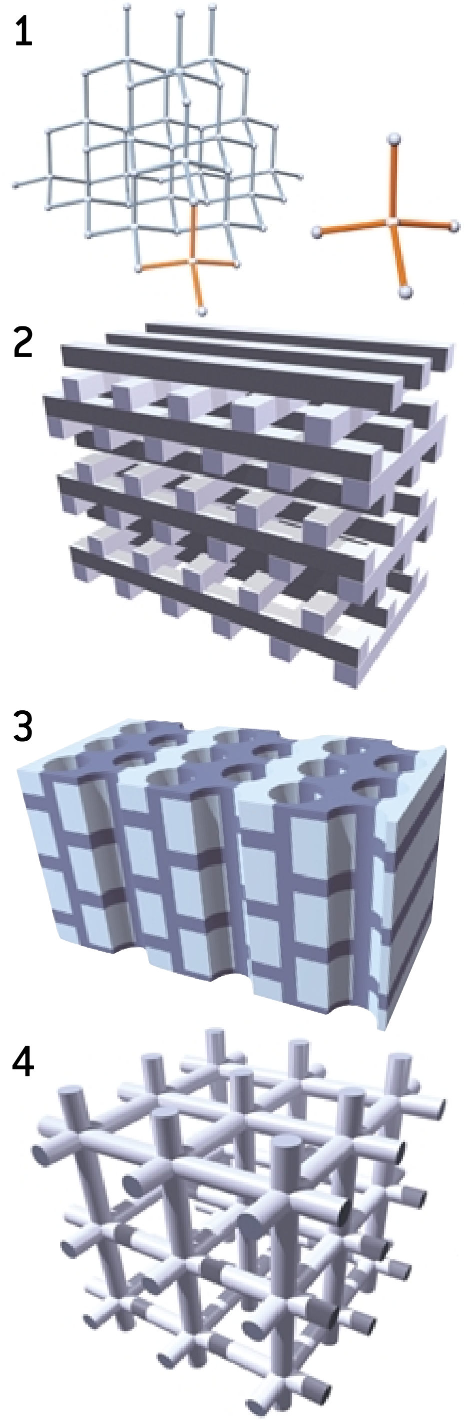
At a wavelength outside the band gap (*1*), the reflected waves are out of phase and cancel out one another (*2*). The light propagates through the material only slightly attenuated (*3*).




## TWO DIMENSIONS

For a two-dimensional band gap, each unit cell of the structure (*1*) produces reflected waves (*not shown*) and refracted waves that must combine to cancel out the incoming wave (*2*) no matter what direction it is traveling (*3*). A full three-dimensional band-gap material works the same way but in all three dimensions.




## THREE DIMENSIONS

Diamond's tetrahedral configuration (*1*) is the most effective geometry for making three-dimensional band-gap materials. This geometry occurs in disguised form in yablonovite (*see pages 46 and 47*), the "stack of logs" (*2*), and this design (*3*), which uses silicon dioxide channels (*light*) in silicon (*dark*). The scaffold structure (*4*) is a rare example that has a different underlying symmetry, but it has only a small band gap.

SLIM FILMS

index as small as 1.87, and many optical materials are available with refractive indices as high as 3.6.

The diamond structure isn't the only structure having a photonic band gap. In 1992 theorist Joseph W. Haus, then at Rensselaer Polytechnic Institute, showed that we had discarded the fcc structures too quickly. Scientists had searched the fcc structures for band gaps only at wavelengths for which about half a wave fits in one cell of the lattice (somewhat like the fundamental vibration of a guitar string). As we saw, only a pseudo-gap occurs at that frequency. Haus, however, also considered a higher frequency, for which a full wavelength fits in a cell (somewhat like the first harmonic of the guitar string), and proved that an fcc band gap would indeed emerge there. In addition, he discovered that even the simple cubic configuration known as the scaffold structure (for its similarity to scaffolding) could have a band gap, albeit a small one.

## Butterflies and Microchips

WE HAVE NOW LEARNED that nature already makes photonic crystals in the sparkling gem opal, in a butterfly's colorful wings and in the hairs of a wormlike creature called the sea mouse. Each of these has a photonic band structure, though not a full band gap, in that light can still propagate in some directions. A complete band gap has eluded nature, perhaps because it requires too much refractive-index contrast.

Nevertheless, an incomplete band gap can be very useful. For example, titanium dioxide particles smaller than a micron can be made to self-assemble in the opal structure. Titanium dioxide is the intensely white pigment used in paint and to make paper white. The coherent scattering of light that occurs from band-gap-structured titanium dioxide can impart more whiteness for less mass of titanium dioxide. One day photonic crystals may be all around us in the painted walls and in the stacks of paper cluttering our desks.

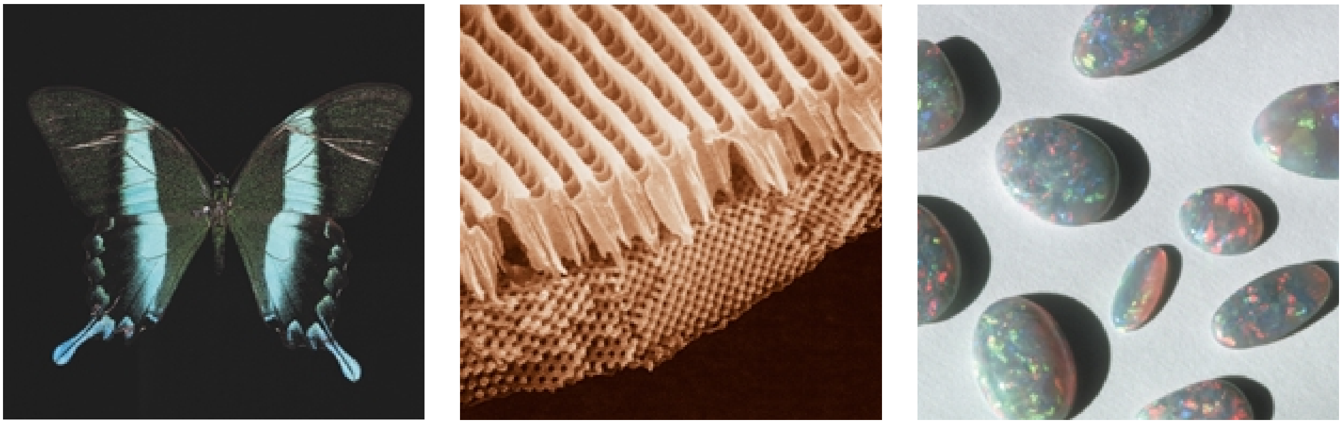
Another very useful type of incomplete band gap material is that of two-dimensional photonic crystals, which can block light from traveling within a plane. Such a structure can be stretched along the third dimension, forming a new kind of optical fiber. Conventional optical fibers have a high refractive index at their core, which confines light by total internal reflection. Philip St. J. Russell of the University of Bath in England demonstrated in 1999 how to make photonic band-gap fibers. In one version, light travels along a central hole in the fiber, confined there by the two-dimensional band gap of the surrounding material. More optical power can be sent through such a central void than through glass, enabling greater information-carrying capacity, perhaps 100 times that of conventional telecommunications fibers. Specialty fibers have advanced the most as commercial photonic band-gap products. Companies in Denmark and the U.K. have already distributed sample quantities and will soon begin volume production.

Instead of stretching out a two-dimensional band-gap structure to make a fiber, one can go to the other extreme and make a two-dimensional thin-film photonic crystal, as was first calculated in 1997 by Shanhui Fan and John D. Joannopoulos, then both at the Massachusetts Institute of Technology. Thin-film photonic crystals can be easily patterned by standard methods used to produce integrated circuits. Introducing

**Integrated circuits that combine conventional electronics and photonic crystals would represent the ultimate limit of optoelectronic miniaturization.**

## APPLICATIONS FOR PHOTONIC CRYSTALS

| DEVICE | DESCRIPTION | STATUS |
|---|---|---|
| OPTICAL FIBERS | 2-D band-gap material stretched along the third dimension | Early versions already commercialized |
| NANOSCOPIC LASERS | World's tiniest optical cavities and tiniest lasers; formed in a thin-film 2-D band-gap material | Demonstrated in the lab |
| ULTRAWHITE PIGMENT | Incomplete 3-D band-gap material, usually patterned as opal structure | Demonstrated; low-cost manufacturing methods under development |
| RADIO-FREQUENCY ANTENNAS, REFLECTORS | Uses inductors and capacitors in place of ordinary dielectric materials | Demonstrated for magnetic resonance imaging and antennas |
| LIGHT-EMITTING DIODES | Photonic band-gap structure can extract light very efficiently (better than 50%) | Demonstrated, but must compete with other methods of achieving the same goal |
| PHOTONIC INTEGRATED CIRCUITS | 2-D thin films can be patterned like conventional integrated circuits to make channel filters, modulators, couplers and so on | Under development |



NATURAL PHOTONIC BAND GAPS occur in some butterfly wings (*left*) and in opals (*right*). In both cases, the band gap is incomplete—it is not effective in every direction—but it produces iridescent colors. A micrograph of a fractured iridescent green butterfly scale (*center*) shows the submicron-size face-centered cubic structure inside. Opals consist of submicron-size silica spheres arranged in a face-centered cubic (close-packed) structure.

PAUL BEARD (*left*); HELEN GHIRADELLA, FROM THE *ANNALS OF THE ENTOMOLOGICAL SOCIETY OF AMERICA*, VOL. 78, 1985 (*center*); THE PURCELL TEAM *Corbis* (*right*)

defects to a band-gap structure is comparable to doping in an electronic semiconductor and opens up a vast range of functions. One example of a dopant is the central hole in photonic crystal optical fibers. Similarly, plugging one of the holes in a thin-film crystal produces a critical element of lasers, namely a small "cavity" that can hold a local electromagnetic mode—imagine a little standing wave of light trapped between mirrors. Recently Axel Scherer's group at the California Institute of Technology used these tiniest of optical cavities to make lasers just 0.03 cubic micron in volume, the smallest ever.

Patterning photonic crystal thin films into optical circuits would represent the ultimate limit of optoelectronic miniaturization. Many researchers believe that integrated circuits that combine conventional electronics and photonics stand ready to extend the integrated-circuit revolution into the domain of high-bandwidth optical signals. This field of band-gap device development will probably draw the most attention in the next few years, but commercial products are still two to three years away.

You might not expect electromagnetic band-gap crystals to be of much use for radio waves, because excessively large crystals would seem to be required. Cellular telephones, for example, may use radio waves that are 35 centimeters long in free space or in air. A crystal with many holes or rods of that size and spacing would hardly be portable. We are rescued by the common LC circuit of electronics, which combines an inductor (a coil; "L") and a capacitor (parallel plates; "C"). Such a circuit can, in effect, cram an electromagnetic wave into a small volume. An array of LC circuits can behave as a photonic crystal and control electromagnetic waves that have free-space wavelengths much larger than the array.

## Backward Light

SHELDON SCHULTZ and David R. Smith, both at the University of California at San Diego, used arrays of LC circuits to create "left-handed" materials, which have a negative refractive index at microwave frequencies. In these materials, electromagnetic waves travel backward: when the wave crests are moving from left to right, the energy of the wave is actually traveling from right to left!

John B. Pendry of Imperial College in England has used LC electromagnetic band-gap arrays for manipulating the radio-frequency magnetic fields used in medical magnetic resonance imaging. Collaborations of researchers from industry, the military and academia (including my group) are studying how LC resonator arrays can also be used for controlling radio waves. Possible advantages of such arrays include making GPS antennas more precise by suppressing signal reflections from Earth and increasing cell-phone handset efficiency by reducing the electromagnetic coupling to the user's head.

It appears likely that these LC circuit concepts can be extended back down to optical wavelengths. These devices would use plasmons, which are currents oscillating at optical frequencies on metallic surfaces. Such tiny LC circuit arrays, smaller than an optical wavelength, may represent the ultimate end point of photonic crystal miniaturization.

Sometimes venturers need to be overconfident, or they would never set off on their quests and persevere to the finish. When I pause to consider the extent of activity in this field today, I am very glad that a decade ago I took those distressing phone calls as an appeal for further research and problem solving. SA

### MORE TO EXPLORE


**Photonic Crystals: Molding the Flow of Light.** John D. Joannopoulos, Robert D. Meade and Joshua N. Winn. Princeton University Press, 1995.

**Optical Properties of Photonic Crystals.** Kazuaki Sakoda. Springer Series in *Optical Sciences*, Vol. 80. Springer Verlag, May 2001.

A thorough photonic and sonic band-gap bibliography is available at **http://home.earthlink.net/~jpdowling/pbgbib.html**

Yurii A. Vlasov's Ultimate Collection of Photonic Band Gap Research Links is at **www.pbglink.com**

Two companies producing photonic crystal fibers are Crystal Fibre A/S (**www.crystal-fibre.com**) and Blaze Photonics (**www.blazephotonics.com**).

# How we came to be

SOME 28,000 YEARS AGO this 60-year-old man was given an elaborate burial, rife with implications of ceremonial practices and of abstract belief. He was interred with rich grave goods and was wearing bracelets, necklaces, pendants, and a tunic on which hundreds of mammoth-ivory beads had been sewn. Along with two juvenile burials from the same site—Sungir in Russia—this is one of the earliest and most resplendent examples of human burials found in Europe.

NOVOSTI *Photo Researchers, Inc.*

# HUMAN

## The acquisition of language and the capacity for symbolic art may lie at the very heart of the extraordinary cognitive abilities that set us apart from the rest of creation

by Ian Tattersall

When we contemplate the extraordinary abilities and accomplishments of *Homo sapiens,* it is certainly hard to avoid a first impression that there must somehow have been an element of inevitability in the process by which we came to be what we are. The product, it's easy to conclude, is so magnificent that it *must* stand as the ultimate expression of a lengthy and gradual process of amelioration and enhancement. How could we have got this way by accident? If we arrived at our exalted state through evolution, then evolution must have worked long and hard at burnishing and improving the breed, must it not? Yet that seems not to be how evolution works; for natural selection is not—it cannot be—in itself a creative process. Natural selection can only work to promote or eliminate novelties that are presented to it by the random genetic changes (influenced, of course, by what was there before) that lie behind all biological innovations. Evolution is best described as opportunistic, simply exploiting or rejecting possibilities as and when they arise, and in turn, the same possibility may be favorable or unfavorable, depending on environmental circumstances (in the broadest definition) at any given moment. There is nothing inherently directional or inevitable about this process, which can smartly reverse itself any time the fickle environment changes.

Indeed, as we'll see a little later, perhaps the most important lesson we can learn from what we know of our own origins involves the significance of what has in recent years increasingly been termed "exaptation." This is a useful name for characteristics that arise in one context before being exploited in another, or for the process by which such novelties are adopted in populations. The classic example of exaptation becoming adaptation is birds' feathers. These structures are essential nowadays to bird flight, but for millions of years before flight came along they were apparently used simply as insulators (and maybe for nothing much at all before that). For a long time, then, feathers were highly useful adaptations for maintaining body temperatures. As adjuncts to flight, on the other hand, they were simply exaptations until, much later, they began to assume an adaptive role in this new function, too. There are many other similar examples, enough that we can't ignore the possibility that maybe our vaunted cognitive capacities originated rather as feathers did: as a very much humbler feature than they became, perhaps only marginally useful, or even as a by-product of something else.

Excerpted from *The Monkey in the Mirror,* by Ian Tattersall, © 2002 by Ian Tattersall, published by Harcourt, Inc.

Let's look at this possibility a little more closely by starting at the beginning. When the first Cro-Magnons arrived in Europe some 40,000 years (kyr) ago, they evidently brought with them more or less the entire panoply of behaviors that distinguishes modern humans from every other species that has ever existed. Sculpture, engraving, painting, body ornamentation, music, notation, subtle understanding of diverse materials, elaborate burial of the dead, painstaking decoration of utilitarian objects—all these and more were an integral part of the day-to-day experience of early *Homo sapiens,* and all are dramatically documented at European sites more than 30 kyr old.

What these behavioral accomplishments most clearly have in common is that all were evidently underwritten by the acquisition of symbolic cognitive processes. There can be little doubt that it was this generalized acquisition, rather than the invention of any one of the specific behaviors I've just listed—or any other—that lay behind the introduction of "modern" behavior patterns into our lineage's repertoire. This new capacity, what's more, stands in the starkest possible contrast to the more modest achievements of the Neanderthals whom the Cro-Magnons so rapidly displaced from their homeland in Europe and western Asia. Indeed, Cro-Magnon behaviors—just like our own—evidently differed totally from those of any other kind of human that had ever previously existed. It is no denigration at all of the Neanderthals and of other now extinct human species—whose attainments were entirely admirable in their own ways—to say that with the arrival on Earth of symbol-centered, behaviorally modern *Homo sapiens,* an entirely new order of being had materialized on the scene. And explaining just how this extraordinary new phenomenon came about is at the same time both the most intriguing question and the most baffling one in all of biology.

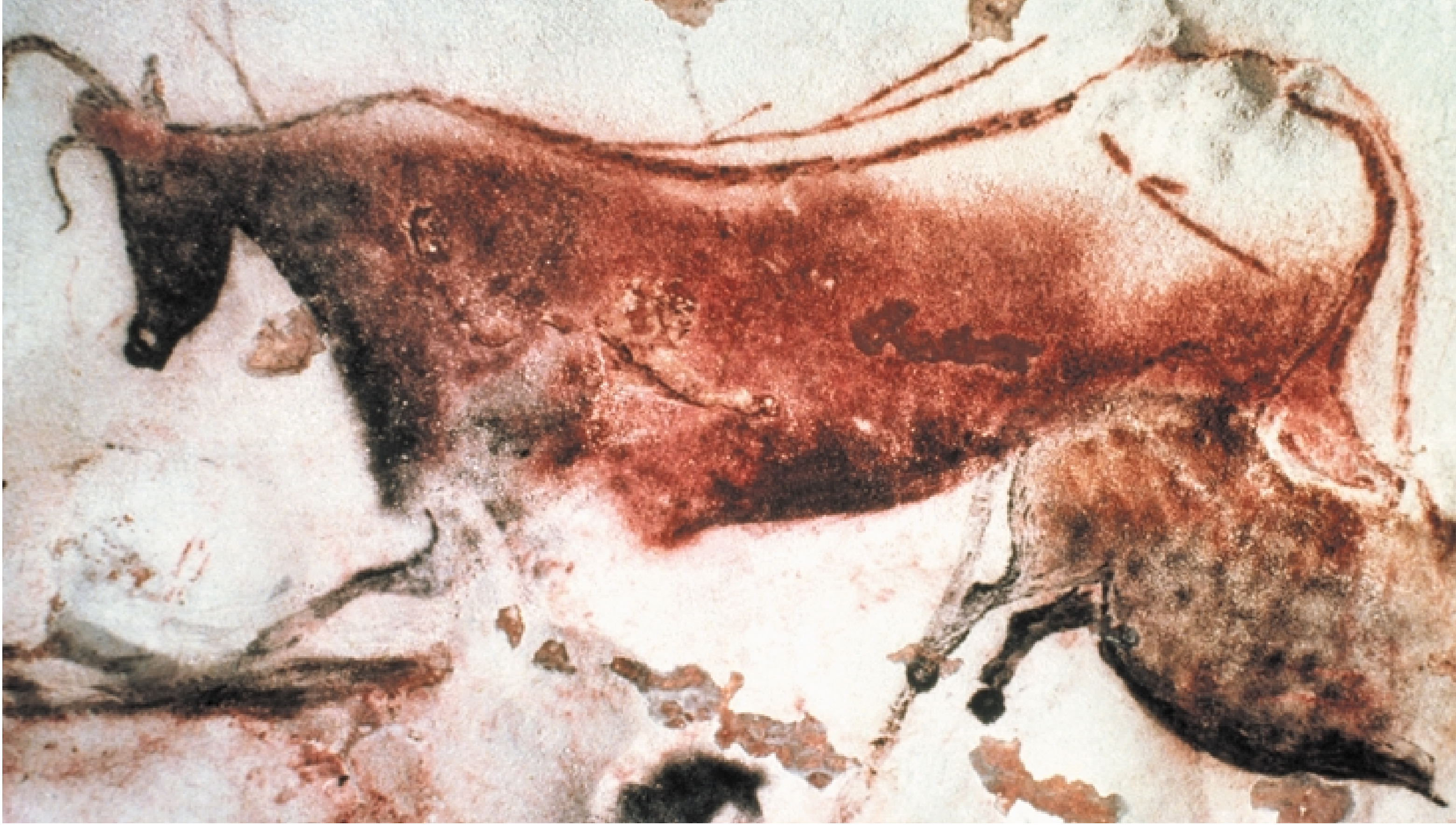
One complicating factor is that there appears to be no correlation whatever between the achievement in the human lineage of behavioral modernity and anatomical modernity. We have evidence of humans who looked exactly like us in the Levant at close to 100 kyr ago. But at the same time, in dramatic contrast to what happened in Europe, the Levantine Neanderthals persisted in the area for some 60 kyr after the anatomical moderns appeared. What's more, throughout this long period of coexistence (whatever form it took, and frankly we have no idea how the different hominids contrived to share the landscape for all those millennia), as far as we can tell from the toolkits they made and the sites they left behind, the two kinds of hominid behaved in more or less identical ways. Suggestively, it was not until right around the time that Cro-Magnon-equivalent stoneworking techniques showed up in the Levant, at about 45 kyr ago, that the Neanderthals finally yielded possession of the area. And it was almost certainly the adoption of symbolic cognitive processes that gave our kind the final—and, for the Neanderthals, fatal—edge. The conclusion thus seems ineluctable that the emergence of anatomically modern *Homo sapiens* considerably predated the arrival of behaviorally modern humans. But while this might sound rather counterintuitive (for wouldn't it be most plausible to "explain" the arrival of a new kind of behavior by that of a new kind of hominid?), it actually makes considerable sense. For where else can any behavioral innovation become established, except within a preexisting species?

Our **VAUNTED COGNITIVE CAPACITIES** may have originated rather as feathers did: as an "exaptation" that arose in one context before being exploited in another.

## The Brain and Innovation

NOBODY WOULD DISPUTE that to understand cognitive processes in any vertebrate species, we have to look to the brain. In the case of our own family, *Homo neanderthalensis* was endowed with a brain as large as our own, albeit housed in a skull of remarkably different shape. And while we know from the very different archaeological records they left behind that Neanderthals and Cro-Magnons behaved in highly distinctive ways, specialists on human brain evolution are hard put to identify any features on the external surface of the brain (as revealed in casts of the interior of the braincase) that would by themselves suggest any major functional difference between Neanderthal and modern *sapiens* brains. The same is obviously true for the brains of those early *sapiens* whose material cultures and ways of life resembled those of the Neanderthals. Clearly, then, we cannot attribute the advent of modern cognitive capacities simply to the culmination of a slow trend in brain improvement over time. Something happened other than a final physical buffing-up of the cognitive mechanism. Of course, by the time modern-looking humans came on the scene the necessary groundwork must have been laid for the adoption of modern cognitive processes, but this is not necessarily the same as saying that a specific neural mechanism had been acquired for them.

Let's look again, for a moment, at what our knowledge of the evolutionary process suggests may have occurred. First, it's important to remember that new structures do not arise *for* anything. They simply come about spontaneously, as byproducts of copying errors that routinely occur as genetic information is passed from one generation to the next. Natural selection is most certainly not a generative force that calls new structures into existence; it can only work on variations that are presented to it, whether to eliminate unfavorable variants or to promote successful ones. We like to speak in



ICE AGE ANIMAL images, such as this aurochs—a form of wild cattle—from the French cave of Lascaux, are frequently accompanied by a wealth of abstract symbols, as we glimpse here in the markings above the neck and back and on the haunches. Lascaux is dated to about 17,000 years ago.

© ARCHIVO ICONOGRAFICO, S.A./CORBIS

terms of "adaptations," since this helps us to make up stories about how and why particular innovations have arisen, or have been successful, in the course of evolution; but in reality, all new genetic variants must come into being as exaptations. The difference is that while adaptations are features that fulfill specific, identifiable functions (which they cannot do, of course, until they are in place), exaptations are simply features that have arisen and are potentially available to be coopted into some new function. This is routine stuff, for many new structures stay around for no better reason than that they just don't get in the way.

