7,000,000-YEAR-OLD SKULL: ANCESTOR? APE? OR DEAD END?

SCIENTIFIC AMERICAN

The Nose-Tickling Science of Bubbly

JANUARY 2003 $4.95
WWW.SCIAM.COM

Micromachines Rewrite the Future of Data Storage

The NANODRIVE

PREDICTING EARTHQUAKES

FIGHTING CANCER WITH LIGHT

THE GOVERNMENT'S FLAWED DIET ADVICE

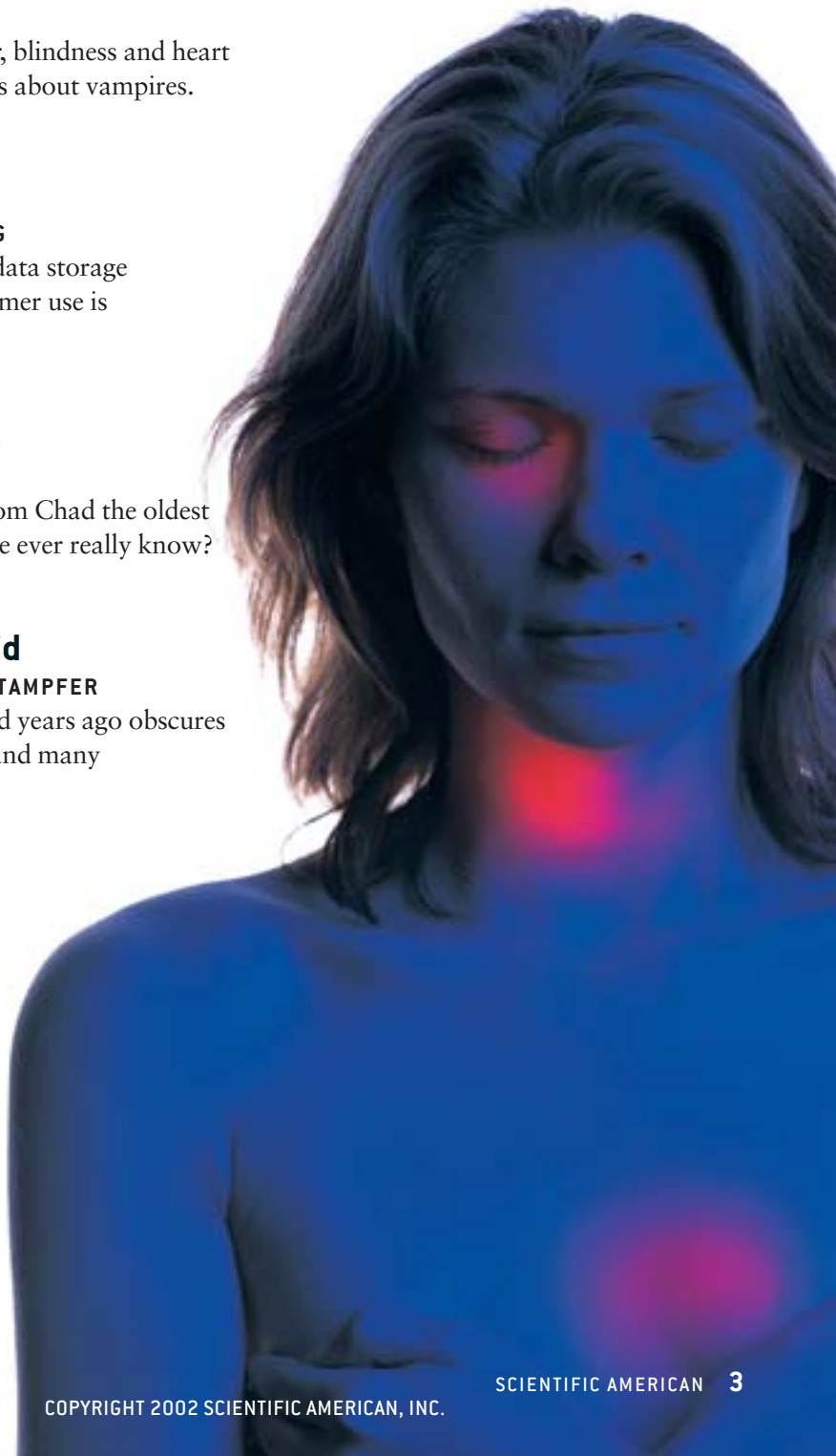
january 2003

contents

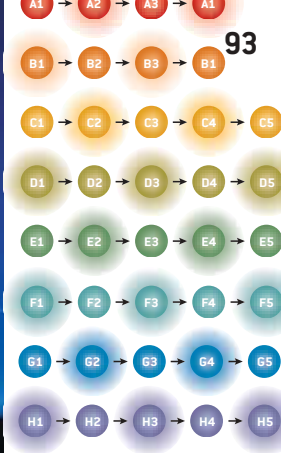
features

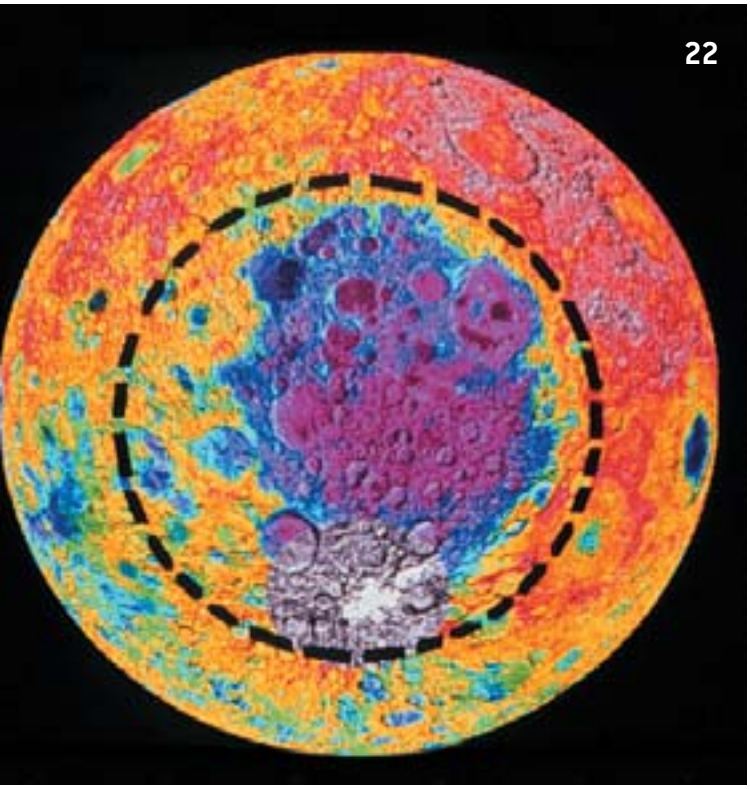
SCIENTIFIC AMERICAN Volume 288 Number 1

38 Light-activated therapies



departments





columns

Cover image by Slim Films

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. 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.