This is the general context in which we are obliged to view both the evolution of the human brain as we are familiar with it today and the appearance of modern cognitive function. There was unquestionably an increase in average hominid brain size over the past two million years, although this doesn't tell us much about the actual events of human brain evolution. But the example of the Neanderthals and, even more tellingly, of the anatomical-but-not-behavioral moderns shows us that the arrival of the modern cognitive capacity did not simply involve adding just a bit more neural material, that last little bit of extra brain size that pushed us over the brink. Still less did it involve adding any major new brain structures, for basic brain design remains remarkably uniform among all the higher primates. Instead an exapted brain, equipped since who knows when with a neglected potential for symbolic thought, was somehow put to use.

Unfortunately, exactly what it was that exapted the brain for modern cognitive purposes remains obscure. This is largely because, while we know a lot about brain structure and about which brain components are active during the performance of particular functions, we have no idea at all about how the brain converts a mass of electrical and chemical signals into what we are individually familiar with as consciousness and thought patterns. And it is this which it will be crucial to understand if we are ever to make the leap to comprehending exactly what it is that enables us to be (and I use the term advisedly) human.

Still, it is possible to talk in general terms about the evolution of modern cognition. It has, for example, been argued that at some time between, say, 60 and 50 kyr ago, a speciation event occurred in the human lineage that gave rise to a new, symbolically expressive entity. By implication, this new species would have possessed neural modifications that permitted modern behavior patterns. It would be nice to believe this, because on one level it would certainly simplify the story. The problem is, though, that the time frame doesn't appear to permit it. For this explanation to work, a new human species, physically identical but intellectually superior to one that al-


THE AUTHOR

*IAN TATTERSALL* is a curator in the division of anthropology at the American Museum of Natural History in New York City. This article is excerpted from his latest book, *The Monkey in the Mirror: Essays on the Science of What Makes Us Human* (Harcourt, 2002). Other recent books include *Becoming Human: Evolution and Human Uniqueness* (Harcourt Brace, 1998), *The Last Neanderthal: The Rise, Success and Mysterious Extinction of Our Closest Human Relatives* (Westview, 1999, revised) and *Extinct Humans,* with Jeffrey Schwartz (Westview, 2000).

ready existed, would have had to appear and then to spread throughout the Old World in a remarkably short space of time, totally eliminating its predecessor species in the process. And there is no indication at all, in an admittedly imperfect record, that anything of this kind occurred. Which leaves us with only one evident alternative.

Instead of some anatomical innovation, perhaps we should be seeking some kind of cultural stimulus to our extraordinary cognition. If the modern human brain, with all its potential capacities, had been born along with modern human skull structure at some time around 150 to 100 kyr ago, it could have persisted for a substantial amount of time as exaptation, even as the neural mass continued to perform in the old ways. We have much less evidence than we would like that directly bears on the origin and spread of *Homo sapiens*. However, we do know that our species originated in this general time frame, probably in Africa. And we know as well that it quite rapidly spread Old World–wide from its center of origin, wherever that was.

Further, if at some point, say around 70 to 60 kyr ago, a cultural innovation occurred in one human population or another that activated a potential for symbolic cognitive processes that had resided in the human brain all along, we can readily explain the rapid spread of symbolic behaviors by a simple mechanism of cultural diffusion. It is much more convincing (and certainly more pleasant) to claim that the new form of behavioral expression spread rapidly among populations that already possessed the potential to absorb it, than it is to contemplate the alternative that the worldwide distribution of the unique human capacity came about through a process of wholesale population replacement. What carnage this latter would undoubtedly have involved! On the other hand, cultural interchange among human populations is a phenomenon that is widely documented throughout recorded history, and it must clearly be the preferred explanation for the rapid success of symbolically mediated human behaviors. It remains, though, to suggest what the new cultural stimulus might have been.

## Cognition and Symbolism

WHEN WE SPEAK OF "symbolic processes" in the brain or in the mind, we are referring to our ability to abstract elements of our experience and to represent them with discrete mental symbols. Other species certainly possess consciousness in some sense, but as far as we know, they live in the world simply as it presents itself to them. Presumably, for them the environment seems very much like a continuum, rather than a place, like ours, that is divided into the huge number of separate elements to which we humans give individual names. By separating out its elements in this way, human beings are able constantly to re-create the world, and individual aspects of it, in their minds. And what makes this possible is the ability to form and to manipulate mental symbols that correspond to elements we perceive in the world within and beyond ourselves. Members of other species often display high levels of intuitive reasoning, reacting to stimuli from the environment in quite complex ways, but only human beings are able arbitrarily to combine and recombine mental symbols and to ask themselves questions such as "What if?" And it is the ability to do this, above everything else, that forms the foundation of our vaunted creativity.

Of course, intuitive reasoning still remains a fundamental component of our mental processes; what we have done is to add the capacity for symbolic manipulation to this basic ability. An intuitive appreciation of the relationships among objects and ideas is, for example, almost certainly as large a force in basic scientific creativity as is symbolic representation; but in the end it is the unique combination of the two that makes science—or art, or technology—possible. Certainly, intuitive reasoning can take you a long way just by itself, as I think it's justifiable to claim the example of the Neanderthals shows. The Neanderthals left behind precious few hints of symbolic abilities in the abundant record they bequeathed us of their lives, and it is clear that symbols were not generally an important factor in their existences. Still, their achievements were hardly less remarkable for that, and as far as we can tell, *Homo neanderthalensis* possessed a mastery of the natural world that had been unexceeded in all of earlier human history. Indeed, it seems fair to regard the Neanderthals as exponents of the most complex—and in many ways admirable—lifestyle that it has ever proved possible to achieve with intuitive processes alone.

This inevitably brings up the question about the Neanderthals that everyone wants answered: Could they talk? Many people, especially looking at the spectacularly beautiful stone tools that the Neanderthals made with such skill, find it hard to believe that they couldn't. How, other than through the use of language, could such remarkable skills have been passed down over the generations? Well, not long ago a group of Japanese researchers made a preliminary stab at addressing this problem. They divided a group of undergraduates in two and taught one half how to make a typical Neanderthal stone tool by using elaborate verbal explanations along with practical demonstrations. The other half they taught by silent example alone. One thing this experiment dramatically revealed was

Humans had a vocal tract that could produce the SOUNDS OF ARTICULATE SPEECH over half a million years before we have evidence our forebears used language.

just how tough it is to make stone tools; some of the undergraduates never became proficient. But more remarkable still was that the two groups showed essentially no difference either in the speed at which they acquired toolmaking skills or in the efficiency with which they did so. Apparently learning by silent example is just fine for passing along even sophisticated stone tool–making techniques.

Although this experiment involved modern humans, not Neanderthals, it does show quite forcefully that, once again, we are making a fundamental mistake by assuming that our way is the only way of doing business in the world. None of this is to suggest, of course, that the Neanderthals did not have some form of vocal communication, even quite sophisticated vocal communication. After all, such communication is common among all mammals. And there can be little doubt that Neanderthals spoke, in some general sense. What they almost certainly did not possess, however, is language as we are familiar with it.

## Language and the Emergence of Human Cognition

IF THERE IS ONE single aspect of human mental function that is more closely tied up with symbolic processes than any other, it is surely our use of language. Language is, indeed, the ultimate symbolic mental function, and it is virtually impossible to conceive of thought as we know it in its absence. For words, it is fair to say, function as the units of human thought, at least as we are aware of it. They are certainly the medium by which we explain our thoughts to one another and, as incomparably social creatures, seek to influence what is going on in one another's brains. Thus, if we are seeking a single cultural releasing factor that opened the way to symbolic cognition, the invention of language is the most obvious candidate. Indeed, it is perhaps the only plausible one that it has so far proved possible to identify. What might have happened? Here we have to return to notions of exaptation, for language is a unique aptitude that doesn't seem to have emerged from apelike "protolanguage" and certainly did not do so directly. Still,

CARVED from mammoth ivory over 32,000 years ago, this tiny (less than five centimeters) sculpture is perhaps the earliest work of art known. Its elegant lines express the essence of the horse rather than rendering exactly the stocky proportions of horses of this period. Found at Vogelherd, Germany.

ALEXANDER MARSHACK

it has been argued that since the general ability to acquire language appears to be deeply and universally embedded in the human psyche, this ability must be hardwired into every healthy human brain, where it resides as a result of "normal" Darwinian processes of adaptation by natural selection.

It is certainly true that language is not reinvented in every generation but is rather re-expressed, as every child learns his native tongue(s) as an ordinary, if astonishing, part of the process of growing up. There is, in other words, no denying the existence in the human mind of a "language instinct." What we need to explain, however, is not only how that innate instinct was acquired but also how it made such a rapid and unprecedented appearance.

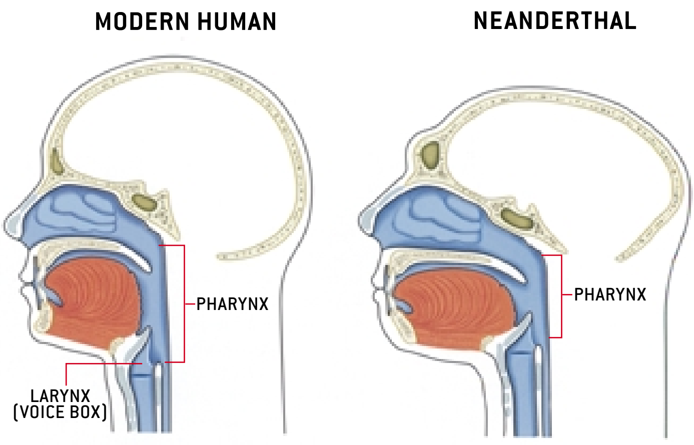
As we've seen, natural selection is not a creative force and can propel nothing into existence by itself. Rather it can only capitalize on what is already there. In a sense, this makes things easier for us since, as far as we can tell, in the emergence of symbolic thought there is no evidence of the kind of slow trend that would be expected under Darwinian selection. What must have happened, instead, is that after a long—and poorly understood—period of erratic brain expansion and reorganization in the human lineage, something occurred that set the stage for language acquisition. This innovation would have depended on the phenomenon of emergence, whereby a chance combination of preexisting elements results in something totally unexpected. The classic example of an emergent quality is water, most of whose remarkable characteristics are entirely unpredicted by those of its constituents, hydrogen and oxygen. Nonetheless, the combination of these ingredients gives rise to something entirely new, and expected only in hindsight. Together with exaptation, emergence provides a powerful mechanism in the evolutionary process, and it truly is a driving force, propelling innovation in new directions.

In the case of linguistic potential, with its innate presence among all humans today, we have to suppose that initially a neural change occurred in some population of the human lineage. This change was presumably rather minor in genetic terms and probably had nothing whatever to do with adaptation in the classical sense. Since during early childhood development the brain rewires itself through the creation of specific pathways from undifferentiated masses of neuronal connections, it is even possible that this event was an epigenetic rather than a genetic one, dependent on developmental stimuli. Whatever the case, it certainly seems to have made no mark on the fossil record, although ultimately its impact on the archaeological traces of the Cro-Magnons and their successors was enormous. Just as the keystone of an arch is a trivial part of the structure, yet is essential to the integrity of the whole, this innovation (whatever it may have been, and we are very far from understanding that) was the final physical element that needed to be in place to make possible language and symbolic thought—and all that has flowed from them, with such fateful consequences for the world. Once it was there, of course, the potential it embodied could lie fallow, simply doing no harm, until released by a cultural stimulus in one particular population. Almost certainly, though it's hard to prove, this stimulus was the invention of language. Everyone today has language, which by itself suggests that it was a highly advantageous acquisition. And if it is as advantageous as we would wish to believe, it is hardly surprising that language, and its associated symbolic behavioral patterns, were subsequently able to spread rapidly among human populations worldwide.

Among the numerous possibilities for how **LANGUAGE MAY HAVE BEEN INVENTED** is that an initial form was created not by adults but **BY CHILDREN**.

So much for the spread of language from its center of origin. Exactly how this fateful novelty may have been invented is a separate question, upon which it is beyond my expertise to speculate. But with the substrate for language in place, the possibilities are numerous. My favorite among them is that an initial form of language may have been invented not by adults but by children. Given the fact that the brain is not a static structure like a rubber ball but is rather a dynamic entity that reorganizes itself during development (and indeed, given the right stimuli, throughout life), it is not implausible that a rudimentary precursor of language as it is familiar today initially arose in a group of children, in the context of play. Such prelanguage might have involved words—sounds—strung together with additive meaning. It is hard to imagine that once this invention had been made, society as a whole would not have eventually adopted it. On a Japanese island, macaque monkeys living along the beach were fed by researchers with sweet potatoes. These delicacies became covered with beach grit, and pretty soon, young macaques started washing them in the sea to remove the sand. It took a while for the adults to catch on: first the females, and only last the dominant males. Doubtless, some of the older and most dominant males never deigned to indulge in this behavior, preferring a familiar life of grit. But a good idea is a good idea—and it is difficult to believe that, in the case of language, once the notion of associating words with objects and ideas had developed, it would not have spread quite rapidly throughout society.

Still, the transition from a nonlinguistic lifestyle to a linguistic one as we are familiar with it involved a huge cognitive and practical leap. It seems probable that the addition of

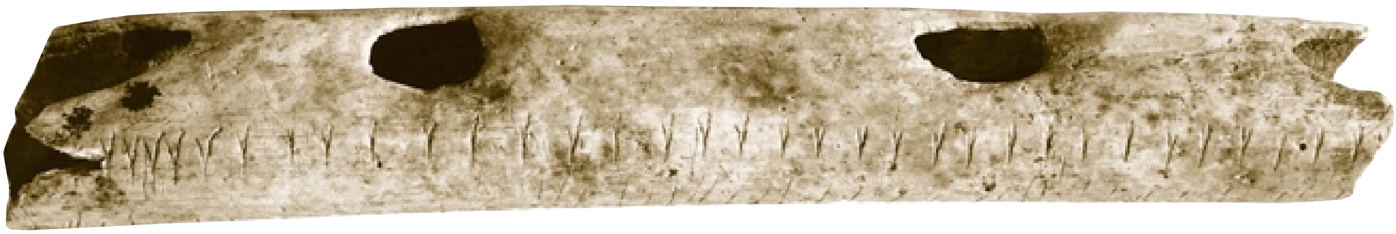



COMPARISON of the head and neck of a modern human and a (reconstructed) Neanderthal shows the differences in the structure of the vocal tract. The much longer pharynx in the modern human is what makes possible the full range of sounds demanded by articulate speech.

ILLUSTRATION BY DIANA SALLES, AFTER SKETCHES BY JEFFREY LAITMAN, © IAN TATTERSALL syntax may have been a separate, and later, event, though perhaps one made inevitable by the arrival of word-object associations. A single-stage progression from inarticulacy to articulate language as we know it seems more than a little implausible, and a multiple-stage process would certainly better mirror the way in which infants acquire language, with the vocabulary beginning to develop (very rapidly) first, and syntax and (later) sentence structuring following after the age of about two years. The history of the emergence of language is undoubtedly complex—indeed, this emergence only seems even possible from our perspective because we *know* it must have occurred. Subsequent to its origin, of course, language quite obviously changed, complexified and diversified hugely, as it became ever more widely adopted among human populations. But its common structure everywhere today, independent of culture, is surely due to the fact that the underlying basis was already there in everyone, long before language itself came along.

But there still remains one other factor to be explained. To speak, you need a brain that will tell your vocal tract what to do, but you also need a vocal tract that will respond appropriately to the brain's instructions. And the primitive primate vocal tract cannot respond in this way. In fact, adult human beings are the only creatures, apes included (though some birds can mimic speech), that can physically make the sounds that are essential to articulate speech. And this ability comes at a price. The principal structures that make up the vocal tract are the larynx, the structure in the neck that houses the vocal cords; the pharynx, a tube that rises above it and opens into the oral and nasal cavities; and the tongue and its associated apparatus. Basic sounds are generated at the vocal cords, and then there is further modulation of those sounds in the pharynx and allied airways above. Among typical mammals, including the apes—and newborn humans—the larynx is positioned high in the neck, and the pharynx is consequently short, limiting what can be done to modulate vocal sounds. In adult humans, in contrast, the larynx lies low in the neck, lengthening the pharynx and increasing the potential for sound modulation. The price I've mentioned is that while the human arrangement makes a vast array of sounds possible, it also prevents simultaneous breathing and swallowing—thereby introducing the unpleasant possibility of choking to death.

This alone suggests that there must be some powerful countervailing advantage in the human conformation of the vocal tract, but the ability to speak, unfortunately, is not it. We know this because the roof of the vocal tract is also the base of the skull. Thus, where this region is preserved in fossils, we can reconstruct in general terms what the vocal tract had looked like in life. The low larynx–high pharynx combination betrays itself in a flexion of the bones of the skull base. We begin to see some evidence of such flexion in *Homo ergaster,* almost 2 myr ago, and a skull of *Homo heidelbergensis* from Ethiopia shows that it had reached virtually its modern degree by about 600 kyr ago. A vocal tract capable of producing the sounds of articulate speech had thus been achieved among humans well over half a million years before we have any independent evidence that our forebears were using language or speaking.



ALEXANDER MARSHACK

MUSICAL INSTRUMENTS, such as this bone flute from a French site, date back at least 32,000 years. They are some of the most striking indicators of a new sensibility in early humans.

Clearly, then, the adult human vocal tract cannot in origin have been an adaptation "for" modern speech—though it might have conferred some advantage in the context of a "prelinguistic" form of vocal communication. So what, then, *is* it "for"? Inevitably we have to come back to exaptation. Despite its disadvantages, basicranial flexion appeared, and it then persisted for a very long time before being capitalized upon for its linguistic qualities. Maybe over that long period it did indeed bestow certain advantages in the production of more archaic forms of speech—forms that we are hardly in a position to characterize. Or maybe it conferred some kind of benefit in terms of respiration, which is an issue that is still very poorly understood among extinct hominids. Still, whatever the case, we have to conclude that the appearance of language and its anatomical correlates was not driven by natural selection, however beneficial these innovations may appear in hindsight to have been.

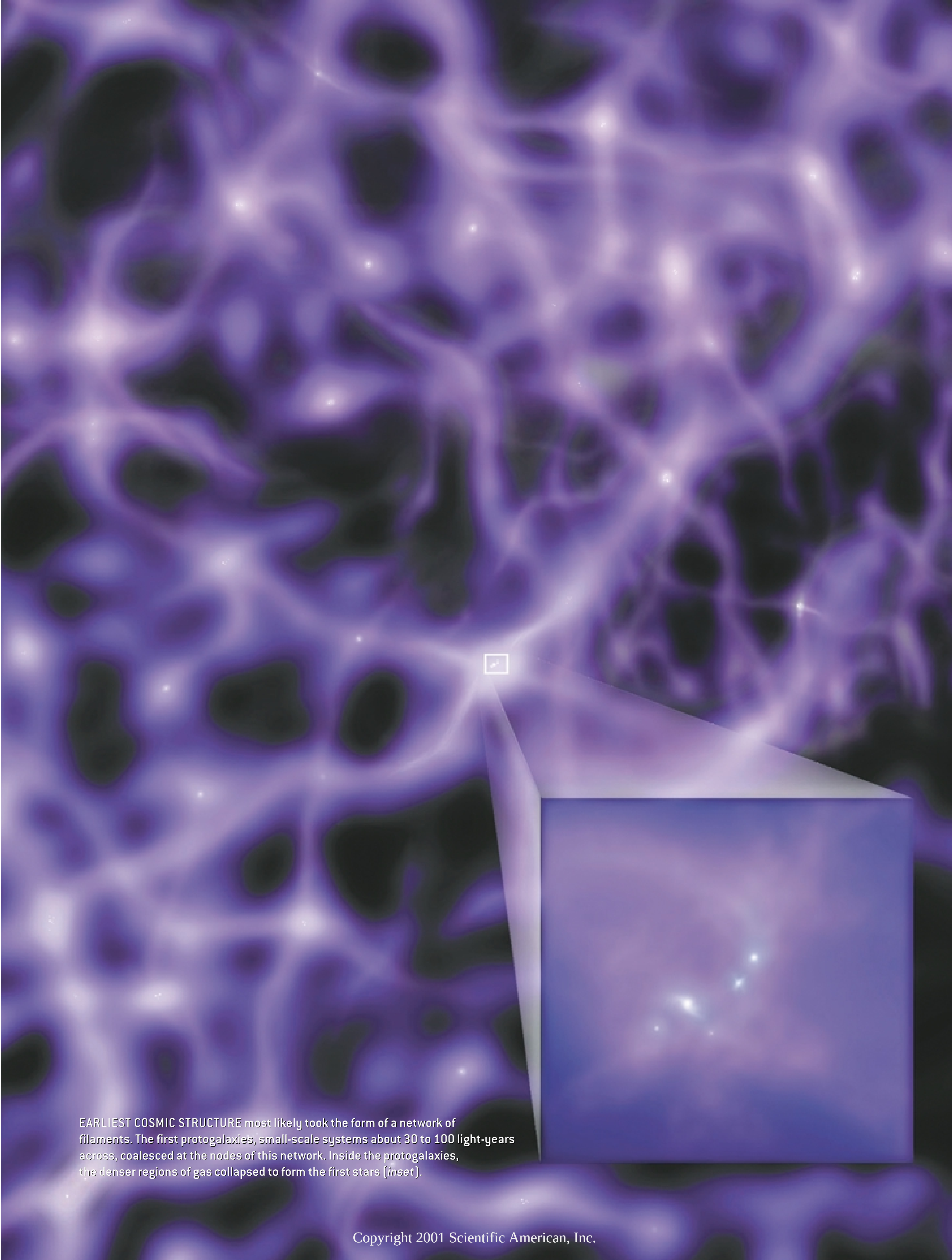
At present, then, there is no way we can come up with any even modestly convincing scenario of what happened in the origination of the extraordinary creature we are, without invoking the humble process of exaptation. Clearly, we are not the result of a constant and careful fine-tuning process over the millennia, and much of our history has been a matter of chance and hazard. Nature never "intended" us to occupy the position of dominance in the living world that, for whatever reasons, we find ourselves in. To a remarkable extent, we are accidental tourists as we cruise through Nature in our bizarre ways. But, of course, we are nonetheless remarkable for that. And still less are we free of responsibility. SA

# THE FIRST STARS IN THE UNIVERSE

**Exceptionally massive and bright, the earliest stars changed the course of cosmic history**

BY RICHARD B. LARSON AND VOLKER BROMM
ILLUSTRATIONS BY DON DIXON

WE LIVE IN A UNIVERSE that is full of bright objects. On a clear night one can see thousands of stars with the naked eye. These stars occupy merely a small nearby part of the Milky Way galaxy; telescopes reveal a much vaster realm that shines with the light from billions of galaxies. According to our current understanding of cosmology, however, the universe was featureless and dark for a long stretch of its early history. The first stars did not appear until perhaps 100 million years after the big bang, and nearly a billion years passed before galaxies proliferated across the cosmos. Astronomers have long wondered: How did this dramatic transition from darkness to light come about?



EARLIEST COSMIC STRUCTURE most likely took the form of a network of filaments. The first protogalaxies, small-scale systems about 30 to 100 light-years across, coalesced at the nodes of this network. Inside the protogalaxies, the denser regions of gas collapsed to form the first stars (*inset*).

After decades of study, researchers have recently made great strides toward answering this question. Using sophisticated computer simulation techniques, cosmologists have devised models that show how the density fluctuations left over from the big bang could have evolved into the first stars. In addition, observations of distant quasars have allowed scientists to probe back in time and catch a glimpse of the final days of the "cosmic dark ages."

The new models indicate that the first stars were most likely quite massive and luminous and that their formation was an epochal event that fundamentally changed the universe and its subsequent evolution. These stars altered the dynamics of the cosmos by heating and ionizing the surrounding gases. The earliest stars also produced and dispersed the first heavy elements, paving the way for the eventual formation of solar systems like our own. And the collapse of some of the first stars may have seeded the growth of supermassive black holes that formed in the hearts of galaxies and became the spectacular power sources of quasars. In short, the earliest stars made possible the emergence of the universe that we see today—everything from galaxies and quasars to planets and people.

## The Dark Ages

THE STUDY of the early universe is hampered by a lack of direct observations. Astronomers have been able to examine much of the universe's history by training their telescopes on distant galaxies and quasars that emitted their light billions of years ago. The age of each object can be determined by the redshift of its light, which shows how much the universe has expanded since the light was produced. The oldest galaxies and quasars that have been observed so far date from about a billion years after the big bang (assuming a present age for the universe of 12 billion to 14 billion years). Researchers will need better telescopes to see more distant objects dating from still earlier times.

Cosmologists, however, can make deductions about the early universe based on the cosmic microwave background radiation, which was emitted about 400,000 years after the big bang. The uniformity of this radiation indicates that matter was distributed very smoothly at that time. Because there were no large luminous objects to disturb the primordial soup, it must have remained smooth and featureless for millions of years afterward. As the cosmos expanded, the background radiation redshifted to longer wavelengths and the universe grew increasingly cold and dark. Astronomers have no observations of this dark era. But by a billion years after the big bang, some bright galaxies and quasars had already appeared, so the first stars must have formed sometime before. When did these first luminous objects arise, and how might they have formed?

Many astrophysicists, including Martin Rees of the University of Cambridge and Abraham Loeb of Harvard University, have made important contributions toward solving these problems. The recent studies begin with the standard cosmological models that describe the evolution of the universe following the big bang. Although the early universe was remarkably smooth, the background radiation shows evidence of small-scale density fluctuations—clumps in the primordial soup. The cosmological models predict that these clumps would gradually evolve into gravitationally bound structures. Smaller systems would form first and then merge into larger agglomerations. The denser regions would take the form of a network of filaments, and the first star-forming systems—small protogalaxies—would coalesce at the nodes of this network. In a similar way, the protogalaxies would then merge to form galaxies, and the galaxies would congregate into galaxy clusters. The process is ongoing: although galaxy formation is now mostly complete, galaxies are still assembling into clusters, which are in turn aggregating into a vast filamentary network that stretches across the universe.

According to the cosmological models, the first small systems capable of forming stars should have appeared between 100 million and 250 million years after the big bang. These protogalaxies would have been 100,000 to one million times more massive than the sun and would have measured about 30 to 100 light-years across. These properties are similar to those of the molecular gas clouds in which stars are currently forming in the Milky Way, but the first protogalaxies would have differed in some fundamental ways. For one, they would have consisted mostly of dark matter, the putative elementary particles that are believed to make up about 90 percent of the universe's mass. In present-day large galaxies, dark matter is segregated from ordinary matter: over time, ordinary matter concentrates in the galaxy's inner region, whereas the dark matter remains scattered throughout an enormous outer halo. But in the protogalaxies, the ordinary matter would still have been mixed with the dark matter.

The second important difference is that the protogalaxies would have contained no significant amounts of any elements besides hydrogen and helium. The big bang produced hydrogen and helium, but most of the heavier elements are created only by the thermonuclear fusion re-

## Overview/*The First Stars*

- Computer simulations show that the first stars should have appeared between 100 million and 250 million years after the big bang. They formed in small protogalaxies that evolved from density fluctuations in the early universe.
- Because the protogalaxies contained virtually no elements besides hydrogen and helium, the physics of star formation favored the creation of bodies that were many times more massive and luminous than the sun.
- Radiation from the earliest stars ionized the surrounding hydrogen gas. Some stars exploded as supernovae, dispersing heavy elements throughout the universe. The most massive stars collapsed into black holes. As protogalaxies merged to form galaxies, the black holes possibly became concentrated in the galactic centers.

actions in stars, so they would not have been present before the first stars had formed. Astronomers use the term "metals" for all these heavier elements. The young metal-rich stars in the Milky Way are called Population I stars, and the old metal-poor stars are called Population II stars; following this terminology, the stars with no metals at all—the very first generation—are sometimes called Population III stars.

In the absence of metals, the physics of the first star-forming systems would have been much simpler than that of present-day molecular gas clouds. Furthermore, the cosmological models can provide, in principle, a complete description of the initial conditions that preceded the first generation of stars. In contrast, the stars that arise from molecular gas clouds are born in complex environments that have been altered by the effects of previous star formation. Therefore, scientists may find it easier to model the formation of the first stars than to model how stars form at present. In any case, the problem is an appealing one for theoretical study, and several research groups have used computer simulations to portray the formation of the earliest stars.

A group consisting of Tom Abel, Greg Bryan and Michael L. Norman (now at Pennsylvania State University, the Massachusetts Institute of Technology and the University of California at San Diego, respectively) has made the most realistic simulations. In collaboration with Paolo Coppi of Yale University, we have done simulations based on simpler assumptions but intended to explore a wider range of possibilities. Toru Tsuribe, now at Osaka University in Japan, has made similar calculations using more powerful computers. Fumitaka Nakamura and Masayuki Umemura (now at Niigata and Tsukuba universities in Japan, respectively) have worked with a more idealized simulation, but it has still yielded instructive results. Although these studies differ in various details, they have all produced similar descriptions of how the earliest stars might have been born.

## Let There Be Light!

THE SIMULATIONS show that the primordial gas clouds would typically form at the nodes of a small-scale filamentary network and then begin to contract because of their gravity. Compression would heat the gas to temperatures above 1,000 kelvins. Some hydrogen atoms would pair up in the dense, hot gas, creating trace amounts of molecular hydrogen. The hydrogen molecules would then start to cool the densest parts of the gas by emitting infrared radiation after they collide with hydrogen atoms. The temperature in the densest parts would drop to about 200 to 300 kelvins, reducing the gas pressure in these regions and hence allowing them to contract into gravitationally bound clumps.

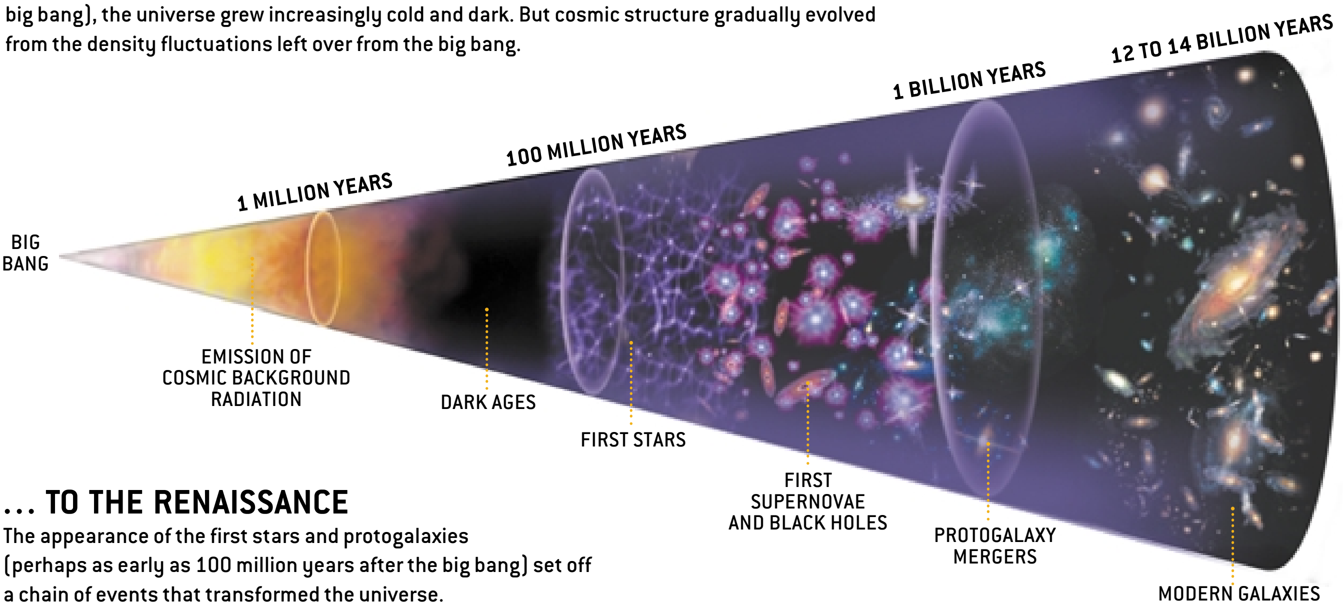
This cooling plays an essential role in allowing the ordinary matter in the primordial system to separate from the dark matter. The cooling hydrogen settles into a flattened rotating configuration that is clumpy and filamentary and possibly shaped like a disk. But because the dark-matter particles would not emit radiation or lose energy, they would remain scattered in the primordial cloud. Thus, the star-forming system would come to resemble a miniature galaxy, with a disk of ordinary matter and a halo of dark matter. Inside the disk, the densest clumps of gas would continue to contract, and eventually some of them would undergo a runaway collapse and become stars.

The first star-forming clumps were

## COSMIC TIMELINE

### FROM THE DARK AGES ...

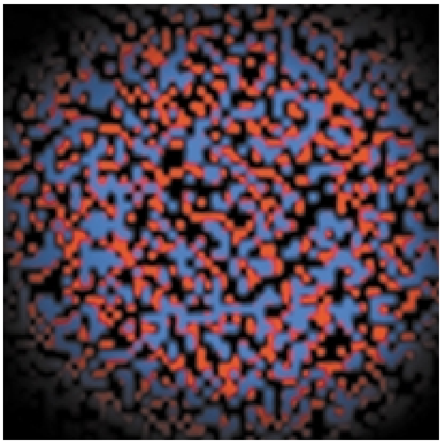
After the emission of the cosmic microwave background radiation (about 400,000 years after the big bang), the universe grew increasingly cold and dark. But cosmic structure gradually evolved from the density fluctuations left over from the big bang.




### ... TO THE RENAISSANCE

The appearance of the first stars and protogalaxies (perhaps as early as 100 million years after the big bang) set off a chain of events that transformed the universe.

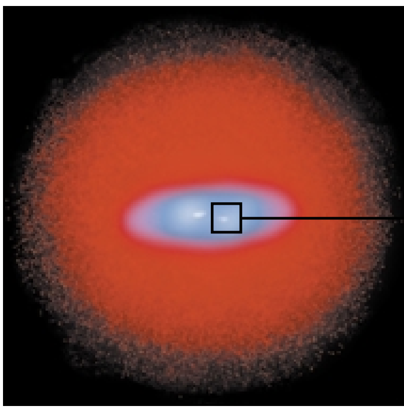
# THE BIRTH AND DEATH OF THE FIRST STARS



## PRIMEVAL TURMOIL

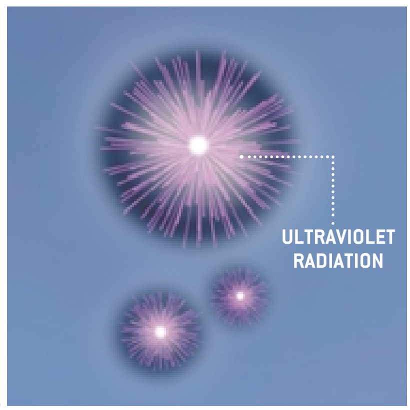
The process that led to the creation of the first stars was very different from present-day star formation. But the violent deaths of some of these stars paved the way for the emergence of the universe that we see today.

**1** The first star-forming systems—small protogalaxies—consisted mostly of the elementary particles known as dark matter (*shown in red*). Ordinary matter—mainly hydrogen gas (*blue*)—was initially mixed with the dark matter.

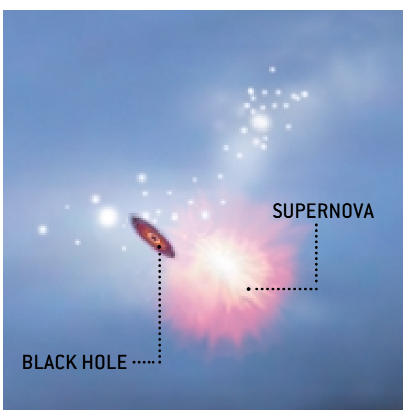


**2** The cooling of the hydrogen allowed the ordinary matter to contract, whereas the dark matter remained dispersed. The hydrogen settled into a disk at the center of the protogalaxy.

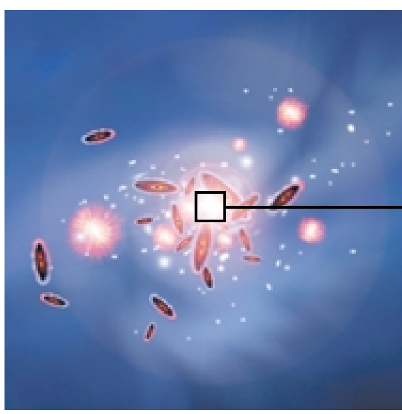
**3** The denser regions of gas contracted into star-forming clumps, each hundreds of times as massive as the sun. Some of the clumps of gas collapsed to form very massive, luminous stars.




**4** Ultraviolet radiation from the stars ionized the surrounding neutral hydrogen gas. As more and more stars formed, the bubbles of ionized gas merged and the intergalactic gas became ionized.




**5** A few million years later, at the end of their brief lives, some of the first stars exploded as supernovae. The most massive stars collapsed into black holes.



**6** Gravitational attraction pulled the protogalaxies toward one another. The collisions most likely triggered star formation, just as galactic mergers do now.

**7** Black holes possibly merged to form a supermassive hole at the protogalaxy's center. Gas swirling into this hole might have generated quasarlike radiation.

much warmer than the molecular gas clouds in which most stars currently form. Dust grains and molecules containing heavy elements cool the present-day clouds much more efficiently to temperatures of only about 10 kelvins. The minimum mass that a clump of gas must have to collapse under its gravity is called the Jeans mass, which is proportional to the square of the gas temperature and inversely proportional to the square root of the gas pressure. The first star-forming systems would have had pressures similar to those of present-day molecular clouds. But because the temperatures of the first collapsing gas clumps were almost 30 times higher than those of molecular clouds, their Jeans mass would have been almost 1,000 times larger.

In molecular clouds in the nearby part of the Milky Way, the Jeans mass is roughly equal to the mass of the sun, and the masses of the prestellar clumps observed in these clouds are about the same. If we scale up by a factor of almost 1,000, we can estimate that the masses of the first star-forming clumps would have been about 500 to 1,000 solar masses. In agreement with this prediction, all the computer simulations mentioned above showed the formation of clumps with masses of several hundred solar masses or more.

Our group's calculations suggest that the predicted masses of the first star-forming clumps are not very sensitive to the assumed cosmological conditions (for example, the exact nature of the initial density fluctuations). In fact, the predicted masses depend primarily on the physics of the hydrogen molecule and only secondarily on the cosmological model or simulation technique. One reason is that molecular hydrogen cannot cool the gas below 200 kelvins, making this a lower limit to the temperature of the first star-forming clumps. Another is that the cooling from molecular hydrogen becomes inefficient at the higher densities encountered when the clumps begin to collapse. At these densities the hydrogen molecules collide with other atoms before they have time to emit an infrared photon; this raises the gas temperature and slows down the contraction until the clumps have built up to at least a few hundred solar masses.

What was the fate of the first collapsing clumps? Did they form stars with similarly large masses, or did they fragment into many smaller parts and form many smaller stars? The research groups have pushed their calculations to the point at which the clumps are well on their way to forming stars, and none of the simulations has yet revealed any tendency for the clumps to fragment. This agrees with our understanding of present-day star formation; observations and simulations show that the fragmentation of star-forming clumps is typically limited to the formation of binary systems (two stars orbiting around each other). Fragmentation seems even less likely to occur in the primordial clumps, because the inefficiency of molecular hydrogen cooling would keep the Jeans mass high. The simulations, however, have not yet determined the final outcome of collapse with certainty, and the formation of binary systems cannot be ruled out.

Different groups have arrived at somewhat different estimates of just how massive the first stars might have been. Abel, Bryan and Norman have argued that the stars probably had masses no greater than 300 solar masses. Our own work suggests that masses as high as 1,000 solar masses might have been possible. Both predictions might be valid in different circumstances: the very first stars to form might have had masses no larger than 300 solar masses, whereas stars that formed a little later from the collapse of larger protogalaxies might have reached the higher estimate. Quantitative predictions are difficult because of feedback effects; as a massive star forms, it produces intense radiation and matter outflows that may blow away some of the gas in the collapsing clump. But these effects depend strongly on the presence of heavy elements in the gas, and therefore they should be less important for the earliest stars. Thus, it seems safe to conclude that the first stars in the universe were typically many times more massive and luminous than the sun.

It seems safe to conclude that the **first stars** in the universe were typically many times more **massive** and **luminous** than the sun.

## The Cosmic Renaissance

WHAT EFFECTS did these first stars have on the rest of the universe? An important property of stars with no metals is that they have higher surface temperatures than stars with compositions like that of the sun. The production of nuclear energy at the center of a star is less efficient without metals, and the star would have to be hotter and more compact to produce enough energy to counteract gravity. Because of the more compact structure, the surface layers of the star would also be hotter. In collaboration with Rolf-Peter Kudritzki of the University of Hawaii and Abraham Loeb of Harvard, one of us (Bromm) devised theoretical models of such stars with mass-

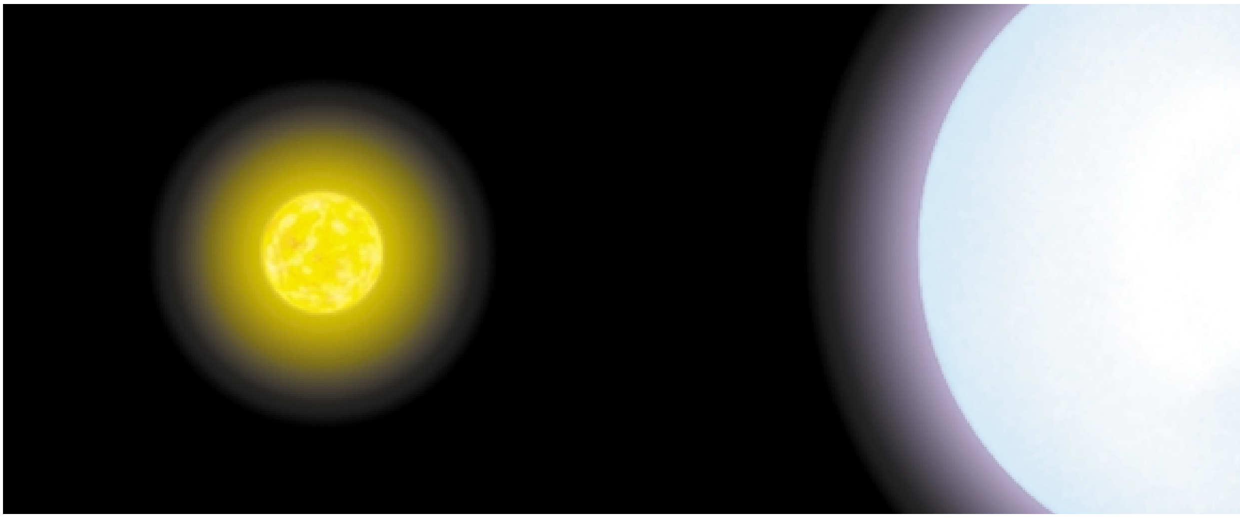
**THE AUTHORS**

*RICHARD B. LARSON* and *VOLKER BROMM* have worked together to understand the processes that ended the "cosmic dark ages" and brought about the birth of the first stars. Larson, a professor of astronomy at Yale University, joined the faculty there in 1968 after receiving his Ph.D. from the California Institute of Technology. His research interests include the theory of star formation as well as the evolution of galaxies. Bromm earned his Ph.D. at Yale in 2000 and is now a postdoctoral researcher at the Harvard-Smithsonian Center for Astrophysics, where he focuses on the emergence of cosmic structure. The authors acknowledge the many contributions of Paolo Coppi, associate professor of astronomy at Yale, to their joint work on the formation of the first stars.

## STAR STATS

### COMPARING CHARACTERISTICS

Computer simulations have given scientists some indication of the possible masses, sizes and other characteristics of the earliest stars. The lists below compare the best estimates for the first stars with those for the sun.



**SUN**

MASS: $1.989 \times 10^{30}$ kilograms
RADIUS: 696,000 kilometers
LUMINOSITY: $3.85 \times 10^{23}$ kilowatts
SURFACE TEMPERATURE: 5,780 kelvins
LIFETIME: 10 billion years

**FIRST STARS**

MASS: 100 to 1,000 solar masses
RADIUS: 4 to 14 solar radii
LUMINOSITY: 1 million to 30 million solar units
SURFACE TEMPERATURE: 100,000 to 110,000 kelvins
LIFETIME: 3 million years

es between 100 and 1,000 solar masses. The models showed that the stars had surface temperatures of about 100,000 kelvins—about 17 times higher than the sun's surface temperature. Therefore, the first starlight in the universe would have been mainly ultraviolet radiation from very hot stars, and it would have begun to heat and ionize the neutral hydrogen and helium gas around these stars soon after they formed.

We call this event the cosmic renaissance. Although astronomers cannot yet estimate how much of the gas in the universe condensed into the first stars, even a fraction as small as one part in 100,000 could have been enough for these stars to ionize much of the remaining gas. Once the first stars started shining, a growing bubble of ionized gas would have formed around each one. As more and more stars formed over hundreds of millions of years, the bubbles of ionized gas would have eventually merged, and the intergalactic gas would have become completely ionized.

Scientists from the California Institute of Technology and the Sloan Digital Sky Survey have recently found evidence for the final stages of this ionization process. The researchers observed strong absorption of ultraviolet light in the spectra of quasars that date from about 900 million years after the big bang. The results suggest that the last patches of neutral hydrogen gas were being ionized at that time. Helium requires more energy to ionize than hydrogen does, but if the first stars were as massive as predicted, they would have ionized helium at the same time. On the other hand, if the first stars were not quite so massive, the helium must have been ionized later by energetic radiation from sources such as quasars. Future observations of distant objects may help determine when the universe's helium was ionized.

If the first stars were indeed very massive, they would also have had relatively short lifetimes—only a few million years. Some of the stars would have exploded as supernovae at the end of their lives, expelling the metals they produced by fusion reactions. Stars that are between 100 and 250 times as massive as the sun are predicted to blow up completely in energetic explosions, and some of the first stars most likely had masses in this range. Because metals are much more effective than hydrogen in cooling star-forming clouds and allowing them to collapse into stars, the production and dispersal of even a small amount could have had a major effect on star formation.

Working in collaboration with Andrea Ferrara of the University of Florence in Italy, we have found that when the abundance of metals in star-forming clouds rises above one thousandth of the metal abundance in the sun, the metals rapidly cool the gas to the temperature of the cosmic background radiation. (This temperature declines as the universe expands, falling to 19 kelvins a billion years after the big bang and to 2.7 kelvins today.) This efficient cooling allows the formation of stars with smaller masses and may also considerably boost the overall rate at which stars are born. In fact, it is possible that the pace of star formation did not accelerate until after the first metals had been produced. In this case, the second-generation stars might have been the ones primarily responsible for lighting up the universe and bringing about the cosmic renaissance.

At the start of this active period of star birth, the cosmic background temperature would have been higher than the temperature in present-day molecular clouds (10 kelvins). Until the temperature dropped to that level—which happened about two billion years after the big bang—the process of star formation may still have favored massive stars. As a result, large numbers of such stars may have formed during the early stages of galaxy building by successive mergers of protogalaxies. A similar phenomenon may occur in the modern universe when two galaxies collide and trigger a starburst—a sudden increase in the rate of star formation. Such events are now fairly rare, but some evidence suggests that they may produce relatively large numbers of massive stars.

## Puzzling Evidence

THIS HYPOTHESIS about early star formation might help explain some puzzling features of the present universe. One unsolved problem is that galaxies contain fewer metal-poor stars than would be expected if metals were produced at a rate proportional to the star formation rate. This discrepancy might be resolved if early star formation had produced relatively more massive stars; on dying, these stars would have dispersed large amounts of metals, which would have then been incorporated into most of the low-mass stars that we now see.

Another puzzling feature is the high metal abundance of the hot x-ray-emitting intergalactic gas in clusters of galaxies. This observation could be accounted for most easily if there had been an early period of rapid formation of massive stars and a correspondingly high supernova rate that chemically enriched the intergalactic gas. The case for a high supernova rate at early times also dovetails with the recent evidence suggesting that most of the ordinary matter and metals in the universe lies in the diffuse intergalactic medium rather than in galaxies. To produce such a distribution of matter, galaxy formation must have been a spectacular process, involving intense bursts of massive star formation and barrages of supernovae that expelled most of the gas and metals out of the galaxies.