Self and Circuitry

Walk down the average American street, and you won't pass many adults who are 100 percent human anymore—at least not physically. Our mouths are studded with dental fillings, posts, crowns and bridgework. More than a few of us have surgical screws, pins and staples holding together fresh or old injuries. Hordes have pacemakers, artificial joints, breast implants and other internal medical devices.

It's likely that if you asked people about merging their living tissue with unliving parts, they would rate the idea as at best odd and at worst horrifying. That so many of us have calmly done so can be attributed to two considerations. First, most of these implants—such as dental fillings—have been minor, minimally invasive and reassuringly simple. Second, pacemakers and other sophisticated implants are medically mandated—we accept them because they save our lives.

IMPLANTABLE VeriChip

This past year an exception to those rules quietly emerged. Applied Digital Solutions in Palm Beach, Fla., introduced its VeriChip, an implantable device the size of a grain of rice that fits under the skin. When a handheld scanner prods it with radio waves, the chip answers with a short burst of identifying data. The immediate applications are for security and identification.

The utility of implantable chips will only grow as they acquire more processing capability, allowing functions such as geolocation. It's easy to picture implantable chips developed for communications, entertainment and even cosmetic purposes. And one regulatory hurdle has already been removed: the Food and Drug Administration has ruled that as long as the current VeriChip does not serve a medical purpose, it will not be regulated as a medical device. (Future devices that broadcast with more power, however, might be subject to safety review. See the news story by David Appell on page 18.)

The new chips come ready-made with controversies. Should sexual offenders or other felons be tagged for permanent identification? What about resident aliens? Could employers require their workers to be implanted? Might laudable applications, such as preventing kidnapping, lead to civil-rights abuses?

All good questions about uses and misuses of the technology, but here's a more fundamental one: Why is there so little uproar over the underlying concept of putting complex microcircuitry into people? This implant isn't an inert dental filling or a lifesaving therapeutic. It's an electronic ID badge stuck permanently inside the body. A couple decades ago a product fitting that description might have been denounced as the first step toward Orwellian mind control and one-world government.

Ah, but 1984 has come and gone. Electronic devices, including ones that track our location, are now commonplace personal accessories. Movies and television have fed us images of friendly robots and cyborgs. The widespread popularity and casual acceptance of cosmetic surgery, body art and ornamental piercings show that the idea of altering the body has become less taboo. If the VeriChip is a landmark social development, then it is one that we've reached by small steps. New devices work their way into our bodies much as they work their way into the rest of our lives—by offering a sensible value. Almost without our realizing it, the merger of human and machine is becoming more routine. Technology gets under our skin in every sense.

THE EDITORS editors@sciam.com

GREG WHITESELL

How to Contact Us

EDITORIAL

For Letters to the Editors:
Letters to the Editors
Scientific American
415 Madison Ave.
New York, NY 10017-1111
or
editors@sciam.com
Please include your name and mailing address, and cite the article and the issue in which it appeared. Letters may be edited for length and clarity. We regret that we cannot answer all correspondence.

For general inquiries:
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-754-0550
fax: 212-755-1976
or
editors@sciam.com

SUBSCRIPTIONS

For new subscriptions, renewals, gifts, payments, and changes of address:
U.S. and Canada
800-333-1199
Outside North America
515-247-7631
or
www.sciam.com
or
Scientific American
Box 3187
Harlan, IA 51537

REPRINTS

To order reprints of articles:
Reprint Department
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-451-8877
fax: 212-355-0408
reprints@sciam.com

PERMISSIONS

For permission to copy or reuse material from *SA*:
permissions@sciam.com
or
212-451-8546 for procedures
or
Permissions Department
Scientific American
415 Madison Ave.
New York, NY 10017-1111
Please allow three to six weeks for processing.

ADVERTISING

www.sciam.com has electronic contact information for sales representatives of Scientific American in all regions of the U.S. and in other countries.

New York
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-451-8893
fax: 212-754-1138
Los Angeles
310-234-2699
fax: 310-234-2670
San Francisco
415-403-9030
fax: 415-403-9033
Midwest
Derr Media Group
847-615-1921
fax: 847-735-1457
Southeast/Southwest
MancheeMedia
972-662-2503
fax: 972-662-2577
Detroit
Karen Teegarden & Associates
248-642-1773
fax: 248-642-6138
Canada
Fenn Company, Inc.
905-833-6200
fax: 905-833-2116
U.K.
The Powers Turner Group
+44-207-592-8331
fax: +44-207-630-9922
France and Switzerland
PEM-PEMA
+33-1-46-37-2117
fax: +33-1-47-38-6329
Germany
Publicitas Germany GmbH
+49-211-862-092-0
fax: +49-211-862-092-21
Sweden
Publicitas Nordic AB
+46-8-442-7050
fax: +49-8-442-7059
Belgium
Publicitas Media S.A.
+32-(0)2-639-8420
fax: +32-(0)2-639-8430
Middle East and India
Peter Smith Media & Marketing
+44-140-484-1321
fax: +44-140-484-1320
Japan
Pacific Business, Inc.
+813-3661-6138
fax: +813-3661-6139
Korea
Biscom, Inc.
+822-739-7840
fax: +822-732-3662
Hong Kong
Hutton Media Limited
+852-2528-9135
fax: +852-2528-9281

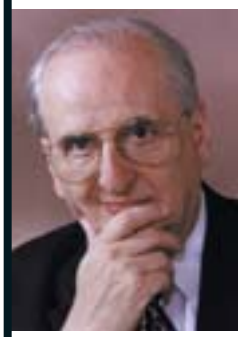
On the Web

WWW.SCIENTIFICAMERICAN.COM

FEATURED THIS MONTH

Visit **www.sciam.com/explore_directory.cfm** to find these recent additions to the site:

Quiet Celebrity:
An Interview with Judah Folkman



The life of Judah Folkman took an unexpected turn one morning in May 1998. That day a front-page article in the *New York Times* announced that Folkman, a professor at Harvard Medical School, had discovered two natural compounds, angiostatin and endostatin, that dramatically shrank tumors in mice by cutting off the cancer's blood supply. The story included a quote from Nobel laureate James D. Watson: "Judah is going to cure cancer in two years." Watson eventually backed off from that assertion, but the media frenzy had already exploded worldwide, transforming Folkman into a reluctant hero in the fight against cancer.