Stars that are more than 250 times more massive than the sun do not explode at the end of their lives; instead they collapse into similarly massive black holes. Several of the computer simulations mentioned above predict that some of the first stars would have had masses this great. Because the first stars formed in the densest parts of the universe, any black holes resulting from their collapse would have become incorporated, via successive mergers, into systems of larger and larger size. It is possible that some of these black holes became concentrated in the inner part of large galaxies and seeded the growth of the supermassive black holes—millions of times more massive than the sun—that are now found in galactic nuclei.

Furthermore, astronomers believe that the energy source for quasars is the gas whirling into the black holes at the centers of large galaxies. If smaller black holes had formed at the centers of some of the first protogalaxies, the accretion of matter into the holes might have generated "mini quasars." Because these objects could have appeared soon after the first stars, they might have provided an additional source of light and ionizing radiation at early times.

Thus, a coherent picture of the universe's early history is emerging, although certain parts remain speculative. The formation of the first stars and protogalaxies began a process of cosmic evolution. Much evidence suggests that the period of most intense star formation, galaxy building and quasar activity occurred a few billion years after the big bang and that all these phenomena have continued at declining rates as the universe has aged. Most of the cosmic structure building has now shifted to larger scales as galaxies assemble into clusters.

In the coming years, researchers hope to learn more about the early stages of the story, when structures started developing on the smallest scales. Because the first stars were most likely very massive and bright, instruments such as the Next Generation Space Telescope—the planned successor to the Hubble Space Telescope—might detect some of these ancient bodies. Then astronomers may be able to observe directly how a dark, featureless universe formed the brilliant panoply of objects that now give us light and life. SA

**The formation of the first stars and protogalaxies began a process of cosmic evolution.**

## MORE TO EXPLORE


**Before the Beginning: Our Universe and Others.** Martin J. Rees. Perseus Books, 1998.

**The Formation of the First Stars.** Richard B. Larson in *Star Formation from the Small to the Large Scale.* Edited by F. Favata, A. A. Kaas and A. Wilson. ESA Publications, 2000. Available on the Web at **www.astro.yale.edu/larson/papers/Noordwijk99.pdf**

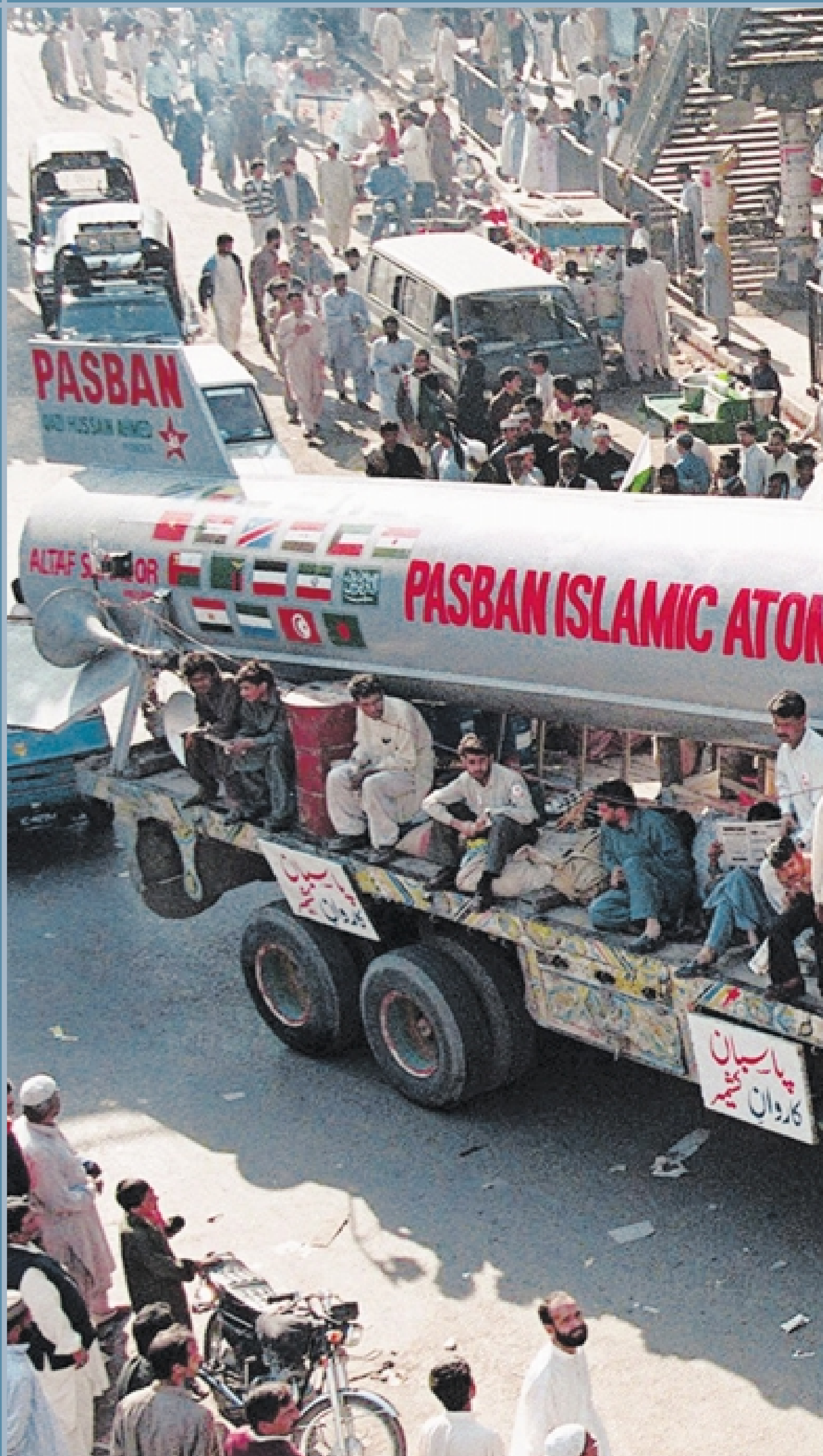
**In the Beginning: The First Sources of Light and the Reionization of the Universe.** R. Barkana and A. Loeb in *Physics Reports,* Vol. 349, No.2, pages 125–238; July 2001. Available on the Web at **aps.arxiv.org/abs/astro-ph/0010468**

Graphics from computer simulations of the formation of the first stars can be found at **www.tomabel.com**

# India, Pak

The Indian subcontinent is the most likely place in the world for a nuclear war

by M. V. Ramana and A. H. Nayyar




AAMIR QURESHI *AFP Photo/Corbis*

THROUGH THE STREETS OF KARACHI, a mock missile is paraded by Pasban, a youth wing of Pakistan's main fundamentalist party, Jamaat-e-Islami. The parade took place in February 1999 on a day of solidarity with Kashmiris in India-administered Kashmir. Such enthusiasm for nuclear weapons is widespread, though not universal, in both India and Pakistan.

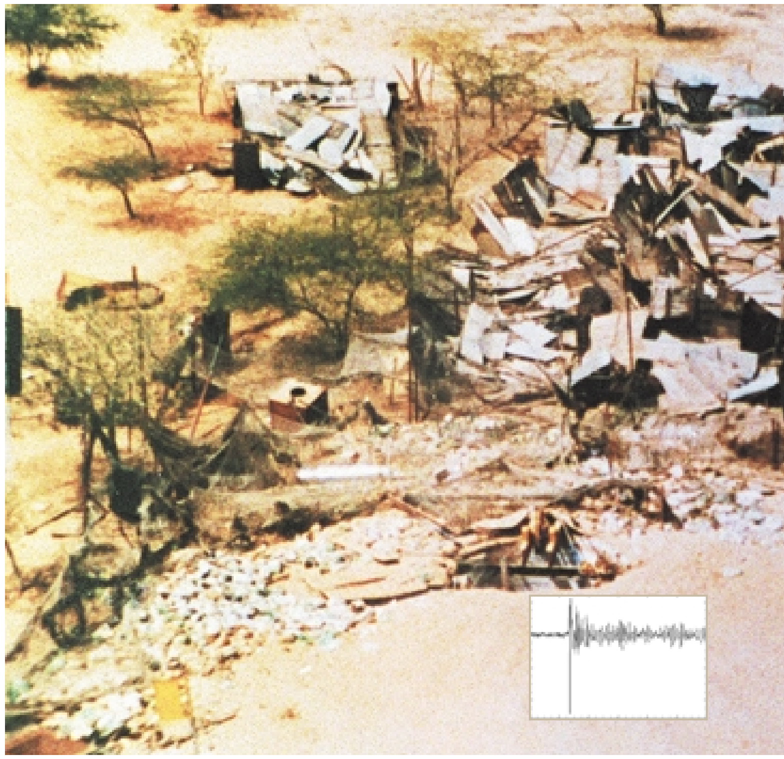
# kistan and the Bomb

**As the U.S. mobilized its armed forces in the aftermath of the terrorist attacks of September 11, the world's attention focused on Pakistan, so crucial to military operations in Afghanistan. When Pakistani president Pervez Musharraf pledged total support for a U.S.-led multinational force on September 14, many people's first thought was: What about Pakistan's nuclear weapons?**

Could they fall into the hands of extremists? In an address to his nation, Musharraf proclaimed that the "safety of nuclear missiles" was one of his priorities. The Bush administration began to consider providing Pakistan with perimeter security and other assistance to guard its nuclear facilities.

The renewed concern about nuclear weapons in South Asia comes a little more than three years after the events of May 1998: the five nuclear tests conducted by India at Pokharan in the northwestern desert state of Rajasthan, followed three weeks later by six nuclear explosions conducted by Pakistan in its southwestern region of Chaghai. These tit-for-tat responses mirrored the nuclear buildup by the U.S. and the former Soviet Union, with a crucial difference: the two cold war superpowers were separated by an ocean and never fought each other openly. Neighboring India and Pakistan have gone to war three times since British India was partitioned in 1947 into Muslim-majority and Hindu-majority states. Even now artillery guns regularly fire over the border (officially, a cease-fire line) in the disputed region of Kashmir.

In May 1999, just one year after the nuclear tests, bitter fighting broke out over the occupation of a mountain ledge near the Kashmiri town of Kargil. The two-month conflict took a toll of between 1,300 (according to the Indian government) and 1,750 (according to Pakistan) lives. For the first time since 1971, India deployed its air force to launch attacks. In response, Pakistani fighter planes were scrambled for fear they might be hit on the ground; air-raid sirens sounded in the capital city of Islamabad. High-level officials in both countries issued at least a dozen nuclear threats. The peace and stability that some historians and political scientists have ascribed to nuclear weapons—because nuclear nations are supposed to be afraid of mutually assured destruction—were nowhere in sight.



UNDERGROUND NUCLEAR EXPLOSIONS conducted by India on May 11, 1998, caused the surface immediately above to collapse. Seismic readings (*inset*) suggest that the total explosive yield was between 16 and 30 kilotons, about half of what India claimed.

AP PHOTO; TERRY C. WALLACE *University of Arizona (inset)*

### SCIENTIFIC AMERICAN *ARTICLES ON NUCLEAR PROLIFERATION*

*Since the dawn of the nuclear age,* SCIENTIFIC AMERICAN *has published articles on nuclear weapons policy. A sampling:*

**The Hydrogen Bomb: II.** Hans A. Bethe. April 1950.

**The Proliferation of Nuclear Weapons.** William Epstein. April 1975.

**Stopping the Production of Fissile Materials for Weapons.** Frank von Hippel, David H. Albright and Barbara G. Levi. September 1985.

**The Future of American Defense.** Philip Morrison, Kosta Tsipis and Jerome Wiesner. February 1994.

**The Real Threat of Nuclear Smuggling.** Phil Williams and Paul N. Woessner. January 1996.

**Iran's Nuclear Puzzle.** David A. Schwarzbach. June 1997.

Wiser counsel eventually prevailed. The end of the Kargil clash, however, was not the end of the nuclear confrontation in South Asia. The planned deployment of nuclear weapons by the two countries heightens the risks. With political instability a real possibility in Pakistan, particularly given the conflict in Afghanistan, the dangers have never been so near.

## Learning to Love the Bomb

BOTH COUNTRIES have been advancing their nuclear programs almost ever since they gained independence from Britain. Understanding this history is crucial in figuring out what to do now, as well as preventing the further proliferation of nuclear weapons. Although the standoff between Pakistan and India has distinct local characteristics, both countries owe much to other nuclear states. The materials used in their bombs were manufactured with Western technology;

both countries' justifications for joining the nuclear club drew heavily on cold war thinking. The continued reliance of the U.S. and Russia on thousands of nuclear weapons on hair-trigger alert only adds to the perceived need for nuclear arsenals in India and Pakistan.

While setting up the Indian Atomic Energy Commission (IAEC) in 1948, Jawaharlal Nehru, India's first prime minister, laid out his desire that the country "develop [atomic energy] for peaceful purposes." But at the same time, he recognized that "if we are compelled as a nation to use it for other purposes, possibly no pious sentiments will stop the nation from using it that way." Such ambivalence remained a central feature of India's nuclear policy as it developed.

To Indian leaders, the program symbolized international political clout and technological modernity. Over the next two decades, India began to construct and operate nuclear reactors, mine uranium, fabricate fuel and extract plutonium. In terms of electricity produced, these activities often proved uneconomical—hardly, one might think, where a developing nation should be putting its resources. Politicians and scientists justified the nuclear program on the grounds that it promoted self-sufficiency, a popular theme in postcolonial India. Rhetoric aside, India solicited and received ample aid from Canada, the U.S. and other countries.

FATHERS OF THE ATOMIC BOMBS: **A. Q. Khan (*left*) set up the Kahuta centrifuge plant, which produces the uranium used in Pakistan's bombs. Homi Jehangir Bhabha (*right*), a theoretical physicist educated at the University of Cambridge, laid the groundwork for India's nuclear capability.**

B. K. BANGASH *AP Photo* (*Khan*); COURTESY OF BHABHA ATOMIC RESEARCH CENTER (*Bhabha*)

After India's defeat in the 1962 border war with China, some right-wing politicians issued the first public calls for developing a nuclear arsenal. These appeals became louder after China's first nuclear test in 1964. Countering this bomb lobby were other prominent figures, who argued that the economic cost would be too high. Many leading scientists advocated the bomb. Homi Bhabha, the theoretical physicist who ran the IAEC, claimed that his organization could build nuclear weapons "within 18 months." Citing a Lawrence Livermore National Laboratory report, Bhabha predicted that nuclear bombs would be cheap. He also promised economic gain from "peaceful nuclear explosions," which many American nuclear researchers extolled for, say, digging canals.

In November 1964 Indian prime minister Lal Bahadur Shastri compromised, permitting the commission to explore the technology for such an explosion. It turned out that Bhabha had already been doing some exploring. In 1960 he reportedly sent Vasudev Iya, a young chemist, to France to absorb as much information as he possibly could about how polonium—a chemical element used to trigger a nuclear explosion—was prepared. Bhabha died in 1966, and design work on the "peaceful" device did not begin for another two years. But by the late 1960s, between 50 and 75 scientists and engineers were actively developing weapons. Their work culminated in India's first atomic test—the detonation on May 11, 1974, of a plutonium weapon with an explosive yield of five to 12 kilotons. For comparison, the bomb dropped on Hiroshima had a yield of about 13 kilotons.

## Nuclear Tipping Point

THE 1974 TEST was greeted with enthusiasm within India and dismay elsewhere. Western countries cut off cooperative efforts on nuclear matters and formed the Nuclear Suppliers Group, which restricts the export of nuclear technologies and materials to nations that refuse to sign the 1968 Nuclear Non-Proliferation Treaty, including both India and Pakistan.

In the years that followed, the bomb lobby pushed for tests of more advanced weapons, such as a boosted-fission design and a hydrogen bomb. It appears that in late 1982 or early 1983, Prime Minister Indira Gandhi tentatively agreed to another test, only to change her mind within 24 hours. One of the causes for the volte-face is said to have been a conversation with the Indian foreign secretary, whom an American official had confronted with satellite evidence of preparations at the test site. The conversation seems to have convinced Gandhi that the U.S. reaction would create economic difficulties for India. Instead, it is reported, she wanted to "develop other things and keep them ready."

The "other things" she had in mind were ballistic missiles. In 1983 the Integrated Guided Missile Development Program was set up under the leadership of Abdul Kalam, a renowned rocket engineer. This followed an earlier, secret attempt to reverse-engineer a Soviet antiaircraft missile that India had purchased in the 1960s. Although that effort did not succeed, it led to the development of several critical technologies, in particular a rocket engine. Kalam adopted an open management style—as compared with the closed military research program—and involved academic institutions and private firms. Anticipating restrictions on imports, India went on a shopping spree for gyroscopes, accelerometers and motion simulators from suppliers in France, Sweden, the U.S. and Germany.

In 1988 India tested its first short-range surface-to-surface missile. A year later came a medium-range missile; in April 1999, a longer-range missile. The latter can fly 2,000 kilometers, well into the heart of China. Despite this ability, India is unlikely to achieve nuclear parity with China. According to various estimates, China has 400 warheads and an additional 200 to 575 weapons' worth of fissile material. If India's plutonium production reactors have been operating on average at 50 to 80 percent of full power, India has somewhere between 55 and 110 weapons' worth of plutonium [*see illustration on opposite page*]. The stockpile could be much larger if commercial reactors earmarked for electricity generation have also been producing plutonium for weapons.

## Eating Grass

PAKISTAN'S NUCLEAR PROGRAM drew on a general desire to match India in whatever it does. The country set up its Atomic Energy Commission in 1954, began operating its first nuclear research reactor in 1965 and opened its first commercial reactor in 1970. As scientific adviser to the government, physicist Abdus Salam, who later won the Nobel Prize in Physics, played an important role.

The program was severely handicapped by a shortage of manpower. In 1958 the commission had only 31 scientists and engineers; it was run by Nazir Ahmad, the former head of the Textile Committee. The commission pursued an active program of training personnel by sending more than 600 scientists and engineers to the U.S., Canada and western Europe. With generous help from these countries, some of which also aided India, Pakistan had a few nuclear research laboratories in place by the mid-1960s.

After the 1965 war with India, many Pakistani politicians, journalists and scientists pressed for the development of nuclear weapons. The most prominent was Foreign Minister Zulfikhar Ali Bhutto, who famously declared that if India developed an atomic bomb, Pakistan would follow "even if we have to eat grass or leaves or to remain hungry." After Pakistan's defeat in the December 1971 war, Bhutto became prime minister. In January 1972 he convened a meeting of Pakistani scientists to discuss making bombs.

As the first prong of their two-pronged effort to obtain weapons material, researchers attempted to purchase plutonium reprocessing plants from France and Belgium. After initially agreeing to the sale, France backed down under American pressure. But a few Pakistani scientists did go to Belgium for training in reprocessing technology. Returning to Pakistan, they constructed a small-scale reprocessing laboratory

THE AUTHORS

*M. V. RAMANA* and *A. H. NAYYAR* are physicists and peace activists who have worked to bridge the divide between India and Pakistan. Ramana, a research staff member in Princeton University's Program on Science and Global Security (www.princeton.edu/~globsec), is a founding member of the Indian Coalition for Nuclear Disarmament and Peace. He was born and raised in southern India and has written extensively on the region's classical music. Nayyar, a physics professor at Quaid-e-Azam University in Islamabad, is co-founder of the Pakistan Peace Coalition. He also runs a project to provide education to underprivileged children.

KHUSHAB NUCLEAR REACTOR in Pakistan produces a few bombs' worth of plutonium every year. Based on the size of the cooling towers visible in this Ikonos commercial satellite image, nuclear analysts estimate that the reactor generates about 50 megawatts of heat.

in the early 1980s. Using spent fuel from a plutonium production reactor that opened in 1998, this lab is capable of producing two to four bombs' worth of plutonium annually.

As the second prong, researchers explored techniques for enriching uranium—that is, for concentrating the bomb-usable isotope uranium 235. In 1975 A. Q. Khan, a Pakistani metallurgist who had worked at an enrichment plant in the Netherlands, joined the group. With him came classified design information and lists of component suppliers in the West, many of which proved quite willing to violate export-control laws [*see box on page 82*]. Success came in 1979 with the enrichment of small quantities of uranium. Since then, Pakistan is estimated to have produced 20 to 40 bombs' worth of enriched uranium. Every year it produces another four to six bombs' worth [*see illustration on opposite page*].

By 1984 designs for aircraft-borne bombs were reportedly complete. Around this time, some American officials started alleging that China had given Pakistan the design for a missile-ready bomb. China and Pakistan have indeed exchanged technology and equipment in several areas, including those related to nuclear weapons and missiles. For example, it is believed that Pakistan has imported short-range missiles from China. But the accusation that China supplied Pakistan with a

COURTESY OF SPACE IMAGING

# MAKING NUCLEAR WEAPONS MATERIAL

**THE MOST DIFFICULT** part of making nuclear weapons is manufacturing the fuel, either plutonium or highly enriched uranium. The starting point is natural uranium, which is 99.3 percent uranium 238 and 0.7 percent uranium 235. Only the latter can sustain a chain reaction. To build a uranium bomb, one needs to increase the uranium 235 content to 80 percent or more. Most modern enrichment facilities, including the ones in Pakistan and India, use high-speed centrifuges [see "The Gas Centrifuge," by Donald R. Olander; Scientific American, August 1978].

The alternative route involves plutonium. This element is not found in nature. It is produced by irradiating uranium fuel in nuclear reactors, then extracted through a chemical process called reprocessing [see "The Reprocessing of Nuclear Fuels," by William P. Bebbington; Scientific American, December 1976]. In the most commonly followed reprocessing scheme, the irradiated fuel is chopped up, dissolved in acid and exposed to a solvent called tributyl phosphate mixed with kerosene. The solvent separates out the plutonium and uranium from other fission products. Plutonium is then precipitated out by a reductant, a chemical that changes it to an insoluble form. *—M.V.R. and A.H.N.*


**INDIAN PLUTONIUM INVENTORY**

| | |
|---|---|
| Cumulative production (in reactors): | 450–722 kg |
| Consumption (in tests and reactors): | 165 kg |
| Net stock: | 285–557 kg (equivalent to 55–110 bombs) |

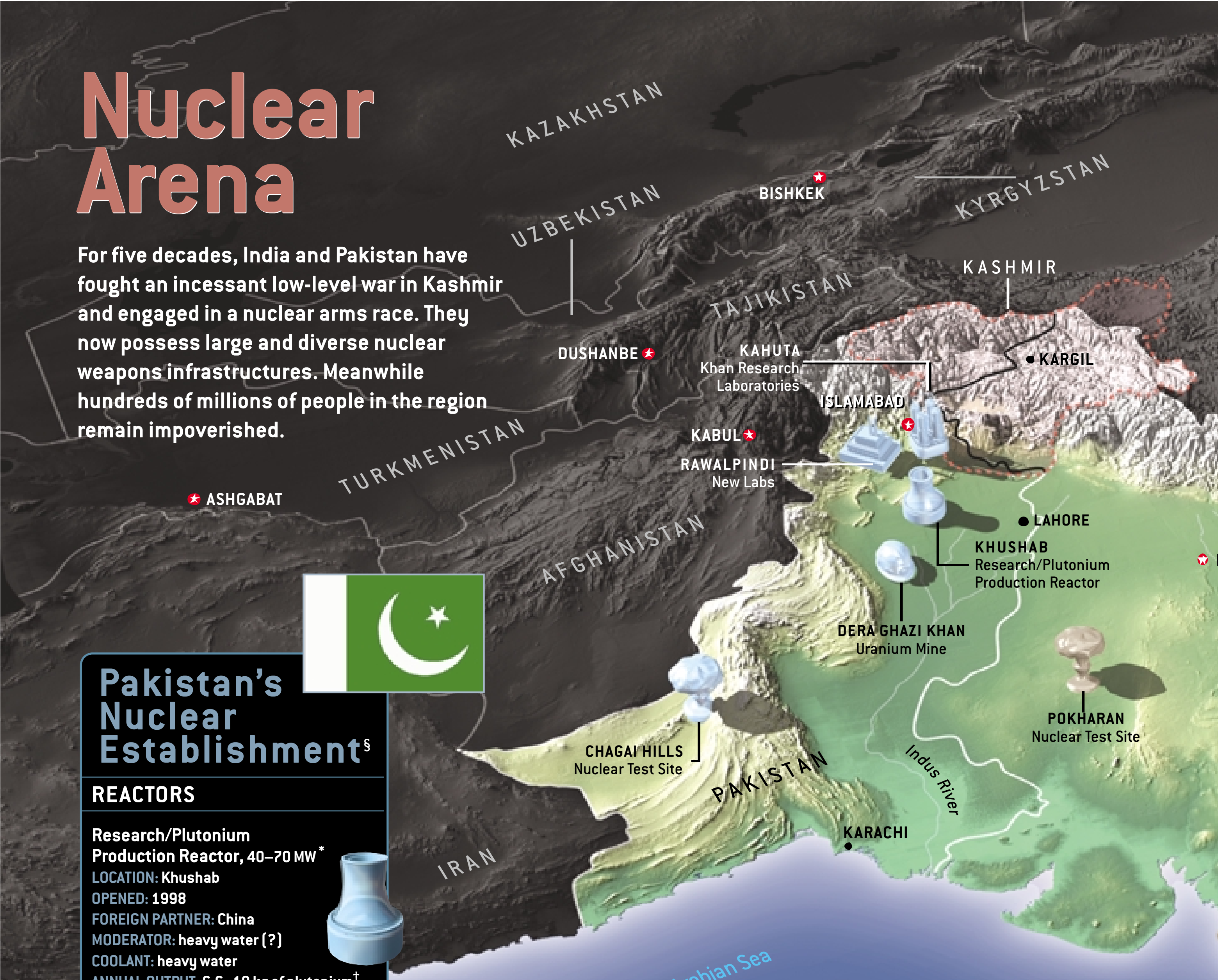
**PAKISTANI ENRICHED-URANIUM INVENTORY**

| | |
|---|---|
| Cumulative production (by enrichment): | 450–750 kg |
| Consumption (during tests): | 120 kg |
| Net stock: | 330–630 kg (equivalent to 20–40 bombs) |

BRYAN CHRISTIE

# Nuclear Arena

For five decades, India and Pakistan have fought an incessant low-level war in Kashmir and engaged in a nuclear arms race. They now possess large and diverse nuclear weapons infrastructures. Meanwhile hundreds of millions of people in the region remain impoverished.




## Pakistan's Nuclear Establishment§

### REACTORS

**Research/Plutonium Production Reactor, 40–70 MW***
LOCATION: Khushab
OPENED: 1998
FOREIGN PARTNER: China
MODERATOR: heavy water (?)
COOLANT: heavy water
ANNUAL OUTPUT: 6.6–18 kg of plutonium†

### PLUTONIUM REPROCESSING

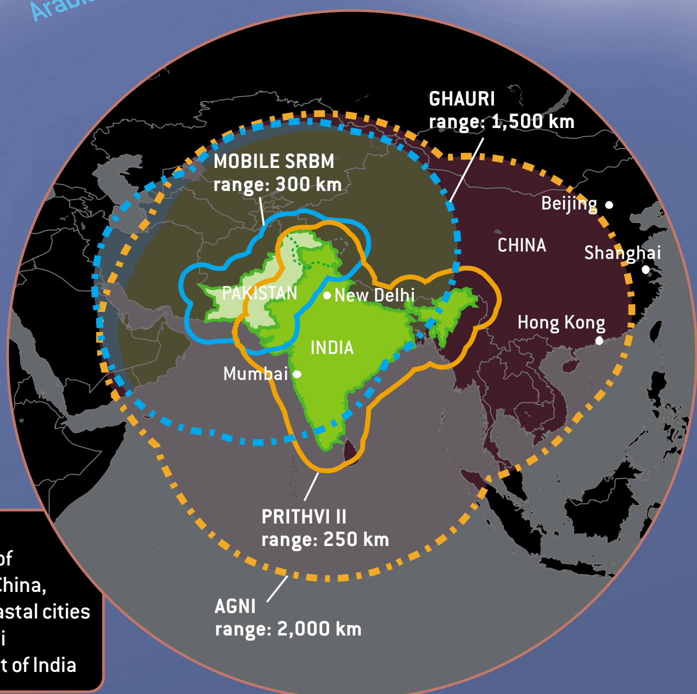
**New Labs**
LOCATION: Rawalpindi
OPENED: early 1980s
ANNUAL OUTPUT: 10–20 kg of plutonium

### URANIUM ENRICHMENT

**Khan Research Laboratories**
LOCATION: Kahuta
OPENED: 1984
ANNUAL OUTPUT: 57–93 kg of highly enriched uranium

### URANIUM MINE

**Dera Ghazi Khan**
OPENED: 1974
ANNUAL PRODUCTION: 23–30 tons




India's longest-range missiles can reach all of Pakistan and most of China, although the major coastal cities are a stretch. Pakistani equivalents cover most of India


# India's Nuclear Establishment§

## REACTORS

**CIRUS, 40 MW***
LOCATION: Mumbai
OPENED: 1960
FOREIGN PARTNER: Canada
MODERATOR: heavy water
COOLANT: light water
ANNUAL OUTPUT: 6.6–10.5 kg of plutonium†

**Dhruva, 100 MW***
LOCATION: Mumbai
OPENED: 1985
MODERATOR: heavy water
COOLANT: heavy water
ANNUAL OUTPUT: 16–26 kg of plutonium†

**Fast Breeder Test Reactor, 40 MW***
LOCATION: Kalpakkam
OPENED: 1983
FOREIGN PARTNER: France
COOLANT: liquid sodium
ANNUAL OUTPUT: 4–6.4 kg of plutonium†

## PLUTONIUM REPROCESSING

**Trombay**
LOCATION: Mumbai
COMMISIONED: 1964
ANNUAL CAPACITY: 30–50 tons of spent metallic fuel

**PREFRE**
LOCATION: Tarapur
COMMISSIONED: 1977
ANNUAL CAPACITY:100 tons of spent oxide fuel

**KARP**
LOCATION: Kalpakkam
COMMISSIONED: 1997
ANNUAL CAPACITY: 100–125 tons of spent oxide fuel

## URANIUM ENRICHMENT

**Rattehalli ††**
LOCATION: Mysore
OPENED: 1990
ANNUAL PRODUCTION: unknown

## URANIUM MINE

**Jadugoda**
OPENED: 1968
ANNUAL PRODUCTION: 200 tons

* thermal power output
† running at 50% – 90% of capacity
†† said to produce fuel for a nuclear submarine
§ bomb-related facilities; commercial power reactors omitted

weapons design has never been substantiated. And understandably, Pakistan's nuclear scientists have denied it.

In spring 1990 events in Kashmir threatened to erupt into another full-scale war. According to a 1993 *New Yorker* article by American journalist Seymour M. Hersh, U.S. satellites detected a convoy of trucks moving out of Kahuta, Pakistan's uranium-enrichment facility, toward an air base where F-16 fighter jets stood ready. Hersh reported that American diplomats conveyed this information to India, which recalled the troops it had amassed at the border. But the overwhelming opinion among scholars who have analyzed these claims is that Pakistan never contemplated the use of nuclear weapons; experts are also skeptical that U.S. satellites ever detected the claimed movement. Nevertheless, the Pakistani bomb lobby has used the allegations to assert that nuclear weapons protect the country from Indian attack. In India, officials have never acknowledged Hersh's story; it would be an admission that Pakistan's nuclear capability had neutralized India's conventional military advantage.

## "Now I Am Become Death"

FUTHER BUILDUP of nuclear capabilities in both countries took place against a background transformed by the end of the cold war. Superpower arsenals shrank, and the Comprehensive Test Ban Treaty, which prohibits explosive tests, was negotiated in 1996. But the five declared nuclear states—the U.S., Russia, Britain, France and China—made it clear that they intend to hold on to their arsenals. This ironic juxtaposition strengthened the bomb lobbies in India and Pakistan.

Domestic developments added to the pressure. India witnessed the rise of Hindu nationalism. For decades, parties subscribing to this ideology, such as the Bharatiya Janata Party (BJP), had espoused the acquisition of greater military capability—and nuclear weapons. It was therefore not surprising that the BJP ordered nuclear tests immediately after coming to power in March 1998.

The Indian tests, in turn, provided Pakistani nuclear advocates with the perfect excuse to test. Here again, religious extremists advocated the bomb. Qazi Hussain Ahmad of the Jamaat-e-Islami, one of the largest Islamist groups in Pakistan, had declared in 1993: "Let us wage jihad for Kashmir. A nuclear-armed Pakistan would deter India from a wider conflict." Meanwhile the military sought nuclear weapons to counter India's vastly larger armed forces.

This lobbying was partially offset by U.S. and Chinese diplomacy after India's tests. In addition, some analysts and activists enumerated the ill effects that would result from the economic sanctions that were sure to follow any test. They suggested that Pakistan not follow India's lead—leaving India to face international wrath alone—but to no avail. Three weeks after India's blasts, Pakistan went ahead with its own tests.

Bombast notwithstanding, the small size of seismic signals from the tests of both countries has cast doubt on the declared explosive yields [*see illustration on page 74*]. The data released by the Indian weapons establishment to support its

DONNING THE MASK OF DEATH and bearing the Indian flag, protesters gather outside the Pakistani Embassy in New Delhi after Pakistan's nuclear tests in 1998. Some are holding up baby bottles to mock Pakistan as an infant nation. It is not known whether the same protesters had objected to India's own nuclear tests several weeks earlier.

AP PHOTO

claims are seriously deficient; for example, a graph said to be of yields of radioactive by-products has no units on the axes. Independent scientists have not been able to verify that the countries set off as many devices as they profess.

Whatever the details, the tests have dramatically changed the military situation in South Asia. They have spurred the development of more advanced weapons, missiles, submarines, antiballistic missile systems, and command-and-control systems. In August 1999 the Indian Draft Nuclear Doctrine called for the deployment of a triad of "aircraft, mobile land-missiles and sea-based assets" to deliver nuclear weapons. Such a system would cost about $8 billion. This past January the Indian government declared that it would deploy its new long-range missile. A month later the Pakistani deputy chief of naval staff announced that Pakistan was thinking about equipping at least one of its submarines with nuclear missiles.

## Critical Mass

DEPLOYMENT INCREASES the risk that nuclear weapons will be used in a crisis through accident or miscalculation. With missile flight times of three to five minutes between the two countries, early-warning systems are useless. Leaders may not learn of a launch until they look out their window and see a blinding flash of light. They will therefore keep their fingers close to the button or authorize others, geographically dispersed, to do so.

Broadly speaking, there are two scenarios. The first pos-

## INDIAN MISSILES

**AGNI ("Fire") I**
TYPE: Solid-fueled, first stage; liquid-fueled, second stage
RANGE: 1,500–2,000 km
WARHEAD: 1,000 kg
DEVELOPMENT STAGE: Suspended

**AGNI II**
TYPE: Solid-fueled, two-stage
RANGE: 2,000 km
WARHEAD: 1,000 kg
DEVELOPMENT STAGE: Tested April 1999

**PRITHVI ("Earth") I**
TYPE: Liquid-fueled, single-stage; engine based on Russian SA-2 air defense missile
RANGE: 150 km
WARHEAD: 1,000 kg
DEVELOPMENT STAGE: Deployed

**PRITHVI II**
TYPE: Liquid-fueled, single-stage
RANGE: 250 km
WARHEAD: 500 kg
DEVELOPMENT STAGE: Tested January 1996

**PRITHVI III**
TYPE: Liquid-fueled, single-stage naval missile
RANGE: 350 km (?)
WARHEAD: Unknown
DEVELOPMENT STAGE: Under development

**SAGARIKA** *(not shown)*
**("Born on the Ocean")**
TYPE: Submarine-launched cruise/ballistic missile
RANGE: 300 km (?)
WARHEAD: Unknown
DEVELOPMENT STAGE: Under development

## PAKISTANI MISSILES

**HATF ("Armor") I**
TYPE: Solid-fueled, single-stage; based on French sounding rocket
RANGE: 60–80 km
WARHEAD: 500 kg
DEVELOPMENT STAGE: Tested January 1989

**HATF II**
TYPE: Solid-fueled, single-stage
RANGE : 280–300 km
WARHEAD: 500 kg
DEVELOPMENT STAGE: Tested January 1989

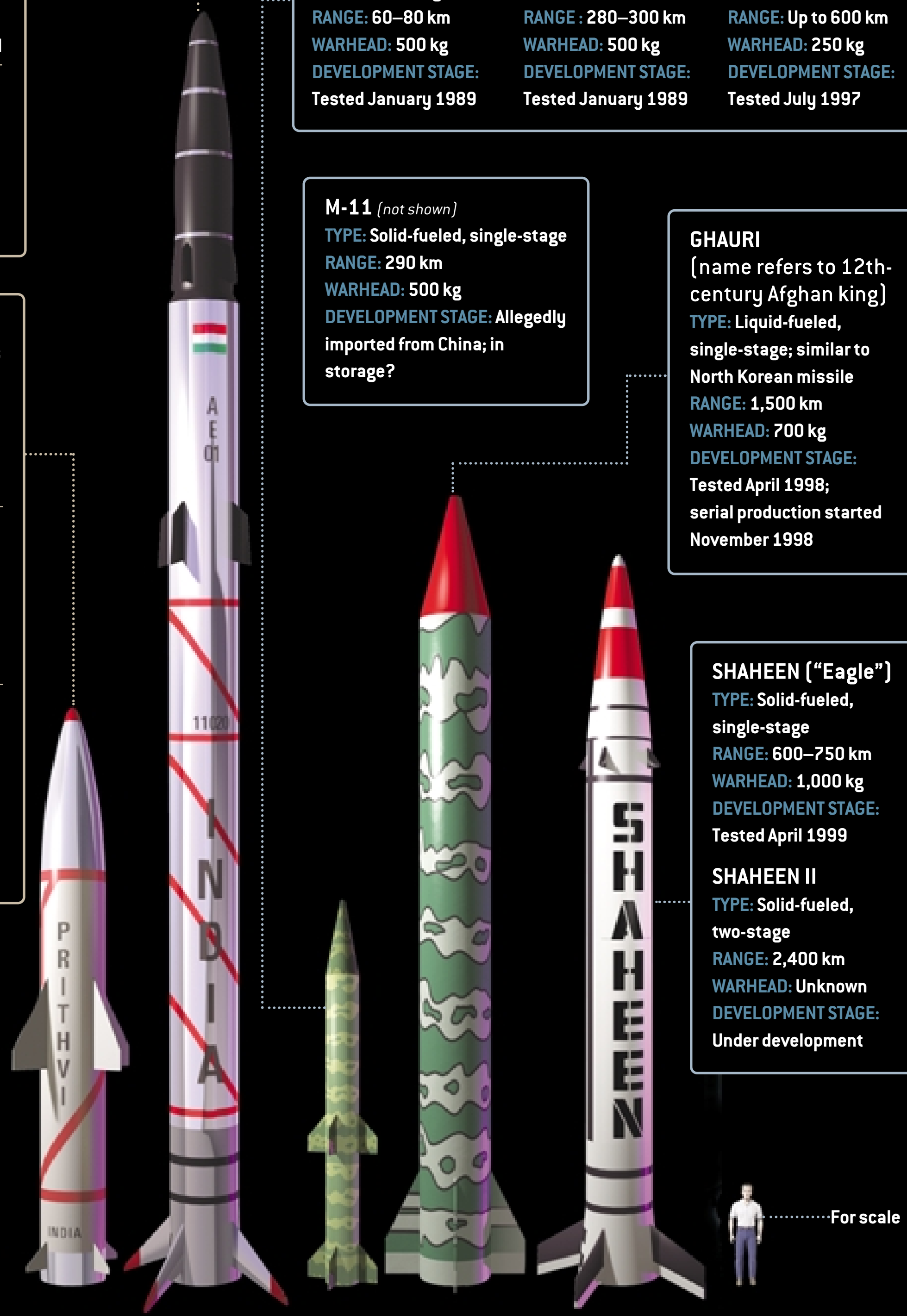
**HATF III**
TYPE: Solid-fueled, single-stage
RANGE: Up to 600 km
WARHEAD: 250 kg
DEVELOPMENT STAGE: Tested July 1997

**M-11** *(not shown)*
TYPE: Solid-fueled, single-stage
RANGE: 290 km
WARHEAD: 500 kg
DEVELOPMENT STAGE: Allegedly imported from China; in storage?

**GHAURI**
(name refers to 12th-century Afghan king)
TYPE: Liquid-fueled, single-stage; similar to North Korean missile
RANGE: 1,500 km
WARHEAD: 700 kg
DEVELOPMENT STAGE: Tested April 1998; serial production started November 1998

**SHAHEEN ("Eagle")**
TYPE: Solid-fueled, single-stage
RANGE: 600–750 km
WARHEAD: 1,000 kg
DEVELOPMENT STAGE: Tested April 1999

**SHAHEEN II**
TYPE: Solid-fueled, two-stage
RANGE: 2,400 km
WARHEAD: Unknown
DEVELOPMENT STAGE: Under development




BRYAN CHRISTIE

tulates that India crosses some threshold during a war—its troops reach the outskirts of Lahore or its ships impose a naval blockade on Karachi—and Pakistan responds with tactical nuclear weapons as a warning shot. The other scenario supposes that under the same circumstances, Pakistan decides that a warning shot would not work and instead attacks an Indian city directly. In 1998 one of us (Ramana) conducted the first scientific study of how much damage a modest, 15-kiloton bomb dropped on Bombay would cause: over the first few months, between 150,000 and 850,000 people would die.

The Indian military is already preparing for these eventualities. This past May it carried out its biggest exercises in more than a decade, called Operation Complete Victory. Tens

# Secrets, What Secrets?

## Terrorists might exploit Pakistan's cavalier attitude toward nuclear information

by David Albright

Over the years, successive Pakistani governments have assured the West that they had a secure grip on the country's nuclear weapons, materials and technology. But nuclear analysts have never been entirely comforted by these assertions. Many people in the Pakistani nuclear weapons program and the military could well be sympathetic to radical Islamist or anti-American causes. What is especially worrisome is that the historical development of Pakistan's program has heightened the risk of illegal assistance and other security violations.

From its inception, the program has relied on illicit procurement and deliberate deception. It has fostered extensive contacts with the world of shady middlemen and companies whose allegiance to Western export controls depends on the price one is willing to pay. In the organizational culture of such a program, disaffected individuals could find plenty of justifications and opportunities to transfer classified information or sensitive items. Others might be disinclined to report on the suspicious actions of colleagues. Some might even feel ownership over parts of the program and believe it is their right to sell their contributions for personal benefit.

Such problems affect India less, because it started its nuclear weapons program earlier than Pakistan did. India obtained much of its nuclear infrastructure from foreign suppliers before Western governments understood the extent to which developing countries were misusing civilian nuclear assistance to make nuclear explosives. To be sure, Pakistan is not alone in dealing with an organizational culture that scorns security guidelines. The German civilian gas-centrifuge program was notorious for its weak security. In the late 1980s German nuclear experts secretly assisted Iraq.

A key component of Pakistan's program, the production of highly enriched uranium for bombs, was born in an act of industrial espionage. In the mid-1970s the father of that effort, A. Q. Khan, worked at a Dutch engineering firm and was given the task of translating classified designs and specifications for gas centrifuges. He gained access to a wide variety of sensitive information. On his return to Pakistan, Khan founded the Engineering Research Laboratories, now known as the Dr. A. Q. Khan Research Laboratories, to transform this knowledge into a bomb factory.

According to a declassified 1983 U.S. State Department memorandum, the enrichment program disguised its activities by providing false statements about the final use of items imported from Western countries. Pakistan once described its gas-centrifuge plant as a synthetic butter factory. In a 1999 interview in the Egyptian newspaper *Al-Ahram,* Khan said that his program purchased items through offshore front companies in Japan, Singapore and elsewhere. Those companies took a cut of 15 to 25 percent of the purchase price.

Khan and his colleagues took a Robin Hood approach to classified information. In the late 1980s they published a series of technical articles in Western journals about gas centrifuges. The intention was to demonstrate Pakistan's self-sufficiency in centrifuges and thereby signal that the country was ready to make a bomb. One paper stated its purpose thus: to "provide useful and practical information, as technical information on balancing of

COURTESY OF URENCO NEDERLAND

DESIGNS FOR GAS CENTRIFUGES like these, operated by the Urenco-Centec industrial group in Europe, were acquired covertly by Pakistani bomb makers.

of thousands of troops, backed by tanks, aircraft and attack helicopters, undertook drills close to the border with Pakistan. The stated aim was to train the armed forces to operate in an "environment of chemical, biological and nuclear assault" and "to teach the enemy a lesson once and for all." In one significant exercise, the military had to "handle a warlike situation wherein an enemy aircraft is encountered carrying a nuclear warhead." Abdul Kalam, head of India's missile program, said that India's nuclear weapons "are being tested for military operations ... for training by our armed forces."

Even before September 11, South Asia had all the ingredients for a nuclear war: possession and continued development of bombs and missiles, imminent deployment of nuclear weapons, inadequate precautions to avoid unauthorized use of these weapons, geographical proximity, ongoing conflict in Kashmir, militaristic religious extremist movements, and leaders who seem sanguine about the dangers of nuclear war.

The responses of India and Pakistan to the events of September 11 and the U.S.-led attack on targets in Afghanistan reflect the strategic competition that has shaped much of their history. India was quick to offer air bases and logistical support to the U.S. military so as to isolate Pakistan. Attempting to tie its own problems in Kashmir with the global concern about terrorism, Indian officials even threatened to launch attacks on Pakistani supply lines and alleged training camps for militants fighting in Kashmir. Pakistan, for its part, realizing both the geopolitical advantage it possessed and the dangers of civil instability, deliberated before agreeing to provide support to fight the Taliban. The diplomatic machinations, war in Afghanistan and violence in Kashmir may well have worsened the prospects for peace on the subcontinent. The lifting of American sanctions, which had been imposed in the 1990s, freed up resources to invest in weapons.

The limitations of Western nonproliferation policy are now painfully obvious. It has relied primarily on supply-side export controls to prevent access to nuclear technologies. But Pakistan's program reveals that these are inadequate. Any effective strategy for nonproliferation must also involve demand-side measures—policies to assure countries that the bomb is not a requisite for true security. The most important demand-side measure is progress toward global nuclear disarmament. Some people argue that global disarmament and nonproliferation are unrelated. But as George Perkovich of the W. Alton Jones Foundation in Charlottesville, Va., observed in his masterly study of the Indian nuclear program, that premise is "the grandest illusion of the nuclear age." It may also be the most dangerous. SA

centrifuge rotors is hardly available because most of the work is shrouded in the clouds of the so-called secrecy." These articles aided other countries, such as South Africa, in their own nuclear programs.

One Pakistani article is the only publicly available study on bellows built from maraging steel, a superstrong type of steel. For years, Urenco—a British, German and Dutch enrichment consortium—considered the mere mention of these bellows a violation of its secrecy rules.

How much further did the Pakistani nuclear scientists go in spreading the art of bomb making? The U.N. arms inspections in Iraq came across a one-page Iraqi intelligence document, marked TOP SECRET, that contained an offer of nuclear weapons assistance from the Pakistanis. According to the document, an intermediary approached Iraqi intelligence in October 1990—two months after the Iraqi invasion of Kuwait and three months before the U.S.-led counterattack—with the following proposition: Khan would give Iraq bomb designs, help to procure materials through a company in Dubai and provide other services. In return, Iraq would pay handsomely.

Arms inspectors were unable to find the middleman, and Pakistan and Khan have denied any involvement. Nevertheless, the Iraqis took this offer as genuine—and apparently rejected it. Khidhir Hamza, a former weapons scientist who left Iraq in 1994 and worked with me in the late 1990s, says he knew of this offer at the time and believes Iraq would not have pursued it, for fear that Khan would gain too much knowledge about, and control over, Iraq's nuclear programs. Khan already had a track record of misleading the Iraqis, having used a contract for a petrochemical facility as a cover to obtain maraging steel.