Folkman's ideas, which were initially met with skepticism by oncologists, are now the basis for an area of research that is attracting enormous interest. At least 20 compounds with an effect on angiogenesis are being tested in humans for a range of pathologies, including cancer, heart disease and vision loss. But the premature hype continues to engender disproportionate hope in the public, the press and the stock market. In this interview with SCIENTIFIC AMERICAN, Folkman talks about his work and the progress of clinical trials on endostatin and angiostatin.

Ask the Experts

How does one determine the exact number of cycles a cesium 133 atom makes in order to define one second?

Physicist Donald B. Sullivan, chief of the time and frequency division of the National Institute of Standards and Technology, explains.
www.sciam.com/askexpert_directory.cfm

FREE E-NEWSLETTERS

Stay current on the latest news in science and technology with FREE weekly e-newsletters delivered straight to your mailbox.

SIGN UP TODAY:
www.sciam.com/newsletter_signup.cfm?saletter=9

PLUS:

DAILY NEWS ■ DAILY TRIVIA ■ WEEKLY POLLS

COURTESY OF HARVARD MEDICAL SCHOOL

SCIENTIFIC AMERICAN®
Established 1845

Letters

EDITORS@SCIAM.COM

IS IT ANY WONDER that "A Matter of Time," the September 2002 single-topic issue, brought out the pensive side of *Scientific American*'s readers? Letter writers reflected, often at great length, on the mysteries of time. "We presume to break time up into little units when we define hours, seconds and nanoseconds," wrote Pete Boardman of Groton, N.Y. "But time is not an object to be divided or a substance that moves. Time is the measuring stick, the ruler, the clock. It is earth rotating on its axis. It is earth orbiting around the sun. It is sand flowing through a narrow hole in an hourglass, the repetitive swing of a pendulum, the decay of cesium atoms." Some even turned to poetry to express their reactions, such as the first of the other musings that await on the following two pages, for those who care to take the time.

EVERYTHING AT ONCE

In your enjoyable issue, I was particularly fascinated by the description of one theory, which holds that everything may actually be happening at once ["That Mysterious Flow," by Paul Davies]. I described this notion to my colleague Joe A. Oppenheimer of the University of Maryland, and he referred me to a poem by T. S. Eliot, "Burnt Norton," which begins:

> Time present and time past
> Are both perhaps present in time future,
> And time future contained in time past.
> If all time is eternally present
> All time is unredeemable.
> What might have been is an abstraction
> Remaining a perpetual possibility
> Only in a world of speculation.

If the theory is eventually accepted, this may be a rather spectacular example of life imitating art.

Norman Frohlich
I. H. Asper School of Business
University of Manitoba

Faced with the unintuitive outcomes of time as defined by Einstein a century ago, I have found that it makes sense to think of motion as the more fundamental quantity than time. The common physics equation velocity = distance/time would be better written, I submit, as time = distance/velocity. The implication is that time is a derived (man-made) quantity that is the ratio of these two fundamentals.

With this adjustment, many phenomena become more intuitive. While it seems strange to think of time slowing in the presence of a strong gravity field (general relativity), it is much easier to think of molecules slowing under the same conditions. Time travel also becomes easier to evaluate: because there is no time, there is no place to travel to.

Andy Hanson
Glen Rock, Pa.

While I read your articles, I alternated between being extremely frustrated and being fascinated. Why should an entity so common and so precious be so maddeningly elusive to understand in scientific terms? In our ordinary living, we all clearly understand the unidirectionality of time. Likewise, the field of engineering is based on spatially varying and rate-dependent phenomena. Is it only theoretical physics and quantum theories that have a problem defining time? Finally, there must be profound spiritual content in our contemplation of time. How else could we embrace the notion of "always was and always will be" and eternity?

Charles E. Harris
NASA Langley Research Center
Hampton, Va.

TIME FOR PHILOSOPHY

Philosophy can be useful to the understanding of physics for the same reason that science scholars often shun the subject. Namely, physics deals with exacti-

tudes, while philosophy is based on a preponderance of available evidence. So, whereas an entire theory in physics can be invalidated by as little as a single erroneous digit, it is much harder to totally discount a philosophical argument.

Isidor Farash
Fort Lee, N.J.

In "That Mysterious Flow," Davies argues that the passage of time may be an illusion. When he suggests that knowing this may eliminate expectation, nostalgia and fear of death, I think he is going too far. Physicists love to point out that we shouldn't try to use our everyday knowledge and experience to understand things like cosmology or nuclear physics. But the argument also works in reverse. Everyday matters such as life and death may be best understood using common sense rather than esoteric cosmological theory. How exactly does Davies propose to eliminate our apprehensions and our sense of living in the present? It seems to me that scientists increasingly try to make obscure theories seem more relevant to our everyday lives by making statements like this, which turn out to be pretty meaningless.

Paul Bracken
Martinez, Calif.

I was intrigued by two claims made in your issue. The first: that physicists "who have read serious philosophy generally doubt its usefulness" ["A Hole at the Heart of Physics," by George Musser]. The second: that "clock researchers have begun to answer some of the most pressing questions raised by human experience in the fourth dimension. Why, for example, a watched pot never boils" ["Times of Our Lives," by Karen Wright].

As a professor of philosophy, I thought that I might be useful by addressing that watched-pot question. So I called my three daughters to witness a science experiment. I poured a small amount of water into a small pot and placed the pot on the hot stovetop. One of us served as timekeeper, and the other three watched the pot. At 130 seconds, the water was at a rolling boil. Triumphantly, I announced that I would publish our fully reproducible findings in a scientific forum no less respectable than the Letters column of *Scientific American*. But then my 11-year-old daughter pointed out that while we did observe the water in the pot boil, we did not actually see the pot itself boil, which is what the adage claims. And if the pot itself actually boiled, my 16-year-old chimed in, it would first have to melt, at which point it would no longer be a pot. Consequently, a pot, let alone a watched pot, could never boil.

One of my sons was asked once whether he had ever taken a philosophy class. He responded that his life was a philosophy class. I regret that as a philosopher I cannot contribute much to pressing science questions, except perhaps teaching young people how to think carefully. Do you think science can find such young people useful?

Murray Hunt
Brigham Young University–Idaho

TROUBLE WITH TIME MACHINES

Paul Davies oversimplifies the so-called twin paradox in "How to Build a Time Machine." He states that Sally, after having made a round-trip to a distant star, would return younger than her twin brother, Sam. This is a curiosity but not a paradox. The real paradox is that according to special relativity, while Sally is traveling at near light-speed, both twins would see each other as aging more slowly, because both frames of reference are equally valid. So who would be older when Sally returns?

The resolution lies in general relativity, which tells us that Sally will experi-

STUART BRADFORD

Letters

ence additional time dilation as a result of her acceleration and will therefore be younger when the twins are reunited.

Edward Hitchcock
Toronto

TIME OUT

I was distressed that "Real Time," by Gary Stix, lent credence to the ridiculous concept of Internet time, a name given by Swatch to the simple translation of the Greenwich Mean Time standard established in 1884. Coupled with an unusable 1,000-unit division, this absurd marketing ploy is meaningless. If you go to your e-mail software, select "source" in the menu and read the headers of most e-mails you've received, you will find the GMT standard being used in most of them to synchronize the time differences. Therefore, we can state that Internet time, as well as the standard used around the world, is the venerable GMT.

Hector Goldin
Via e-mail

SPREAD SPECTRUM'S SECRET

Experience shows that spread spectrum won't work as advertised by "Radio Space," by Wendy M. Grossman [News Scan]. As a space-hardware developer and IEEE senior member, I have been involved with numerous modes of spread spectrum since the 1950s. Frankly, all of them can be jammed either by a carrier frequency near their center frequency or by any signal generating slightly more total power than they do. The only way out is frequency hopping. But other "hoppers" in the area can still jam that frequency. This is a dirty little secret of the communications industry.

Robert Wilson
Big Lake, Alaska

ERRATA Andrewes ["A Chronicle of Timekeeping," by William J. H. Andrewes] edited *The Quest for Longitude* and co-wrote *The Illustrated Longitude* with Dava Sobel.

A tuning fork vibrates 44, not four, times per tenth of a second ["Instantaneous to Eternal," by David Labrador].

50, 100 & 150 Years Ago

FROM SCIENTIFIC AMERICAN

Radio Astronomy ▪ Radio Commerce ▪ Industrial Luxury

JANUARY 1953

RADIO TELESCOPES—"As the sky has been plotted in greater and greater detail with radio telescopes of improved resolving power, it has become clear that the regions with the greatest concentrations of stars generate the most intense radio waves. Even in our present state of uncertainty regarding the source of the radio waves, this relationship is of the utmost importance to astronomy. The work needs high resolution, and this requires very large radio telescopes. The new telescope at Jodrell Bank station of the University of Manchester, England, is based on the radio telescope which has been in use there for several years, but it will be much bigger, and it can be trained on any part of the sky."

TREATING SCHIZOPHRENICS—"In the face of the overwhelming size of the problem, most psychiatrists today are disposed to resort to the quick, drastic treatments developed during the past 20 years—shock treatments of various kinds (with electricity, Metrazol, insulin, carbon dioxide) or prefrontal lobotomy. Although they produce dramatic immediate results, after years of experience it has now become clear that the results are often temporary; a large proportion of shock-treated patients sooner or later relapse. Within the past 10 years more psychiatrists, especially among the younger ones, have been treating schizophrenia by psychotherapy. In recent years it has been shown that, contrary to Freud's early conclusion, it is possible to achieve a workable transference relationship between a schizophrenic and his therapist. The treatment takes at least two years, and usually longer; it is incomparably more expensive than the quick method of shock treatments."

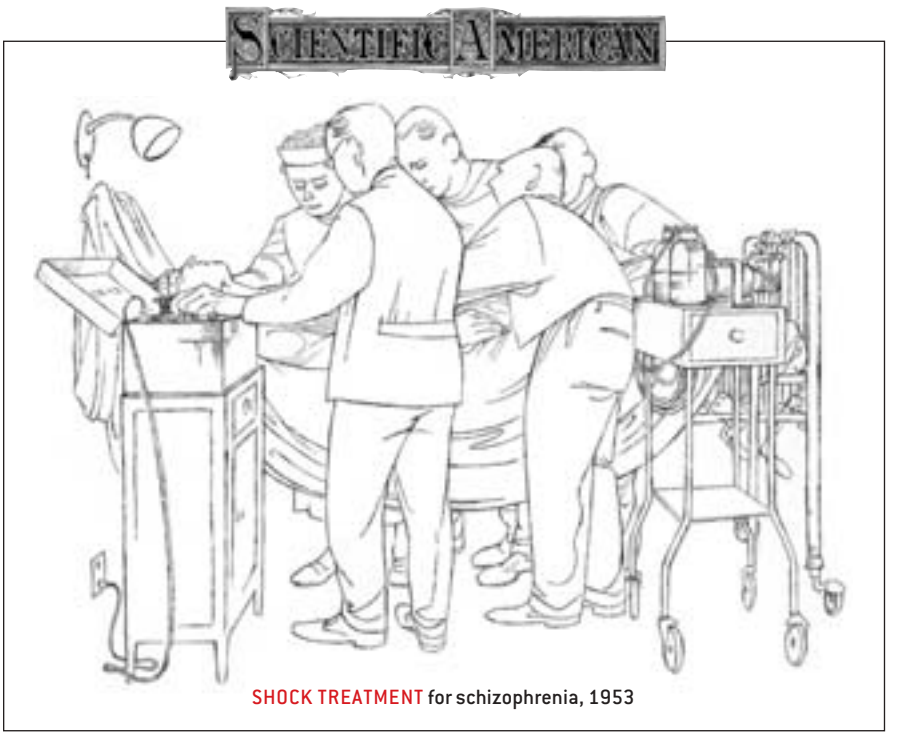
JANUARY 1903

WIRELESS WONDER—"On a barren headland on the eastern shores of Cape Breton, Canada, a few days before Christmas, Guglielmo Marconi exchanged messages of congratulation by wireless telegraphy with some of the crowned heads of Europe. That the brilliant young Anglo-Italian should stand to-day prepared to transmit commercial messages across the Atlantic, must be regarded as certainly the most remarkable scientific achievement of the year."

USEFUL FOR DRUNKS—"A prize of £50 was offered at the Grocers' Exhibition in London for a safe kerosene lamp, that is, for those who use lamps as missiles. The desire of the directors was to produce a cheap lamp, which could be sold even in the poorest districts, and which could be used with the maximum of safety. One of the most serious problems of London was how they could protect those afflicted with drunkenness against themselves. They wanted to find a lamp which, if thrown by a drunken man at his wife or children, would automatically put itself out, so that the man, if he unfortunately inflicted any injury on his wife, should not, at the same time, burn down his house and set fire to his children."

JANUARY 1853

FRUITS OF INDUSTRY—"The Providence (R.I.) Journal laments, with rueful voice, the inordinate progress of luxury: 'The sum necessary, now, to set up a young couple in housekeeping, would have been a fortune to their grandfathers. The furniture, the plate, and the senseless gewgaws with which every bride thinks she must decorate her home, if put into bank stock at interest, would make a handsome provision against mercantile disaster. The taste for showy furniture is the worst and the most vulgar of all. The man who would not rather have his grandfather's clock ticking behind the door, than a gaudy French mantel clock in every room in his house, does not deserve to know the hour of the day.' Yet while we agree with some of its remarks, we dissent from others. We like to see progress in building, dress, and everything that is not immoral."