In March of this year the government of Pakistan removed Khan as head of the nuclear laboratory and offered him a position as a special science and technology adviser. The move is widely viewed as an attempt to rein him in. This past summer, however, reports emerged that the laboratory has kept up its ties with North Korea's ballistic-missile program, reviving fears of nuclear cooperation. Pakistani officials have denied any connection.

No evidence links elements in the Pakistani government with any terrorist group, but the Pakistani government has had extensive contact with the Taliban. It is conceivable that terrorists could exploit these connections to gain access to sensitive nuclear items. The culture within the nuclear program increases this risk.

*David Albright is a physicist, president of the Institute for Science and International Security in Washington, D.C., and a former U.N. weapons inspector in Iraq.*

## MORE TO EXPLORE

**Fissile Material Production Potential in South Asia.** A. H. Nayyar, A. H. Toor and Zia Mian in *Science and Global Security,* Vol. 6, No. 2, pages 189–203; 1997.

**The Making of the Indian Atomic Bomb: Science, Secrecy and the Postcolonial State.** Itty Abraham. Zed Books, 1998.

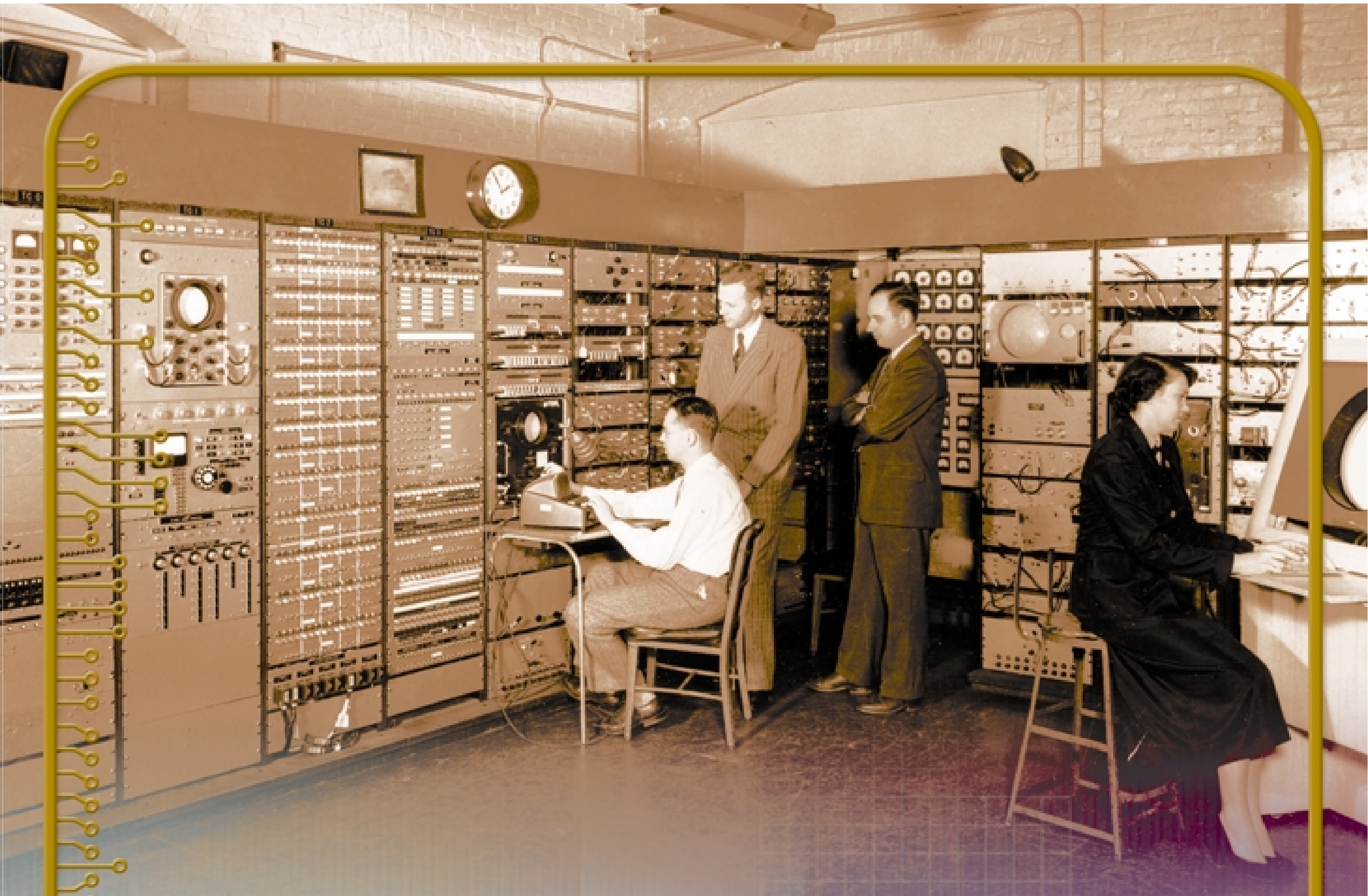
**India's Nuclear Bomb: The Impact on Global Proliferation.** George Perkovich. University of California Press, 1999.

**Out of the Nuclear Shadow.** Edited by Smitu Kothari and Zia Mian. Zed Books, 2001.

For articles by M. V. Ramana, visit
**www.geocities.com/CollegePark/5409/nuclear.html**

For extensive information on both countries' nuclear weapons, visit
**www.isis-online.org/publications/southasia/index.html**
**www.ceip.org/files/Publications/trackingTOC.asp**
**www.fas.org/nuke/hew/**

# the ORIGINS



1944

1945

1944
Flight simulator
project begins at M.I.T.

# of PERSONAL COMPUTING

## Forget Gates, Jobs and Wozniak. The foundations of modern interactive computers were laid decades earlier

**By M. Mitchell Waldrop**

Just over a quarter of a century ago, according to the standard accounts, a pack of techno-savvy kids with names like Gates, Jobs and Wozniak began to play around in their garages and dorm rooms with the new technology of microprocessors—and ended up pioneering the personal computer revolution almost by accident.

In reality, however, the story actually didn't start with these young entrepreneurs. After all, what really put the fire in the PC revolution (and in the Internet revolution that would follow) wasn't the hardware or the software per se but the message those products embodied. This was the idea that computers didn't have to be huge, ominous machines sitting off in a back room somewhere, processing punch cards for some large institution. Instead they could be humane, intimate machines, responding to us and helping us as individuals. Computers could enhance human creativity, democratize access to information, foster wider communities and build a new global commons for communications and commerce. Computers, in short, could be instruments of individual empowerment. The irony is that the foundations of that vision had been laid more than three decades earlier—by the very same government and the very same establishment that the 1970s generation so distrusted.

### Real-Time Responses

CONSIDER, FOR EXAMPLE, that personal computing as we know it today would be inconceivable without the much more basic notion of *interactive* computing: that a machine could respond to a user's actions right now, as opposed to delivering a fanfold printout six hours from now.

This was certainly not an obvious idea in the early days, not when computers were still being thought of as little more than superfast calculating machines. From the first all-electronic digital computers developed at the end of World War II through the first big wave of corporate computerization in the early 1960s, virtually every computer in existence was designed to grind away at a given problem until it spit out an answer and then—like a standard desktop calculator—wait for new input.

Even in the beginning, however, there was one exception to the rule: Whirlwind, an experimental computer developed at the Massachusetts Institute of Technology with funding from the U.S. Navy. Significantly, the Whirlwind project had started in 1944 as a wartime effort to build not a calculator but an all-electronic flight simulator—a machine for which there was never an "answer," just a constantly changing sequence of pilot actions and simulated aircraft responses. Team leader Jay Forrester and his colleagues quickly realized that the computer they built to control the simulator would have to be interactive from the ground up and capable of responding to events as fast as they occurred. That is, it would have to be the world's first real-time computer.

Forrester and his team also foresaw that interactive, real-time computing might be far more important than the flight simulator itself; potential applications ranged from logistics and the coordination of naval task forces to antiballistic missile defense and air-traffic control. So in 1948 the researchers talked the navy into upgrading the Whirlwind project, turning it into a demonstration of general-purpose real-time computing funded at $1 million a year—by far the largest and most expensive computer effort of its day.

The machine had the bulk to match. When Whirlwind finally became operational in 1951, its eight tall racks of vacuum-tube electronics occupied the space of a small house, with enough room for researchers to walk around inside. Its performance was likewise impressive, being roughly equivalent to a 1980-vintage personal computer such as the TRS-80. Indeed, Whirlwind was the first machine to be used as a personal computer: individuals signed up for 15-minute sessions, during which they could sit at Whirlwind's cathode-ray-tube (CRT) display and write code, run simulations or just play around.

Unfortunately, by that point the navy had gotten tired of paying for Whirlwind and was threatening to shut it down. What saved the project—and with it, the future of interactive

TOM DRAPER DESIGN; M.I.T. MUSEUM (*flight simulator*)

1946

1947

1948

1948
Flight simulator becomes Whirlwind (*left*), a general-purpose real-time computer

1949

1951–1958
SAGE (*left*) development at M.I.T.

1951
Whirlwind becomes operational

## The ARPA Community

1957
J.C.R. Licklider (*below*) conceives the "Truly SAGE System"

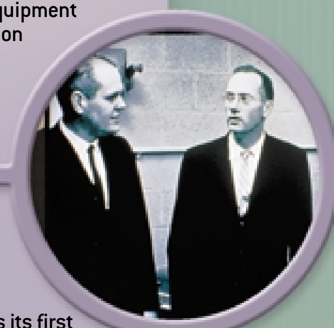
1960
Licklider publishes "Man-Computer Symbiosis"

1962
Licklider organizes computer research office at the Advanced Research Projects Agency (ARPA)

1963
Project MAC begins at M.I.T.; Licklider writes memo proposing a nationwide "Intergalactic Computer Network"

## DEC and Minicomputers

1957
Kenneth Olsen (*below left*) and Harlan Anderson found Digital Equipment Corporation



1960
DEC ships its first computer, the PDP-1 (for programmable data processor)

1957

1958

1959

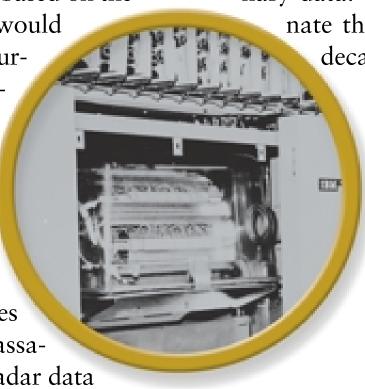
1960

1961

1962

1963

computing—was the Soviet Union's first atomic bomb test in August 1949, which raised the specter of a surprise attack by long-range Soviet bombers. In 1951 the U.S. Air Force accordingly commissioned M.I.T. to design a brand-new, state-of-the-art early-warning system in which real-time computers based on the Whirlwind design would coordinate radar surveillance, target tracking and all other operations. On April 20, 1951, Whirlwind demonstrated the feasibility of that concept by tracking three propeller planes in the skies over Massachusetts, taking in radar data and computing interception trajectories that steered the "defending fighter" to within 1,000 yards of the "attackers." Soon thereafter, the machine became the centerpiece of a full-scale development program, dubbed Project Lincoln.



Millions of dollars and more than half a decade later, the result was SAGE (for Semi-Automatic Ground Environment): a continent-spanning system of 23 direction centers, each housing up to 50 human operators, plus two redundant real-time computers capable of tracking up to 400 airplanes at once. Because it was never tested in combat, SAGE's military effectiveness is debatable. (It was decommissioned in 1984.) But there is no doubt about its impact on the history of computing.

First, it helped to create the Silicon Valley of the East. In 1952, most notably, Project Lincoln acquired a new name—Lincoln Laboratory—and a new home in suburban Lexington, not too far from a major ring road around Boston: Route 128. Other high-tech organizations soon followed. Second, SAGE was the pipeline that transported Whirlwind's technology into the commercial world. By the late 1950s, for example, IBM was using its experience as prime contractor for the SAGE computers to create a nationwide, real-time ticketing system for American Airlines. It went into operation in 1964 as SABRE (for Semi-Automatic Business-Related Environment) and would be a model for all point-of-sale transaction systems to come. In 1955 IBM likewise became the first manufacturer to market a business computer using Whirlwind's "magnetic core memory" technology, which was far cheaper and more reliable than any previous method of storing binary data. Core memory would dominate the industry for another two decades to come, giving way to semiconductor memory chips only in the mid-1970s.

SAGE produced a standard operator's console that would ultimately evolve into the modern desktop computing environment.

Perhaps most relevant to this story, however, SAGE produced a standard operator's console that would ultimately evolve into the modern desktop computing environment. Each radar controller's console included a CRT-display screen, a keyboard and a handheld device that allowed the operator to select various items on the screen (it was a light gun, not a mouse, but it was used in a similar way). Behind the scenes, moreover, the computers at all 23 direction centers were linked together by a long-distance digital network operating over telephone lines. Indeed, to transmit digital signals through lines meant for analog signals, the SAGE designers had to develop another familiar gadget: the modem.

Of course, the path from SAGE to the modern PC was hardly a straight line. It was more like two lines of development in parallel: one focused on hardware and the other on how people used that hardware.

## Hands-On Hardware

THE HARDWARE TRACK grew out of those 15-minute chunks of personal time on Whirlwind and the other machines developed for the SAGE project. Using the machines in that way had inspired some of the project's younger participants to think that computers ought to be just as much fun for everyone. So in 1957 two of them, Kenneth Olsen and Harlan Anderson, founded a little company to bring such interactive computers to market. They called it DEC: Digital Equipment Corporation.

The market was lukewarm at first. When DEC introduced its first programmable data processor, the PDP-1, in 1960, the company sold only 49 units—a respectable number but hardly spectacular. Among scientists and engineers, however, the PDP-1 was a hit. The machine was fully interactive, had a built-in CRT display, could fit into a single, smallish room and delivered a remarkable amount of computational power, considering that it cost a mere $120,000. More tantalizing still, the PDP-1 was "open": all the details of the hardware were publicly spelled out, so that technically savvy users could modify their machine or add to it in any way they wanted—which they did.

By 1964, in fact, users' enthusiasm for the PDP-1 and the company's success with other products led DEC to start development of a "tabletop" computer intended for small groups or even individuals. Much of the design was based on LINC, an experimental laboratory computer developed at Lincoln Laboratory by Olsen and Anderson's former colleague Wesley Clark, who had since moved his team to Washington University. But inspiration also came from the rapid advances being made in semiconductor and storage technologies, not to mention assembly-line manufacturing techniques.

The result was the PDP-8, a computer so unbelievably small and light—just 250 pounds—that one of DEC's ads showed it riding in the backseat of a Volkswagen Beetle convertible. It also had an unbelievably low price: $18,000. And it was irresistible to customers: even users without technical backgrounds loved a computer they could get their hands on.

M.I.T. MUSEUM (*Whirlwind, SAGE, Licklider*); COMPAQ COMPUTER CORPORATION (*Olsen and Anderson, PDP-1*)

1964

1965

1966

1967

1968

1969

1970

1971

1972

1973

1974

1975

1976

1977

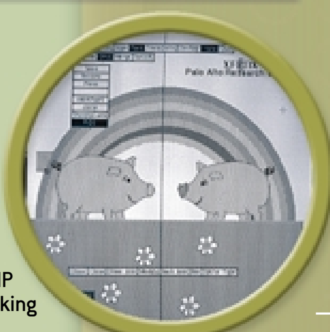
## Xerox PARC

1966
Licklider's successor, Robert W. Taylor, decides to build the Intergalactic Computer Network

1968
Douglas Engelbart (*above*) demonstrates mouse, on-screen windows and much else

1969
Arpanet begins operation

1970
Xerox founds Palo Alto Research Center, hires many top students from ARPA-funded universities



1974
Robert E. Kahn and Vinton Cerf devise TCP/IP inter-networking protocol

1976
Cerf (*below*) becomes ARPA net-working chief, begins switch to TCP/IP

1971–1977
PARC creates modern desktop computing environment: stand-alone personal computer; graphical user interface (*above*) with windows, icons, menus and mouse; Ethernet local-area network; laser printing; WYSIWYG word processing; and more

## Microcomputers

1965
DEC ships PDP-8, the first minicomputer

1969
Data General ships NOVA, its first 16-bit minicomputer

1970
DEC responds with PDP-11

1975
*Popular Electronics* Altair cover; Gates and Allen write Altair BASIC, found Micro Soft; Silicon Valley hobbyists launch Homebrew Computer Club


Bill Gates (*left*) and Paul Allen

Steve Jobs

Steve Wozniak

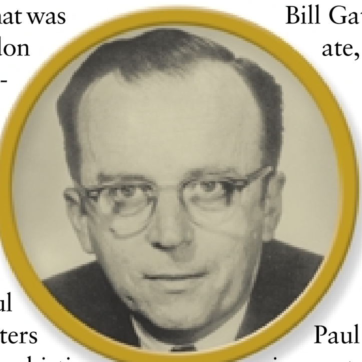
1976
Steve Jobs and Steve Wozniak found Apple Computer Company; Gary Kildall announces CP/M, an early micro-computer operating system


1977
DEC announces VAX, its first 32-bit minicomputer

1977
First consumer micros, including Apple II, TRS-80

The machine was first shipped in April 1965 and was soon showing up in chemical plants, newspapers, laboratories, refineries and even schools.

With competitors by the dozens rushing to offer similar machines, it soon became clear that the PDP-8 was the prototype for a whole new genre of minicomputers—a name that was coined at DEC's London office and that was apparently inspired by a hot new fashion item called the miniskirt. By the mid-1970s, moreover, with the competition producing ever more powerful waves of minicomputers based on ever more sophisticated semiconductor technology, DEC and its rivals had begun to make serious inroads into the market for IBM-style mainframes.



This same aura of hands-on independence had also intrigued electronics hobbyists. Many of them had encountered minicomputers at work or at school and now wanted one they could tinker with at home. The now famous issue of *Popular Electronics* magazine hit the newsstands in January 1975. The cover photo showed a pale blue box with an array of switches and diodes on the front and a name in the upper left-hand corner: Altair 8800. "World's First Minicomputer Kit to Rival Commercial Models," the headline proclaimed.

Inside, readers learned that the $397 kit could be ordered from MITS, a hobbyist firm in Albuquerque, N.M., and that it was based on the Intel 8080, a microprocessor that placed a computer's entire central processing unit, or CPU, on a single chip. (Conventional minicomputers had CPUs comprising several chips or even several circuit boards.) In modern parlance, that made the Altair a *micro*computer. In fact, it would prove to be the first commercially successful microcomputer. But as the *Popular Electronics* banner suggested, the Altair had been conceived of from the beginning as a mini. It looked like a mini. It had the open architecture of a mini. And it could use the same peripherals as a mini. Indeed, except for the 8080 chip, the Altair *was* a minicomputer.

Even MITS's later choice for an official programming language was reminiscent of the minis: created in the spring of 1975 by two Seattle natives who had been inspired by the *Popular Electronics* article—Bill Gates, a Harvard undergraduate, and his high school buddy Paul Allen, a programmer working outside Boston—Altair BASIC took a number of key features from DEC's BASIC for the PDP-11. (Once the language was ready, Allen quit his job, Gates dropped out of school, they both moved to Albuquerque to be near MITS, and together they formed a little company called Micro Soft to market it.)

The Altair was a smash hit in the hobbyist world, ultimately selling more than 10,000 units. It inspired the formation of many user groups, including the legendary Homebrew Computer Club, which held its first meeting in a Palo Alto, Calif., garage in March 1975.

And within a year or two, with competing microcomputers appearing by the dozens, a few young entrepreneurs had begun to think about marketing their machines to consumers. True, this would require offering not just a kit but something more like an appliance, which would work as soon as the customer plugged it in. But several firms took up the challenge—most memorably the Apple Computer Company, founded in 1976 by Homebrew Computer Club members Steve Wozniak and Steve Jobs, longtime friends from the Silicon Valley town of Cupertino. Their Apple II machine, introduced in April 1977, had a built-in keyboard and a professionally designed beige case. It cost just $1,195 without the monitor. And best of all, it was great for playing video games.

By decade's end, Apple had become the fastest-growing company on record. And the rest, of course, is history—albeit a history that was strongly shaped by the second line of development from the SAGE project.

## Man-Machine Symbiosis

THE PIVOTAL MOMENT in this story had come in 1962, when the Pentagon's Advanced Research Projects Agency (ARPA) hired an experimental psychologist named J.C.R. Licklider to organize a new research program on command and control. A decade earlier, at M.I.T., Licklider had been a member of the SAGE console design team, specializing in the human-factors aspects. By 1957 this experience had led him to envision a "Truly SAGE System" that would be focused not on national security but on enhancing the power of the mind. In place of the 23 air defense centers, he imagined a nationwide network of "thinking centers," with computers containing vast libraries covering every subject imaginable. And in place of the radar consoles, he imagined a multitude of interactive terminals, each capable of displaying text, equations, pictures, diagrams or any other form of information.

By 1958 Licklider had begun to talk about this vision as a "symbiosis" of humans and machines, each preeminent in

**By 1958 Licklider had begun to talk about a "symbiosis" of humans and machines.**


THE AUTHOR

*M. MITCHELL WALDROP* is the author of a new book about the history of computing, *The Dream Machine*, from which this article is adapted. He earned a Ph.D. in elementary particle physics and a master's degree in journalism at the University of Wisconsin–Madison. He has been a staff reporter at *Chemical and Engineering News* and *Science* magazines. He is now an independent writer living in Washington, D.C. He is the author of two earlier books, *Man-Made Minds* (1987), about artificial intelligence, and *Complexity* (1992), about the Santa Fe Institute and the sciences of complexity.



COMPAQ COMPUTER CORPORATION (*PDP-8, VAX*); BOOTSTRAP INSTITUTE (*Engelbart*); COURTESY OF XEROX PARC (*Xerox PARC founders, graphical user interface*); DOUG WILSON *Corbis* (*Gates and Allen*); AP PHOTO/APPLE COMPUTERS, INC. (*Jobs*); STEVE CASTILLO *AP Photo* (*Wozniak*); REPRODUCED FROM *POPULAR ELECTRONICS*, JAN. 1975, BY PERMISSION OF GERNSBACK PUBLICATIONS, INC. (*cover*); COURTESY OF VINTON CERF (*Cerf*)

1979
VisiCalc, the first electronic spreadsheet and first "killer app," written for Apple II

EARLY 1980s
PARC technology inspires new breed of "workstations" from Sun, SGI and many others

1981
Xerox introduces commercial version of PARC system: the Xerox Star (*left*). The market is lukewarm

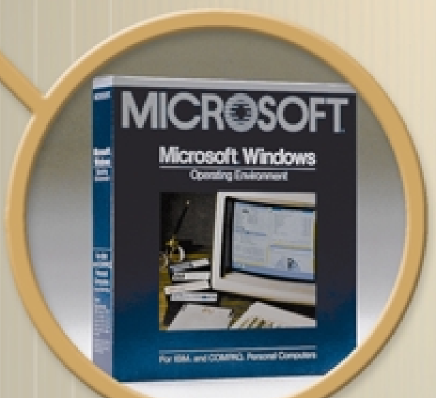
1981
Original IBM PC; micros begin invasion of corporate offices

1983
Arpanet officially converts to TCP/IP

1983
*Time* declares computer "Machine of the Year"; Lotus introduces 1-2-3, first killer app for the PC

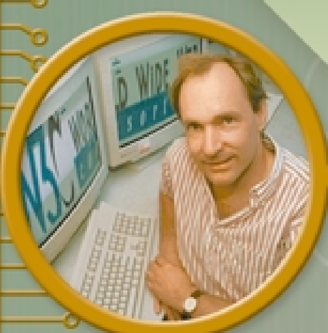
1984
Apple introduces Macintosh, using PARC-inspired GUI (graphical user interface)


**Internet**



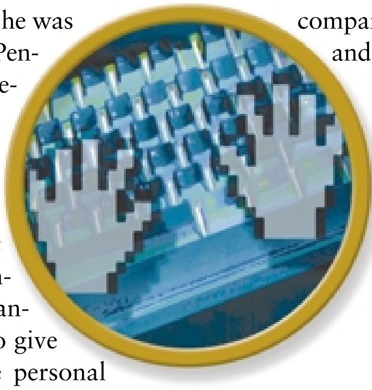

1985
Microsoft introduces Windows 1.0; PARC technology begins to move into the mass market

1986
National Science Foundation launches NSFnet; campus networking explodes



1990
Arpanet formally ends; Tim Berners-Lee (*above*) creates World Wide Web; Internet expands into mass market

its own sphere—rote algorithms for computers, creative heuristics for humans—but together far more powerful than either could be separately. By 1960 he had written down these ideas in detail (in the classic article "Man-Computer Symbiosis"), in effect laying out a research agenda for how to make his vision a reality. And now, at ARPA, he was determined to use the Pentagon's money to implement that agenda.