SHOCK TREATMENT for schizophrenia, 1953

INFOTECH

Getting under Your Skin

REGULATORY QUESTIONS ABOUT IMPLANTABLE CHIPS PERSIST **BY DAVID APPELL**

When identification microchips were implanted in members of the Jacobs family on the *Today* show last May, George Orwell's surveillance society seemed to be another step closer. But confusion over the chip's medical status and even safety, among other stumbling blocks, has left many wondering if the era of the embedded human ID really is at hand.

For several months, Applied Digital Solutions (ADS) in Palm Beach, Fla., has been offering an integrated chip, called the VeriChip, that is about the size of a grain of rice and is injected beneath the dermal layers. Operating just like those in millions of pets, the chip returns a radio-frequency signal from a wand passed over it. The chip can serve as basic identification or possibly link to a database containing the user's medical records. ADS is also planning a chip with broadcasting capabilities—a kind of human "lojack" system that could signal the bearer's GPS coordinates, perhaps serving as a victim beacon in a kidnapping.

As of mid-November 2002, 11 people in the U.S. and several people overseas had been "chipped," says ADS president Scott R. Silverman. But the company ran into problems after the VeriChip's May rollout. Because ADS had said the chip data could be transmitted to an "FDA compliant" site, the Food and Drug Administration insisted on taking a closer look. (Adding to ADS's woes, stockholders filed class action lawsuits alleging that ADS had falsely claimed that some Florida-area hospitals were equipped with scanning devices. And the company's stock price, which rose by almost 400 percent last April and May, tanked last summer and was delisted from the Nasdaq.)

On October 22, 2002, the FDA somewhat surpris-

GREG WHITESELL

FAST FACTS: CHIPS, ANYONE?

Some features of Applied Digital Solutions's VeriChip:

- **12 millimeters long by 2.1 millimeters wide**
- **Encased in glass**
- **Special polyethylene sheath bonds to skin**
- **Cost of being "chipped": $200**
- **Expected lifetime: 20 years**

ingly announced that the VeriChip would not be considered an FDA-regulated device if it were used for "security, financial, and personal identification/safety applications." But it is a regulated medical device "when marketed to provide information to assist in the diagnosis or treatment of injury or illness." The FDA's Office of Compliance is now studying what will be required in the latter case.

Many question the FDA's decision not to regulate the implant fully, because some research suggests that it is not completely safe in animals. A 1990 study in the journal *Toxicologic Pathology* by Ghanta N. Rao and Jennifer Edmondson of the National Institute of Environmental Health Sciences in Research Triangle Park, N.C., reported that a subcutaneous tissue reaction occurred in mice implanted with a glass-sealed microchip device (not unlike the VeriChip). No problems were seen in the 140 mice studied after 24 months, except in those mice that had a genetic mutation in their *p53* gene. In that case, if the device was kept in too long, "these mice develop subcutaneous tumors called fibrosarcomas," Rao says.

HOLD STILL: Like millions of other pets before it, a Labrador mix has a tracking device implanted under its skin. Humans are getting chipped as well.

In humans, the corresponding *p53* mutation causes Li-Fraumeni syndrome, a rare disorder that predisposes patients to a wide variety of cancers. Rao sees reason for caution: "More evaluation may be necessary before they are used in humans."

Implanted microchips may not be so stable under the skin either, according to a 1999 paper in the *Veterinary Record* by Jans Jansen and his colleagues at the University of Nijmegen in the Netherlands. Jansen found that in just 16 weeks chips inserted in shoulder locations of 15 beagles had moved, a few by as much as 11 centimeters. (Transponders implanted in the dogs' heads, however, hardly moved at all.) An inflammatory response is also a risk, Jansen observes, similar to what is sometimes seen with other devices implanted in humans. "I'm not claiming it's not safe," he says, "but you have to be completely sure it will not damage patients in the end."

ADS insists that the implanted chips are safe. The bulk of that evidence "obviously lies with the animal application," Silverman remarks. Other research, such as that reported in 1991 by the Sandoz Research Institute in East Hanover, N.J., found no adverse reactions in rats, although the rodents were observed for only a year. Still, more than 25 million dogs, cats, racehorses and other animals have been chipped without reports of significant problems. Silverman also points out that "no side effects or ramifications whatsoever" have come to those people who received chips in May, including himself and other company executives.

Since the FDA's partial ruling, ADS has signed up distributors in Latin America, Europe and China and has four hospitals in Florida testing the scanners now. Silverman says that the firm has received inquiries from "several hundred people" interested in getting chipped. The procedure can take place at a doctor's office or in ADS's "Chipmobile"; it was scheduled to begin this past December.

The VeriChip faces other hurdles as well. Unless a substantial fraction of the population is chipped, hospitals may not bother installing scanning devices; if that's the case, what good is being chipped? And, of course, the chip introduces the potential for unsolicited surveillance and various privacy violations, a possibility that makes many people's skin crawl, even without a chip under it.

David Appell lives in Ogunquit, Me.

SILICON STITCHING

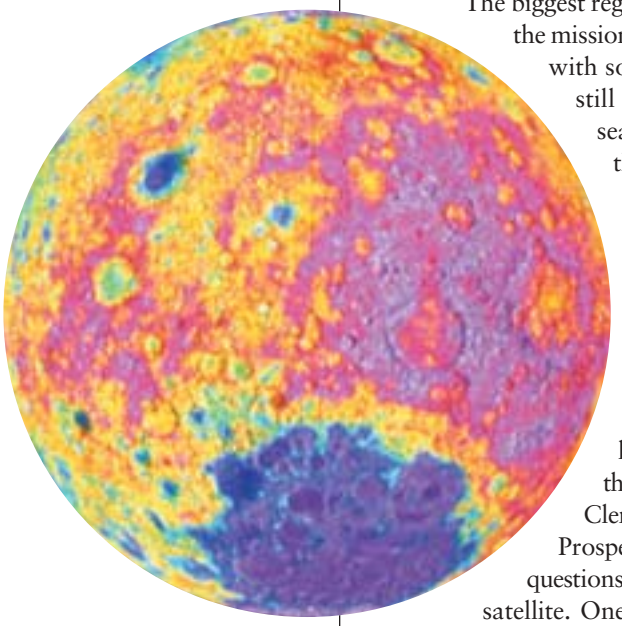
Uses for implantable devices may go far beyond those envisioned for the VeriChip. British engineers are experimenting with a "tooth phone," a chip implanted in a tooth. Futurist Ian D. Pearson of BTexact Technologies in Adastral Park, England, foresees circuitry tattooed into the skin. Such "active skin" would display television pictures, serve as cosmetics, provide a virtual-reality interface without data gloves or goggles, or even—in the ultimate in cyber-isolation—deliver orgasms by e-mail.

ANNIE McCORMICK *AP Photo*

SPACE TRAVEL

Back to the Moon?

PROBES MAY GO, BUT ASTRONAUTS WILL HAVE TO WAIT **BY MARK ALPERT**



FAR SIDE topographic map of the moon shows the immense South Pole–Aitken Basin (*bottom*).

Scientists who study Earth's moon have two big regrets about the six Apollo missions that landed a dozen astronauts on the lunar surface between 1969 and 1972. The biggest regret, of course, is that the missions ended so abruptly, with so much of the moon still unexplored. But researchers also lament that the great triumph of Apollo led to a popular misconception: because astronauts have visited the moon, there is no compelling reason to go back.

In the 1990s, however, two probes that orbited the moon—Clementine and Lunar Prospector—raised new questions about Earth's airless satellite. One stunning discovery was strong evidence of water ice in the perpetually shadowed areas near the moon's poles. Because scientists believe that comets deposited water and organic compounds on both Earth and its moon, well-preserved ice at the lunar poles could yield clues to the origins of life. Just as important, though, was the detection of an immense basin stretching 2,500 kilometers across the moon's far side. Carved out by an asteroid or comet collision, the South Pole–Aitken Basin is a 13-kilometer-deep gouge into the lunar crust that may expose the moon's mantle. It is the largest impact crater in the entire solar system.

Thanks to rock samples collected by Apollo astronauts, lunar geologists know that impact basins on the moon's near side were created about 3.9 billion years ago. South Pole–Aitken is believed to be the moon's oldest basin, so determining its age is crucial. If it turns out to be not much older than the near-side basins, it would bolster the "lunar cataclysm" theory, which posits that Earth and its moon endured a relatively brief but intense bombardment about half a billion years *after* the creation of the solar system. Planetary scientists are at a loss to explain how such a deluge could have occurred.

These discoveries have put the moon back on the exploration agenda, but some scientists are unenthusiastic about the lunar missions that have been scheduled so far. The European Space Agency expects to launch a lunar orbiter called SMART-1 in March, but the craft's primary goal is to test an ion engine similar to the one already tested in NASA's Deep Space 1 mission. Lunar-A, a Japanese probe to be launched this summer, is designed to implant seismometers on the moon by hurling missile-shaped penetrators into the surface, but technical difficulties have limited the craft to only two penetrators, so the risk of failure is high. The Japanese space agency is planning a more ambitious mission named SELENE for 2005, but this lunar orbiter will not be able to answer the fundamental questions posed by the Clementine and Lunar Prospector findings. Says Alan Binder, the principal investigator for Lunar Prospector: "We need to get to the surface, dig it up and see what's there."

An upcoming series of NASA missions, called New Frontiers, will most likely include an unmanned lunar lander that could scoop up about one kilogram of rock fragments from the South Pole–Aitken Basin and then rocket the samples back to Earth for detailed analysis. Michael B. Duke, a Colorado School of Mines geologist who proposed a similar mission in 2000, says the selection of the landing site is critical. Ideally, the site would have impact melt rocks revealing both the age of the basin and the composition of the lunar mantle and would also be close enough to the South Pole so that researchers could test for the presence of ice. To minimize risk, the best solution would be to send landers to more than one site, but that might break the mission's budget, which will probably not exceed $650 million.

The notion of sending astronauts back to the moon seems even more far-fetched given NASA's money troubles. But the agency offered a glimmer of hope last October when Gary L. Martin, NASA's first "space architect," sketched out a possible next step for

COURTESY OF THE LUNAR AND PLANETARY INSTITUTE

MOON PIE IN THE SKY

A few years ago, when space entrepreneurship was all the rage, several companies promised to launch privately funded probes that would explore the moon and make a profit, too. But so far the moon business has remained earthbound. One example: LunaCorp in Fairfax, Va., had intended to finance an ice-hunting mission on the lunar surface by selling television and Internet rights to commercial sponsors. Now the company is focused on a more modest plan—putting a camera-carrying satellite into orbit around the moon—but the effort hinges on persuading NASA to buy the satellite's maps of the lunar surface. David Gump, LunaCorp's president, admits that commercial interest alone is not strong enough to cover the mission's projected $20-million cost.

human exploration: positioning a small space station at the L1 Earth-moon Lagrangian point, where the gravitational pulls of Earth and its moon cancel each other out. Located only 65,000 kilometers from the moon—one sixth the distance between the moon and Earth—this point would provide easy access to the lunar surface (and to Mars as well). But with NASA still struggling to assemble the International Space Station in low-Earth orbit, nobody is expecting to see a replay of Apollo anytime soon.

BIOTECH

Fixing Food

ALLERGEN-FREE COMESTIBLES MIGHT BE ON THE WAY **BY CAROL EZZELL**

A bite of a cookie containing peanuts could cause the airway to constrict fatally. Sharing a toy with another child who had earlier eaten a peanut butter and jelly sandwich could raise a case of hives. A peanut butter cup dropped in a Halloween bag could contaminate the rest of the treats, posing an unknown risk.

These are the scenarios that "make your bone marrow turn cold," according to L. Val Giddings, vice president for food and agriculture of the Biotechnology Industry Organization. Besides representing the policy interests of food biotech companies in Washington, D.C., Giddings is the father of a four-year-old boy with a severe peanut allergy. Peanuts are among the most allergenic foods; estimates of the number of people who experience a reaction to the legumes hover around 2 percent of the population.

Giddings says that peanuts are only one of several foods that biotechnologists are altering genetically in an attempt to eliminate the proteins that wreak havoc with some people's immune systems. Although soy allergies do not usually cause life-threatening reactions, the scientists are also targeting soybeans, which can be found in two thirds of all manufactured food, making the supermarket a minefield for people allergic to soy. Biotechnologists are zeroing in on wheat, too, and might soon expand their research to the rest of the "big eight" allergy-inducing foods: tree nuts, milk, eggs, shellfish and fish.