By the mid-1960s Project MAC would evolve into the world's first online community—complete with hackers.

Licklider's flagship endeavor was M.I.T.'s Project MAC, the first large-scale experiment ever in personal computing. The project managers couldn't hope to give anyone a stand-alone personal computer, of course, not with the cheapest machines still costing hundreds of thousands of dollars. But they could scatter dozens of remote terminals around the campus and in people's homes. And then through the technology of time-sharing they could tell their big, central machine to dole out little slices of processing time very, very rapidly, so that each user would feel as if it were responding to him or her as an individual, in real time. By the mid-1960s Project MAC would evolve into the world's first online community, complete with bulletin boards, e-mail, "virtual" friendships, a "freeware" exchange—and hackers.

Another of Licklider's beneficiaries was Douglas C. Engelbart, a soft-spoken engineer at SRI International, a large high-tech consulting firm in Menlo Park, Calif., near what would soon become Silicon Valley. Engelbart's bosses had found his ideas on "augmenting the human intellect" to be incomprehensible, but Licklider had immediately recognized them as identical to his own symbiosis vision. With funding from ARPA (as well as from the air force and NASA), Engelbart would go on to invent the mouse, on-screen windows, hypertext, full-screen word processing and a host of other innovations. Indeed, Engelbart's December 1968 demonstration of these marvels at the Fall Joint Computer Conference in San Francisco would be remembered as one of the turning points in computer history: the moment when large numbers of computer professionals finally began to understand what interactive computing could do.

In general, Licklider's strategy at ARPA was to seek out isolated research groups that were already doing work consistent with his vision, nurture them with his comparatively ample funding, and forge them into a nationwide movement that would carry on after he was gone. (He would actually leave in 1964, first for IBM and then for a new position at M.I.T.) On April 25, 1963, in a memo to "the members and affiliates of the Intergalactic Computer Network"—that is, his principal investigators—Licklider outlined a key part of that strategy: connect all their individual computers and time-sharing systems to a single computer network spanning the continent. By the late 1960s Licklider's handpicked successors at ARPA would begin to implement his intergalactic network as the Arpanet, a nationwide digital network that connected all the ARPA-funded computer research sites. By the 1970s, moreover, they would begin to expand the Arpanet into the network of networks known today as the Internet.

Another, arguably even more important part of Licklider's strategy was to pour most of his research funding into universities. His successors would continue that policy, and as a result, Licklider's vision of human-computer symbiosis was soon being carried into the mainstream computer industry by a whole generation of computer science graduates. This was the generation that would build the Arpanet. This was the generation that would gather in the 1970s at PARC, the Xerox Corporation's legendary Palo Alto Research Center, where they would put Licklider's symbiosis vision into the form we still use today: a stand-alone PC equipped with a graphics-display screen and a mouse. A laser printer to print things out. The Ethernet local-area network to tie everyone together. And of course, the user interface that Apple later made famous with the Macintosh computer—the one with windows, icons, menus, scroll bars and all the rest.

Finally, this was the generation, together with the students they taught, who would engineer the personal computer revolution of the 1980s and the networking revolution of the 1990s—more than 50 years after Jay Forrester and his colleagues first began to think about computing in real time. SA

## MORE TO EXPLORE


**Bit by Bit: An Illustrated History of Computers.** Stan Augarten. Ticknor and Fields, 1984.

**A History of Personal Workstations.** Edited by Adele Goldberg. ACM Press History Series. ACM Press, 1988.

**Man-Computer Symbiosis.** J.C.R. Licklider in *In Memoriam: J.C.R. Licklider 1915–1990.* Edited by Robert W. Taylor. Digital Systems Research Center Reports, Vol. 61. Palo Alto, Calif., 1990.

**Digital at Work: Snapshots from the First Thirty-Five Years.** Edited by Jamie Parker Pearson. Digital Press, 1992.

**Transforming Computer Technology: Information Processing for the Pentagon, 1962–1986.** Arthur L. Norberg and Judy E. O'Neill. Johns Hopkins University Press, 1996.

**A History of Modern Computing.** Paul E. Ceruzzi. MIT Press, 1998.

**The Dream Machine: J.C.R. Licklider and the Revolution That Made Computing Personal.** M. Mitchell Waldrop. Viking, 2001.

The Computer Museum History Center Web site is at **www.computerhistory.org/**

ELECTRONIC TOLL COLLECTION

# In the Fast Lane

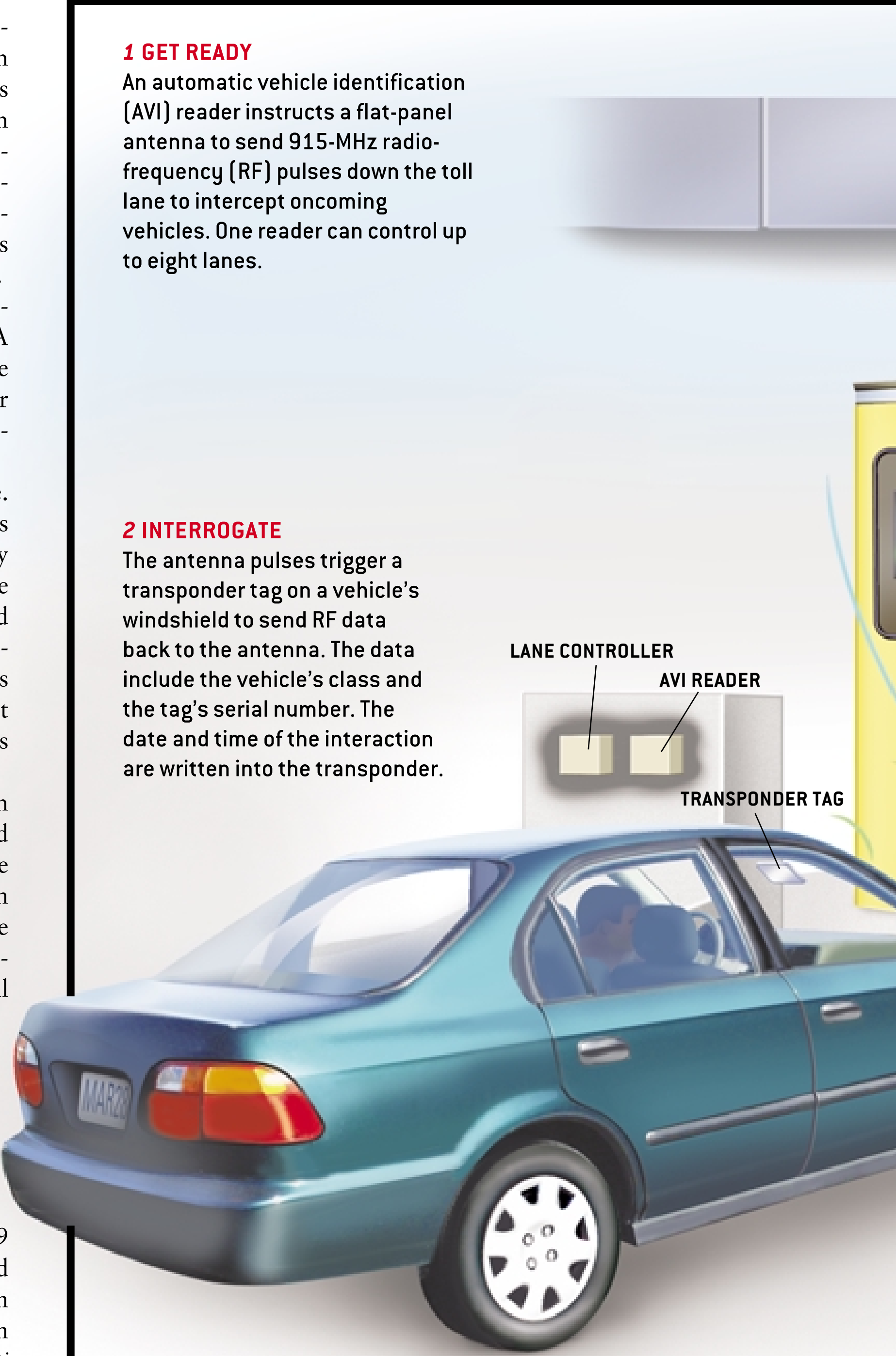
**More than seven million Americans** with electronic "tags" on their car windshields now cruise through deliciously vacant tollbooths while fellow drivers wait in lines. The convenience requires tollbooth equipment that interrogates a vehicle's tag with radio-frequency waves, validates the tag holder's account and deducts the toll from the account's prepaid balance—while imaging the vehicle with lasers and videotaping its license plates to catch cheaters.

All the action can happen in a few seconds because your account is not updated in real time. A database at a given toll plaza debits your balance there, but the network of toll plazas updates your central account only once a day. Violations are processed later, too.

Electronic toll systems have proved accurate. "We have only one error in 10,000 reads," says Walter Kristlibas, who oversees the Port Authority of New York & New Jersey's use of E-ZPass, the nation's largest system. The Port Authority and agencies in five neighboring states from Massachusetts to West Virginia now honor one another's tags, so drivers can traverse the Northeast without stopping. The group is talking with other states about establishing a national network.

Manufacturers are testing future tags into which a driver would insert a "smart" card. Tolls would be deducted from the smart card's balance; there would be no driver account. "That way, no one can track where you're going," says Peter Oomen, vice president of engineering at Mark IV IVHS in Ontario, which makes the E-ZPass tag. And the toll agency saves money because it doesn't have to maintain an account for you. Australia is using a similar scheme, and Europe is testing a second-generation system, Adept II, in several countries.

Radio-frequency technology is also being employed by collision-avoidance systems under development for vehicles. The Federal Communications Commission has allocated a band of spectrum at 5.9 GHz for this task. In a few years a new car could come with a standard radio transponder set, Oomen says. "It could pay your tolls and provide collision avoidance nationwide." —*Mark Fischetti*



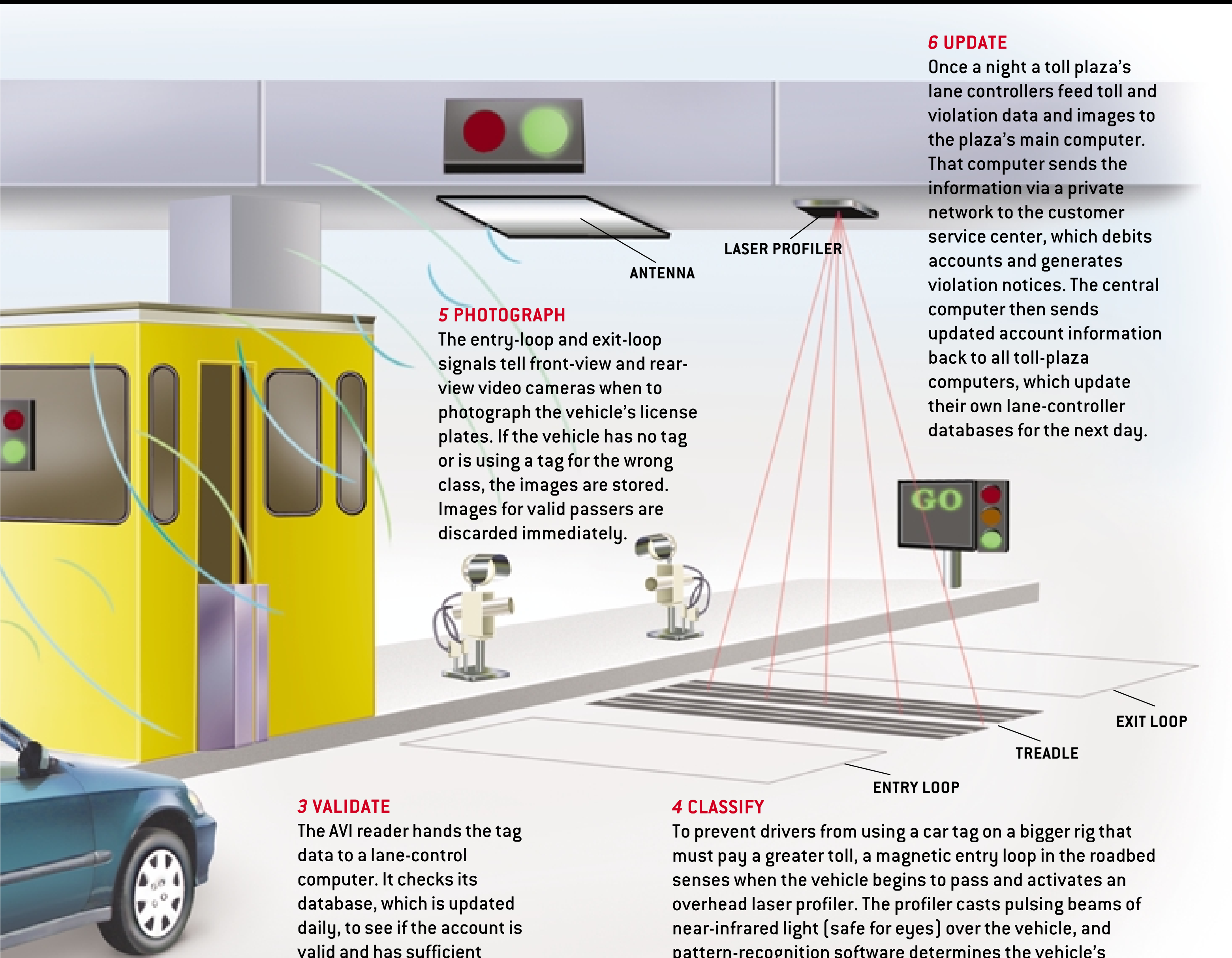
## DID YOU KNOW . . .

➤ NAME GAME: Promoters of electronic toll systems try to capture people's interest with zippy system names. Some work, some don't. You be the judge: FastLane (Massachusetts), E-ZPass (New York and New Jersey), CruiseCard (Georgia), FasTrak (California), K-Tag (Kansas), CityLink (Australia), TelePass (Italy), PASE (Argentina).

➤ 100 MPH: Signs at electronic tollbooths demand that drivers slow to speeds of around 15 mph as they pass. The reason is purely to protect drivers and toll-plaza personnel. "The E-ZPass equipment was designed to work at up to 100 miles per hour," says regional director Walter Kristlibas, "so it could be converted to open-highway toll collection. But I don't recommend you drive through a tollbooth that fast."

➤ FASTER FOOD: Owners of McDonald's burger franchises are testing whether to allow customers in their drive-through lanes to automatically pay for orders with electronic toll-collection tags, bypassing the cashier's window. Field trials have been held with FasTrak in California and E-ZPass in New Jersey.

➤ SEPTEMBER 11: The central computer that maintains the Port Authority of New York & New Jersey's vast E-ZPass system was located in the World Trade Center. In a now eerie statement, E-ZPass's own system description says that communications lines to the host computer can instantly switch to a backup site in New Jersey "even if an entire router at the World Trade Center" were to fail "in the event of a disaster."




**6 UPDATE**
Once a night a toll plaza's lane controllers feed toll and violation data and images to the plaza's main computer. That computer sends the information via a private network to the customer service center, which debits accounts and generates violation notices. The central computer then sends updated account information back to all toll-plaza computers, which update their own lane-controller databases for the next day.

**5 PHOTOGRAPH**
The entry-loop and exit-loop signals tell front-view and rear-view video cameras when to photograph the vehicle's license plates. If the vehicle has no tag or is using a tag for the wrong class, the images are stored. Images for valid passers are discarded immediately.

**3 VALIDATE**
The AVI reader hands the tag data to a lane-control computer. It checks its database, which is updated daily, to see if the account is valid and has sufficient funds. It will turn on a green "go" light or a yellow "account low" light on the driver sign. If the tag is invalid or there is no tag, it will flash a yellow or red "toll unpaid" light.

**4 CLASSIFY**
To prevent drivers from using a car tag on a bigger rig that must pay a greater toll, a magnetic entry loop in the roadbed senses when the vehicle begins to pass and activates an overhead laser profiler. The profiler casts pulsing beams of near-infrared light (safe for eyes) over the vehicle, and pattern-recognition software determines the vehicle's approximate shape. A pressure-sensitive treadle simultaneously counts wheel impacts to determine the number of axles. All the information is fed to the lane controller.

ILLUSTRATION BY MARK SCHROEDER

*Have a topic for a future column?*
*Send it to workingknowledge@sciam.com*

# Long-Distance Robots

THE TECHNOLOGY OF TELEPRESENCE MAKES THE WORLD EVEN SMALLER **BY MARK ALPERT**

**A week after** the World Trade Center disaster, I drove from New York City to Somerville, Mass., to visit the offices of iRobot, one of the country's leading robotics companies. I'd originally planned to fly there, but with the horrific terrorist attacks of September 11 fresh in my mind, I decided it would be prudent to rent a car. As I drove down the Massachusetts Turnpike, gazing at the American flags that hung from nearly every overpass, it seemed quite clear that traveling across the U.S., whether for business or for pleasure, would be more arduous and anxiety-provoking from now on. Coincidentally, this issue was related to the purpose of my trip: I was evaluating a new kind of robot that could allow a travel-weary executive to visit any office in the world without ever leaving his or her own desk.


The technology is called telepresence, and it takes advantage of the vast information-carrying capacity of the Internet. A telepresence robot is typically equipped with a video camera, a microphone, and a wireless transmitter that enables it to send signals to an Internet connection. If a user at a remote location logs on to the right Web page, he or she can see what the robot sees and hear what the robot hears. What's more, the user can move the machine from place to place simply by clicking on the mouse. With the help of artificial-intelligence software and various sensors, telepresence robots can roam down hallways without bumping into walls and even climb flights of stairs.

Until now, businesspeople have relied on techniques such as videoconferencing to participate in meetings that they can't attend. Anyone who's seen a videoconference, though, knows how frustrating the experience can be. Unless the participants are sitting right in front of the camera, it's often difficult to understand what they're saying. Researchers are developing new systems that may make videoconferences more realistic [see "Virtually There," by Jaron Lanier; Scientific American, April 2001]. But there's another problem with videoconferencing: the equipment isn't very mobile. In contrast, a telepresence robot can travel nearly anywhere and train its camera on whatever the user wishes to see. The robot would allow you to observe the activity in a company's warehouse, for example, or to inspect deliveries on the loading dock.

The idea for iRobot's machines originated at the Massachusetts Institute of Technology's Artificial Intelligence Laboratory. Rodney Brooks, the lab's director, co-founded the company in 1990 with M.I.T. graduates Colin Angle and Helen Greiner. iRobot's offices are on the second floor of a nondescript strip mall, just above a store selling children's clothing. It's the kind of office that an eight-year-old would adore—machines that look like miniature tanks lurk in every corner, as if awaiting orders to attack. The robots are tested in a large, high-ceilinged room called the High Bay, which is where I encountered a telepresence robot named Cobalt 2.

The machine resembles a futuristic wheeled animal with a long neck and a bubblelike head. When the robot raised its head to train its camera on me, it looked kind of cute, like a baby giraffe. Angle, who is iRobot's chief executive, says the company designed the machine to appear friendly and unthreatening. "We wanted to create a device that would be easy for people to interact

PHOTOGRAPHS COURTESY OF iROBOT

with," he says. The robot rides on six wheels and has a pair of "flippers" that it can extend forward for climbing stairs. The antenna is fixed to the back of the machine like a short black tail.

After I finished admiring Cobalt 2, I turned to a nearby computer monitor that showed the robot's Web page. In the center of the screen was the video that the robot was transmitting over the Internet. The machine was still staring at me, so I had a nice view of my own backside. The video was grainy and jerky; because the system transmits data at about 300 kilobits per second, the user sees only five or six frames per second (television-quality video shows 30 frames per second). "You're trading off the frame-update rate for the ability to move and control the camera," Angle explains. Transmitting audio over the Internet is more troublesome because of time lags, but users can easily get around this problem by equipping the robot with a cellular phone.

Now I was ready to give Cobalt 2 a road test. Using the mouse, I clicked on the area of the video screen where I wanted the robot to go. The machine's motors whirred loudly as they turned the wheels, first pointing the robot in the right direction and then driving it to the indicated spot. Then I devised a tougher challenge: I directed the machine to smash into the wall on the other side of the room. Fortunately for Cobalt 2, its compact torso is studded with sensors. The machine's acoustic sensor acts like a ship's sonar, detecting obstacles by sending out sound waves and listening to the echoes. Infrared sensors gauge the distance to the obstacles and can also warn the robot if it's heading toward a drop-off. Cobalt 2 stopped just shy of the wall, thwarting my destructive intentions.

The machine that iRobot plans to sell to businesses looks a little different from Cobalt 2. Called the CoWorker, it resembles a small bulldozer—it actually has a shovel for pushing objects out of its path. "It's a robot with a hard hat," Angle says. In addition to a video camera, the machine has a laser pointer and a robotic arm that remote users can manipulate. iRobot has not set a price for the CoWorker yet, but it is already shipping prototype versions to businesses that want to evaluate the technology. The company also plans to introduce a telepresence robot for home use. Such a device could be a lifeline for senior citizens living alone; the robot would allow nurses and relatives to see whether an elderly person is ill or needs immediate help.

TELEPRESENCE ROBOT called the Packbot is designed to do reconnaissance in dangerous environments. iRobot, a company based in Somerville, Mass., has built other mobile machines that can transmit video over the Internet (*opposite page*).

Will these mechanical avatars soon be knocking on your door? The fundamental challenge of telepresence is not technological but psychological: I, for one, would have a lot of trouble keeping a straight face if a robot sat next to me at one of our magazine's staff meetings. And can you imagine how most senior citizens would react to the wheeled contraptions? Nevertheless, people may eventually accept the technology if the potential benefits are great enough. For example, an elderly person may decide to tolerate the intrusions of a camera-wielding robot if the only safe alternative is living in a nursing home.

As I wandered through iRobot's offices, I got a glimpse of another telepresence robot called the Packbot. About the size of a small suitcase, this low-slung machine moves on caterpillar treads and, like Cobalt 2, has extendable flippers that allow it to climb over obstacles. The Defense Advanced Research Projects Agency (DARPA)—the U.S. military's research and development arm—is funding the development of the Packbot, which is designed to do reconnaissance and surveillance in environments where it would not be safe for humans to go.

In the aftermath of the September 11 attacks, military officials recognized that telepresence robots could aid the search-and-rescue efforts. So the engineers at iRobot attached video and infrared cameras to the prototype Packbots and rushed them to New York. At the Somerville office I watched an engineer fasten two flashlights and a camera to a Packbot that would soon be taken to the World Trade Center site.

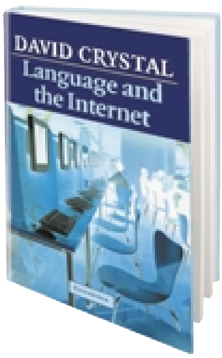
Although the Packbots were too large to burrow into the wreckage, the iRobot engineers used one machine to search a parking garage. Smaller telepresence robots called the MicroTrac and the MicroVGTV—machines made by Inuktun, a Canadian company that sells robots for inspecting pipes and ducts—were able to crawl through the holes in the rubble. The machines found no survivors but located the bodies of several victims.

This grim task was perhaps the best demonstration of the value of telepresence. As I drove back to New York, I felt a grudging respect for the robots—and for the men and women who'd built them. SA

# REVIEWS

# Spontaneous, Unedited, Naked

A LINGUIST LOOKS AT DISCOURSE ON THE INTERNET **BY ANNE EISENBERG**




**LANGUAGE AND THE INTERNET**
by David Crystal
Cambridge University Press, 2001 ($19.95)

**Never mind** those anxieties about the Internet's impact on privacy, intellectual property and the recreational habits of 12-year-olds. What is it doing to the future of the English language? Will it really lead to the end of literacy as we know it—not to mention spelling?