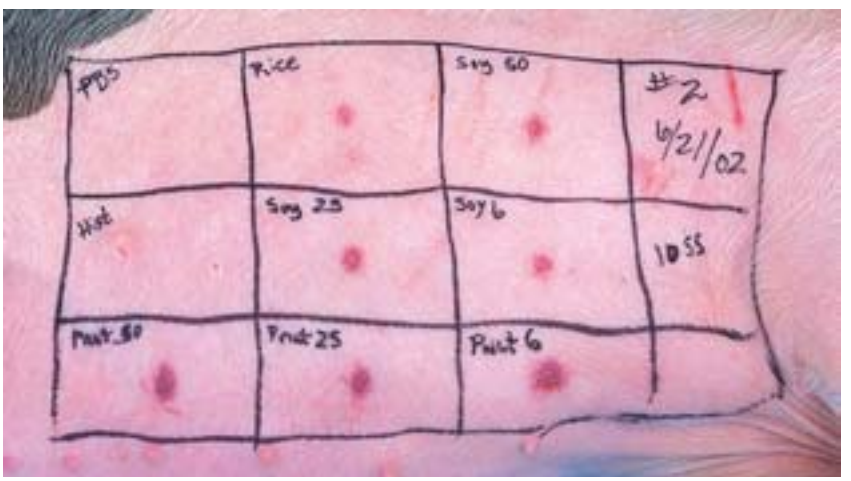
Last September, for example, Anthony J. Kinney, a crop genetics researcher at DuPont Experimental Station in Wilmington, Del., and his colleagues reported using a technique called RNA interference (RNAi) to silence the genes that encode p34, a protein responsible for causing 65 percent of all soybean allergies. RNAi exploits the mechanism that cells use to protect themselves against foreign genetic material; it causes a cell to destroy RNA transcribed from a given gene, effectively turning off the gene.

Whether the public will accept food genetically modified to be low-allergen is still unknown. Courtney Chabot Dreyer, a spokesperson for Pioneer Hi-Bred International, a subsidiary of DuPont, says that the company will conduct studies to determine whether a niche market exists for low-allergen soy before developing the seeds for sale to farmers. She estimates that Pioneer Hi-Bred is seven years away from commercializing the altered soybeans.

Doug Gurian-Sherman, scientific director of the biotechnology project at the Center for Science in the Public Interest—a group that has advocated enhanced Food and Drug Administration oversight for genetically modified

COUNTERING INTOLERANCE

Although relatively few people have outright food allergies—in which the body raises an immune attack against proteins within a food—many more have difficulty digesting some foods. Dairy products are already on the market for those who develop gas, bloating and diarrhea from drinking milk or eating ice cream. A similar product could soon emerge for those allergic to, or intolerant of, wheat gluten. Scientists led by Chaitan Khosla of Stanford University have found an enzyme that when orally administered might allow people with celiac sprue, an allergy to gluten, to eat some wheat products. It could also help buffer the effect of less severe gluten intolerance.



PORCINE PREDICTOR: Various proteins from foods such as soybeans are injected underneath the skin of pigs to help identify those items that are most allergenic.

AGRICULTURAL RESEARCH SERVICE, USDA; SCOTT BAUER

foods—comments that his organization would not oppose low-allergen foods if they prove to be safe. But he wonders about "identity preservation," a term used in the food industry to describe the deliberate separation of genetically engineered and nonengineered products. A batch of nonengineered peanuts or soybeans might contaminate machinery reserved for low-allergen versions, he suggests, reducing the benefit of the gene-altered food. Such issues of identity preservation could make low-allergen genetically modified foods too costly to produce, Chabot Dreyer admits. But, she says, "it's still too early to see if that's true."

Biotech's contributions to fighting food allergies need not require gene modification of the foods themselves, Giddings notes. He suggests that another approach might be to design monoclonal antibodies that bind to and eliminate the complexes formed between allergens and the subclass of the body's own antibodies that trigger allergic reactions. "Those sorts of therapeutics could offer a huge potential," Giddings states, and may be more acceptable to a public wary of genetically modified foods. He definitely sees an untapped specialty market. "When you find out your child has a life-threatening food allergy, your life changes in an instant," Giddings remarks. "You never relax." Not having to worry about every bite will enable Giddings—and his son—to breathe a lot easier.

MATHEMATICS

Color Madness

ODDBALL MAPS CAN REQUIRE MORE THAN FOUR COLORS

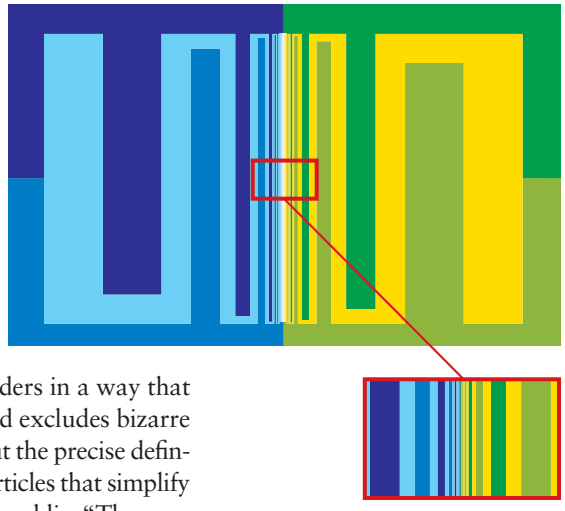
BY GEORGE MUSSER

Science operates according to a law of conservation of difficulty. The simplest questions have the hardest answers; to get an easier answer, you need to ask a more complicated question. The four-color theorem in math is a particularly egregious case. Are four colors enough to identify the countries on a planar map, so that two bordering countries (not counting those that meet at a point) never have the same color? The answer is yes, but the proof took a century to develop and filled a 50-page article plus hundreds of pages of supplementary material.

More complicated versions of the theorem are easier to prove. For instance, it takes a single page to show that a map on a torus requires at most seven colors. The latest example of unconventional cartography comes from philosopher Hud Hudson of Western Washington University in a forthcoming *American Mathematical Monthly* paper. He presents a hypothetical rectangular island with six countries. Four occupy the corners, and two are buffer states that zigzag across the island. The twist is that the zigs and zags change in size and spacing as they go from the outskirts toward the middle of the island: each zigzag is half the width of the previous one. As the zigzags narrow to nothingness, an infinite number of them get squeezed in.

Consequently, the border that runs down the middle of the island is a surveyor's nightmare. If you draw a circle around any point of the border, all six countries will have slivers of territory within that circle. No matter how small you draw the circle, it will still intersect six countries. In this sense, the border is the meeting place of all six countries. So you need six colors to fill in the map. By extending the procedure, you can prepare maps that require any number of colors.

The standard four-color theorem defines borders in a way that hews to common sense and excludes bizarre cases such as Hudson's. But the precise definition is usually left out of articles that simplify the theorem for the general public. "The popular formulation of the four-color problem, 'Every map of countries can be four-colored,' is not true, unless properly stated," says Robin Thomas of the Georgia Institute of Technology. Thomas is one of the co-authors of a shorter proof of the theorem—just 42 pages.



SIX COLORS are needed to fill in this map, with its infinitely meandering countries.