Not according to David Crystal, a linguist who says in this witty, thoughtful book that, on the contrary, the discourse of the Internet—with its new, informal, even bizarre forms of language—neither threatens nor replaces existing varieties of English but instead enriches them, extending our range of expression and showing us "*homo loquens* at its best."

Crystal, the Welsh author of the *Cambridge Encyclopedia of the English Language* who is known to many in the U.S. through his comments on National Public Radio, analyzes the discourse of Web pages, e-mail, chatgroups and virtual-reality games. At first glance, much of this text certainly looks like a primer on linguistic irresponsibility: the shedding of capital letters; the minimalist punctuation; the perverse spellings and goofy abbreviations like RUOK ("are you okay?"); the smileys, such as :-), representing humor; the coining of terms at a rate that has no parallel in contemporary language.

For Crystal, though, these phenomena are not portents of linguistic doom but examples of a set of language tactics developed for a new medium he calls computer-mediated communication. The innovative, sometimes screwball varieties of English expressed in computer-mediated channels, he says, have evolved as users have adapted their language creatively to meet changing circumstances.

Smileys, for instance, appeared early in the language of e-mail as people struggled to replace many characteristics of speech, like pitch and tone, with symbols, using ;-) for winking or :-( for sadness. Most other forms of written language suffer under the same burden as e-mail, of course—they are not face-to-face and are therefore always ambiguous in their omission of cues such as intonation. So why are there no smileys in other forms of writing? Crystal argues that the answer lies in the immediacy of computer-mediated communication. Traditional writing entails time to revise, to make personal attitudes clear, to tinker with phrases. Smileys and other, related devices stand in for this extra work in the more spontaneous, fluid world of the new medium, which combines properties not only of speaking and writing but of rapid electronic exchange.

Crystal is unbothered by typical usage issues—for instance, whether the form "email," "e-mail" or "E-mail" will prevail. He's willing to leave such matters to a future editorial consensus. And he does not worry about whether using "Dear Bob" instead of "Bob" at the beginning of an e-mail will make him a fuddy-duddy, as one handbook on e-mail usage advises. In fact, Crystal laughs at this prescriptive approach, arguing that

## A Guide to Netspeak

BOTH UPPERCASE AND LOWERCASE FORMS ARE USED.

| | | | |
|---|---|---|---|
| afaik | as far as I know | mtfbwu | may the force be with you |
| awhfy | are we having fun yet? | obtw | oh, by the way |
| b4 | before | rotf | rolling on the floor |
| bg | big grin | rtfm | read the f---ing manual |
| cm | call me | smtoe | sets my teeth on edge |
| dur? | do you remember? | t+ | think positive |
| fwiw | for what it's worth | tttt | to tell the truth |
| gal | get a life | tx | thanks |
| gmta | great minds think alike | wb | welcome back |
| ianal | I'm not a lawyer, but. . . | X! | typical woman |
| icwum | I see what you mean | Y! | typical man |
| imo | in my opinion | 2bctnd | to be continued |
| j4f | just for fun | 2g4u | too good for you |
| jk | just kidding | 4yeo | for your eyes only |

EXCERPTED FROM TABLE 3.2 IN *LANGUAGE AND THE INTERNET* to condemn one style as bad is to deny English users the stylistic option of switching, thereby reducing the versatility and richness of language. No single recommendation, he says, can suit the expectations of the range of audiences the Internet is reaching.

His interest, instead, is in the readiness with which people are adapting spelling, grammar and semantics to meet the needs of Internet-based situations. The chapters on specific adaptations are studded with linguistic delights to satisfy anyone who has ever wondered what TTFN means ("ta ta for now") or tia ("thanks in advance") or gal ("get a life"). (Many more of these abbreviations are explained in highly entertaining tables, as are the varieties of smileys.) He tackles etymologies, too, and the derivations shed light on much that may otherwise have been mysterious: cc, for example, has a new gloss as "complimentary copy," now that carbon copies are a distant memory. He examines the plural ending "-en" that is popular on the Internet—as in "vaxen" for VAX computers—saying that such suffixes are a development that "will cause delight to all Anglo-Saxonists."

Crystal devotes a chapter to the discourse of chatgroups—"gossip groups" is a more accurate description for most of what goes on within them, he says—which he characterizes as a "perpetual linguistic party, where you bring your language, not a bottle." He is fascinated by chatgroup language in part because it provides a domain in which to see written language in its most primitive state—banal, repetitive and untouched (as most writing is) by editing. "Chatgroups are the nearest we are likely to get to seeing writing in its spontaneous, unedited, naked state."

He also reports on the scholarly literature of computer-mediated communication, including such gems as the finding that, in contrast to females, males on academic newslists sent longer messages, made stronger assertions and engaged in more self-promotion, while making fewer apologies and asking fewer questions.

Crystal is definitely upbeat, discovering the still evolving discourse of the Internet an area of huge potential enrichment. He uses the analogy of a gift he received—a new informal shirt. This shirt didn't destroy his sense of the value of formal and informal—it just made his previously satisfactory, informal shirts look somewhat staid. He sees the language of the Internet, too, as similarly extending the range of communication options. RUOK with this? SA

***Editors' note:*** *Among the many concerns we had on September 11 was December's book reviewer Anne Eisenberg. We knew that Eisenberg, who writes for the "Circuits" section of the* New York Times *and teaches writing at the Polytechnic University in Brooklyn, lives a few short blocks from the World Trade Center. We discovered that she was safe: she had dashed from her apartment, into a cloud of black smoke and debris, just as the second tower began to pancake. We asked her what she took: "Money, passport, a sweater and David Crystal's book, because it was my next assignment."*

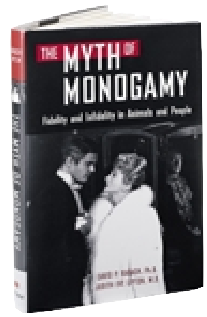


NASA

## THE EDITORS **RECOMMEND**

**THE MYTH OF MONOGAMY: FIDELITY AND INFIDELITY IN ANIMALS AND PEOPLE**

***by David P. Barash and Judith Eve Lipton. W. H. Freeman and Company, New York, 2001 ($24.95)***

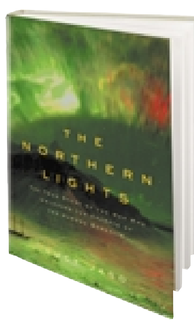



Monogamists, this husband-wife team says, "are going against some of the deepest-seated evolutionary inclinations with which biology has endowed most creatures, *Homo sapiens* included." Barash, professor of psychology at the University of Washington, and Lipton, a psychiatrist, note how rare monogamy is in the animal kingdom. One could not have been so sure about humans until the advent of DNA fingerprinting, which makes it possible to "specify, with certainty, whether a particular individual is or is not the parent." And a "key point" is that women as well as men stray from monogamous relationships. The argument leads one inevitably to ask why monogamy exists at all and why human societies show such concern about it. Barash and Lipton suggest that it may occur as a means for males to minimize the risk "that someone else's sperm will fertilize the eggs of a given female" and that society's many strictures against adultery arise because monogamy is not automatic "but needs to be enforced and reinforced."

---

**THE NORTHERN LIGHTS**

***by Lucy Jago. Alfred A. Knopf, New York, 2001 ($24)***

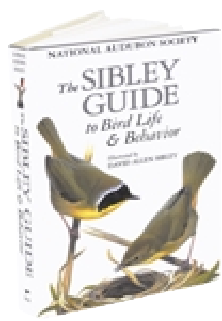



In *The Northern Lights,* Jago uncovers a subject that has been all but buried: the true story of Kristian Birkeland, a man more than half a century ahead of his time in his scientific pursuits. The book details the life of the Norwegian scientist as he struggles, at the turn of the 19th century, to solidify his theories about the aurora borealis, or northern lights. (This luminous phenomenon of the upper atmosphere occurs in the Northern Hemisphere; in the Southern Hemisphere it is known as the aurora australis, or southern lights.) Jago, a former producer for the BBC, deftly paints a historical background for some of the most important concepts in electromagnetic theory today, breathing life into a subject traditionally presented with a drab countenance.

---

**THE SIBLEY GUIDE TO BIRD LIFE & BEHAVIOR**

***Illustrated by David Allen Sibley. Edited by Chris Elphick, John B. Dunning, Jr., and David Allen Sibley. Alfred A. Knopf, New York, 2001 ($45)***




One year after publication of the widely praised *Sibley Guide to Birds* (reviewed here in January) comes this companion volume, which tells how birds live and what they do. Readers will learn about feathers, feet and flight dynamics; food, foraging and courtship; breeding, migration and conservation threats. The text-to-illustration ratio is, quite logically, much larger in this new volume, and the text is a powerhouse of information compiled by Sibley and his co-editors from 48 leading birders and biologists. The 796 small, enchanting, full-color paintings are by Sibley.

*All the books reviewed are available for purchase through www.sciam.com*

# Fashion Gang

BY DENNIS E. SHASHA

**A group of mathematicians** who happen to be teenage girls decide to form a fashion gang. The rules of the gang are that every day each girl must wear a tank top that is either blue or black; sunglasses whose rims are either black or brown; capri pants that are black, red, white or pink; and lipstick that is pink, red or brown. Each pair of girls must differ in at least two of these items. For example, if they wear the same color tank top and lipstick, then they must differ in their choice of sunglass rims and capris. Differing in more than two items is also acceptable.

There are two challenges: What is the largest number of girls who could be in this gang, and what might each girl wear in that case? What is the minimum number of girls that could be in this gang so that they satisfy the difference constraint but adding one girl would violate the constraint? Again, show a possible outfit.

Here is a warm-up: Suppose there are just three attributes and they are all binary: a tank top that is blue or black, sunglass rims that are black or brown, and capris that are either black or red. Try to find maximal gangs as small as two and as large as four. SA

**Answer to Last Month's Puzzle**

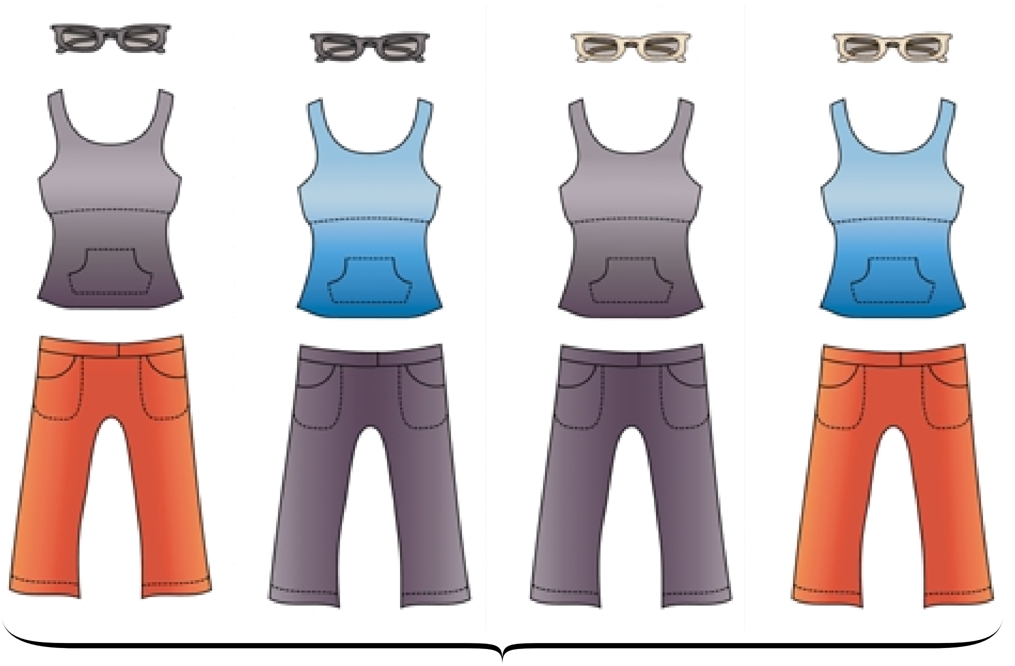
The delivery trucks can reach their destinations in six minutes using the routing schedule shown in detail at www.sciam.com

This is in fact a minimum-time solution. Because 11 traversals across BC are necessary and only two are possible in each minute, no solution can do better than six minutes.

**Web Solution**

For a peek at the answer to this month's problem, visit www.sciam.com

## WARM-UP PROBLEMS



LARGEST GANG

SMALLEST GANG

SARA CHEN

# The Importance of Being Ernst

THOUGHTS ON AN EVOLUTION EXPERT WHO HAS TAKEN SERIOUSLY THE WARNING "PUBLISH OR PERISH" BY STEVE MIRSKY

**In 1928 a young man** named Ernst Mayr performed the first survey of the birds of the Cyclops Mountains, a forbidding region of New Guinea. The 23-year-old got malaria. He got dengue. He got dysentery. He experienced a trip down a waterfall. He was said to have been killed by natives, but the reports of his death were exaggerated. (The Twainian expression usually describes a death report as "greatly exaggerated," but Mayr came close enough to dying to warrant the deletion of the qualifier.)

A mere 73 years later Mayr, who went on to become the great evolutionary biologist of the 20th century (and the 21st so far), has a new book out. The now 97-year-old Mayr found time amid the organizing of his lifetime achievement awards to write *What Evolution Is,* published in October by Basic Books.

Having Ernst Mayr still among us, and still publishing for us, brings to mind the words of science historian Gerald Holton, who, with tongue in cheek, somehow still clearly said, "In the sciences, we are now uniquely privileged to sit side by side with the giants on whose shoulders we stand."

Mayr's latest book extends a publishing streak that began with his first book in 1942. In his preface, Mayr describes the work as being intended for three groups of readers: anyone interested in evolution; anyone who accepts evolution but is doubtful about the Darwinian particulars; and finally, "creationists who want to know more about the current paradigm of evolutionary science, if for no other reason than to be able to better argue against it." (If you're looking for the difference between a scientific attitude and a dogmatic one, it can pretty much be found in that last sentiment.)

Mayr would still be having a fairly prolific 2001 were he half, or a third, or, what the heck, a quarter of his age. In addition to the new book, Mayr's article "Darwin's Influence on Modern Thought," which first appeared in the July 2000 issue of *Scientific American,* was chosen for inclusion in *The Best American Science Writing 2001*. I had the privilege of meeting Mayr when he


was preparing the article for us, and I asked him to sign my copy of his 1991 book on Darwinism, *One Long Argument*. His daughter, who accompanied him, noted that the work was her favorite of his, for its simple, direct style. Her statement prompted him to say, "My books are so straightforward that the *New York Review of Books* has never been interested in them." (Nonreaders of the *New York Review* may find that comment merely wry; readers of the *New York Review* may find it hilarious.)

Mayr's *Scientific American* article included a sentence that motivated quite a bit of reader response. That sentence read, "No educated person any longer questions the validity of the so-called theory of evolution, which we now know to be a simple fact." (And the quill pens are now once again angrily hovering above the parchment.)

A few of the letters we received noted that Mayr was astoundingly wrong, as the letter writer didn't buy evolution despite being quite well educated, the proof of which was the correspondent's degree, or multiple degrees, from a fine, perhaps even Ivy League, institution. To such readers, I belatedly recommend perhaps the most cogent commentary on the relationship between degrees and education, as delivered to the scarecrow in the movie *The Wizard of Oz:*

"Why, anybody can have a brain. That's a very mediocre commodity. Every pusillanimous creature that crawls on the earth or slinks through slimy seas has a brain. Back where I come from, we have universities, seats of great learning where men go to become great thinkers. And when they come out, they think deep thoughts, and with no more brains than you have! But they have one thing you haven't got: a diploma. Therefore, by virtue of the authority vested in me by the Universitatus Committeatum E Pluribus Unum, I hereby confer upon you the honorary degree of Th.D. … that's Doctor of Thinkology."

Toto, even the school board in Kansas has seen the light on evolution. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

# ANNUAL INDEX 2001

## AUTHORS

### ARTICLES

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION (required by 39 U.S.C. 3685). *1.* Publication title: Scientific American. *2.* Publication number: 509-530. *3.* Filing date: September 10, 2001. *4.* Issue frequency: monthly. *5.* Number of issues published annually: 12. *6.* Annual subscription price: U.S. and its possessions, 1 year, $34.97; Canada, 1 year, $49; all other countries, 1 year, $55. *7.* Complete mailing address of known office of publication: 415 Madison Avenue, New York, NY 10017-1111. *8.* Complete mailing address of the headquarters or general business office of the publisher: 415 Madison Avenue, New York, NY 10017-1111. *9.* Full names and complete mailing address of publisher, editor and managing editor: Publisher, Gretchen Teichgraeber, 415 Madison Avenue, New York, NY 10017-1111. Editor, John Rennie, 415 Madison Avenue, New York, NY 10017-1111. Managing Editor, Michelle Press, 415 Madison Avenue, New York, NY 10017-1111. *10.* Owner: Scientific American, Inc., 415 Madison Avenue, New York, NY 10017-1111; Holtzbrinck Publishing Holdings Limited Partnership, 123 West 18th Street, 8th Floor, New York, NY 10011: (a) Holtzbrinck Publishing Group, Inc. (General Partner), 100 West 10th Street, Wilmington, DE; (b) Georg von Holtzbrinck GmbH & Co. (Limited Partner), Gaensheidestrasse 26, 70184 Stuttgart, Germany. *11.* Known bondholders, mortgagees and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages or other securities: none. *12.* Tax status: not applicable. *13.* Publication title: Scientific American. *14.* Issue date for circulation data below: September 2001. *15.* Extent and nature of circulation: *a.* Total number of copies (net press run): average number of copies each issue during preceding 12 months, 934,585; number of copies of single issue published nearest to filing date, 1,003,000. *b.* Paid and/or requested circulation: (*1*) Paid/requested outside-county mail subscriptions stated on Form 3541 (include advertiser's proof and exchange copies): average number of copies each issue during preceding 12 months, 545,677; number of copies of single issue published nearest to filing date, 567,707. (*2*) Paid in-county subscriptions stated on Form 3541 (include advertiser's proof and exchange copies): average number of copies each issue during preceding 12 months, 0; number of copies of single issue published nearest to filing date, 0. (*3*) Sales through dealers and carriers, street vendors, counter sales and other non-USPS paid distribution: average number of copies each issue during preceding 12 months, 156,978; number of copies of single issue published nearest to filing date, 186,000. (*4*) Other classes mailed through USPS: average number of copies each issue during preceding 12 months, 0; number of copies of single issue published nearest to filing date, 0. *c.* Total paid and/or requested circulation (sum of *15b* (*1*), (*2*), (*3*) and (*4*)): average number of copies each issue during preceding 12 months, 702,655; number of copies of single issue published nearest to filing date, 753,707. *d.* Free distribution by mail (samples, complimentary and other free): (*1*) Outside-county as stated on Form 3541: average number of copies each issue during preceding 12 months, 0; number of copies of single issue published nearest to filing date, 0. (*2*) In-county as stated on Form 3541: average number of copies each issue during preceding 12 months, 0; number of copies of single issue published nearest to filing date, 0. (*3*) Other classes mailed through the USPS: average number of copies each issue during preceding 12 months, 19,670; number of copies of single issue published nearest to filing date, 44,616. *e.* Free distribution outside the mail (carriers or other means): average number of copies each issue during preceding 12 months, 1,000; number of copies of single issue published nearest to filing date, 1,000. *f.* Total free distribution (sum of *15d* and *15e*): average number of copies each issue during preceding 12 months, 20,670; number of copies of single issue published nearest to filing date, 45,616. *g.* Total distribution (sum of *15c* and *15f*): average number of copies each issue during preceding 12 months, 723,325; number of copies of single issue published nearest to filing date, 799,323. *h.* Copies not distributed: average number of copies each issue during preceding 12 months, 211,260; number of copies of single issue published nearest to filing date, 203,677. *i.* Total (sum of *15g* and *15h*): average number of copies each issue during preceding 12 months, 934,585; number of copies of single issue published nearest to filing date, 1,003,000. Percent paid and/or requested circulation (*15c/15g* × 100): average number of each issue during preceding 12 months, 97.14%; number of single issue published nearest to filing date, 94.29%. *16.* Publication of statement of ownership is required. Will be printed in the December 2001 issue of this publication. *17.* I certify that all information furnished above is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties). Signature and title of Editor, Publisher, Business Manager or Owner: (signed) Gretchen Teichgraeber, President and CEO. Date: September 10, 2001.

# END POINTS

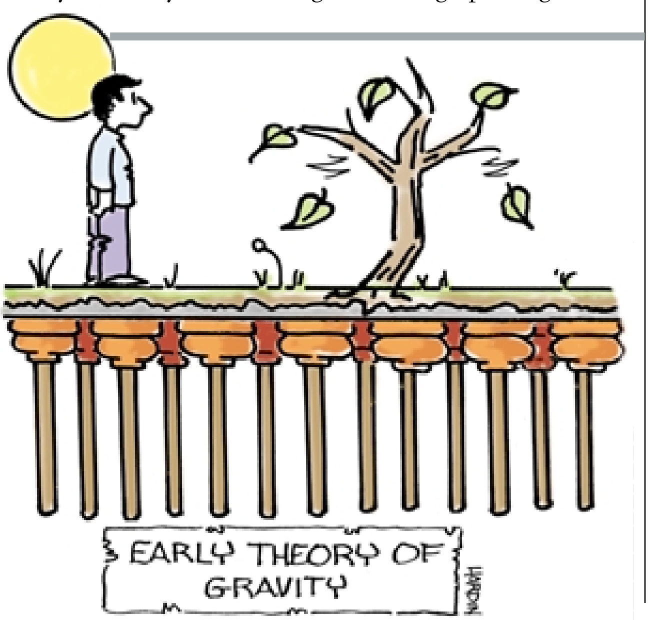
## What happens when lightning strikes an airplane?

—J. McGill, Tampa, Fla.

**Edward J. Rupke, senior engineer at Lightning Technologies in Pittsfield, Mass., provides the following explanation:**

On average, each airplane in the U.S. commercial fleet is struck lightly by lightning more than once a year. In fact, aircraft often trigger lightning when flying through a heavily charged region of a cloud. Business and private airplanes are thought to be struck less frequently because of their smaller size and because they often can avoid weather conducive to lightning strikes.

The last confirmed commercial plane crash in the U.S. directly attributed to lightning occurred in 1967. Today airplanes receive a rigorous set of lightning certification tests to verify the safety of their designs. Although passengers and crew may see a flash and hear a loud noise if lightning strikes their plane, nothing serious should happen, because of the careful protection engineered into the aircraft.




Most aircraft skins consist primarily of aluminum, which conducts electricity very well. By guaranteeing that no gaps exist in this conductive path, the engineer can ensure that most of the lightning current will remain on the exterior of the aircraft. Some modern aircraft are made of advanced composite materials, which contain an embedded layer of conductive fibers or screens designed to carry lightning currents. The engineer must ascertain that no damaging electrical surges or transients can reach the sensitive equipment inside the aircraft. Careful shielding, grounding and surge-suppression devices avert problems caused by indirect effects in cables and equipment. Every circuit and piece of equipment that is critical to the safe flight and landing of an aircraft must be verified to be protected against lightning in accordance with regulations set by the Federal Aviation Administration or a similar authority in the country of the aircraft's origin.

The other main area of concern is the fuel system, where even a tiny spark could be disastrous. Engineers ensure that lightning currents cannot cause sparks in any part of an aircraft's fuel system. The aircraft skin around the fuel tanks must be thick enough to withstand a burn-through. All the pipes and fuel lines that carry fuel to the engines, and the engines themselves, must also be protected against lightning. In addition, new fuels that produce less explosive vapors are now widely used. SA

*For the complete text of this and many other answers, visit Ask the Experts (www.sciam.com/askexpert).*

PATRICK HARDIN

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 2001 by Scientific American, Inc. All rights reserved. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97, Canada $49, International $55. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. **Reprints available:** write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; (212) 451-8877; fax: (212) 355-0408 or send e-mail to sacust@sciam.com **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.** Printed in U.S.A.