HUD HUDSON *Western Washington University*

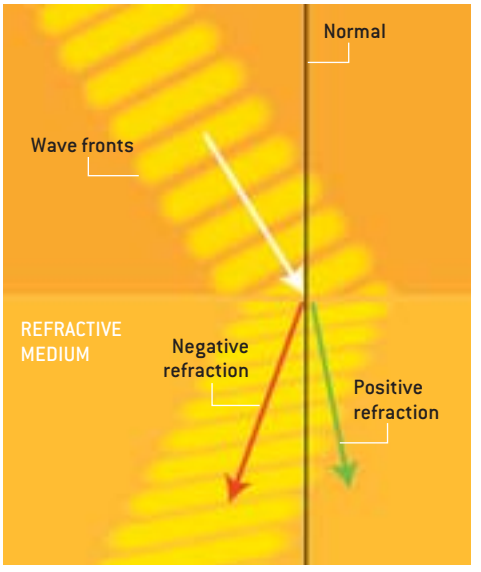
PHYSICS

Heat and Light

DOES NEGATIVE REFRACTION REALLY EXIST?

BY GRAHAM P. COLLINS

Bending light through water or other media is high school science, hardly a subject that would appear to be controversial. But what happens to light in very special media that have a negative index of refraction is currently being hotly debated in leading physics journals and preprints. Ordinary materials such as glass lenses bend light so that the refracted ray is on the opposite side of the "normal," the imaginary line perpendicular to the surface of the medium. In a negative index material, also known as a left-handed material, light is refracted back on the same side of the normal. According to John Pendry of Imperial College, London, an ideal slab of such a material would act like a perfect lens, creating an image that would include details well below the stopping point for conventional lenses, called the diffraction limit. It sounds too good to be true, but for each complaint raised, proponents of negative index materials have at least a partial answer.



NEGATIVE REFRACTION bends light back to the same side of the "normal," the line perpendicular to the refractive medium. Ordinarily, light bends on the opposite side (positive refraction). Note that the modulation wave fronts in negative refraction have the same orientation expected for positive refraction.

The refractive index of a substance is determined by two properties known as the electrical permittivity and the magnetic permeability. In the 1960s Russian physicist Victor Veselago noted that in a material where both those quantities are negative, the refractive index will also be negative. In a negative index material (NIM), the peaks and troughs of an electromagnetic wave travel backward even though the energy of the wave continues to travel forward. This behavior leads to predictions of numerous strange phenomena, such as the reversal of the usual Doppler effect (traveling toward a wave results in a redshift instead of a blueshift).

Many materials, including plasmas and metals, have a negative permittivity, but no natural substance has a negative permeability as well. In 2001 a group led by David Smith of the University of California at San Diego demonstrated that a "metamaterial" could be built to have the requisite negative permeability and permittivity for a narrow band of microwaves. The metamaterial is made of an array of tiny copper loops and wires. The San Diego group showed that microwaves passing through a small prism of the metamaterial were refracted in the opposite direction of waves passing through a similarly shaped prism of Teflon (Teflon is to microwaves as glass is to visible light).

The negative index interpretation was soon challenged, however. Nicolas Garcia and Manuel Nieto-Vesperinas of the National Research Council of Spain in Madrid claimed that the prism was merely absorbing more microwaves at its thick end. They carried out an experiment with a thin wedge of gold and visible light to demonstrate similar results with no negative refraction. Smith counters that the gold wedge has many orders of magnitude greater absorption than his group's metamaterial and fails to reproduce the propagating refracted beam his group detects.

In a paper published last May, Prashant Valanju and his co-workers at the University of Texas at Austin disputed the basic theory of negative index materials. They pointed out that modulation wave fronts (such as those in the front of a light pulse) in a negative index material are aligned the same way as they are in a positive index material—anything else would violate basic causality and

IAN WORPOLE

PROBLEMS WITH PERFECT LENSES

A flat slab of a negative index material causes light from a nearby point source to converge to a focus rather than diverging. Waves with features shorter than the light's wavelength increase in amplitude inside the slab instead of attenuating. With an ideal slab, those effects would lead to perfect, sub–diffraction limit imaging but would also cause infinite energy buildup in a very thick slab—an absurdity. In a real slab, absorption losses and dispersion prevent the infinite energy buildup but might still allow superimaging. The perfect imaging occurs only at a single frequency of light—that for which the material's refractive index is exactly –1, the opposite of a vacuum. If superimaging were possible for visible light, DVDs could be made to store 100 times more data and semiconductors could be fabricated with features one tenth the size of those possible today.

require parts of the wave to travel with infinite velocity. Also, any negatively refracted wave would be rapidly smeared out in only a few wavelengths by dispersion, which Valanju maintains will always be a major problem in a negative index material.

Smith and Pendry agree that the wave fronts are aligned the way Valanju says but contend that the waves nonetheless travel in the direction of negative refraction. As for dispersion, Smith notes that NIMs are not intrinsically worse than positive index materials in that regard and that within narrow but useful bandwidths the negatively refracted waves can persist. Indeed, the U.C.S.D. team got some validation after a group at Boeing's Phantom Works, including physicist Claudio Parazzoli, repeated the prism experiment out to about 30 wavelengths—much farther than was done in the original experiment.

The question of a perfectly imaging slab remains less clear. Some papers claim that absorption and dispersion will completely spoil the effect. Others argue that although *perfect* imaging is impossible, sub–diffraction limit imaging is still feasible, provided the metamaterial meets a stringent set of conditions. Pendry's latest contribution is to suggest that slicing up the slab and alternating thin pieces of NIM with free space will greatly enhance the focusing effect and that such an arrangement will behave somewhat like a fiber-optic bundle channeling the electromagnetic waves, including the sub–diffraction limit components. Groups such as Smith's are working toward testing the superimaging effect. Judging by past form, however, even experimental results are unlikely to settle the debate quickly.

BIOPHYSICS

Shake, Waddle and Stroll

VIBRATING SHOE INSERTS FOR SURER FOOTING

BY CHARLES CHOI

KEEPING BALANCED is helped by exercise—and someday perhaps by tiny vibrations underfoot.

PATRICK MOLNAR *Getty Images*

People begin to lose their balance in their old age just as their bones get more fragile, a deadly combination that can lead to crippling or fatal falls. The elderly grow wobbly in part because their nervous systems become less sensitive to the changes in foot pressure whenever they lean one way or another. No one keeps perfect posture—everyone sways at least a little—and the brain needs the cues from the soles to stay balanced.

Foot massages could help those who have balance problems. Research led by bioengineer James J. Collins of Boston University shows that gentle stimulation of the feet helps elderly study subjects. The key is that the vibrations must be random—or, put another way, noisy. Usually, noise interferes with the main signal—think of static drowning out a television picture or attempts at conversation in a crowded room. Under the right circumstances, however, noise can actually boost weak signals. The effect is known as stochastic resonance, and it occurs in electronic circuits, global climate models and nerve cells. To see how it works, imagine a frog in a jar: by itself the amphibian might not be able to jump out, but if the jar is in a rumbling truck the frog might get the boost it needs to make it. In the same way, a faint background of random pulses could amplify weak signals sent from the feet to the brain.

The researchers built a platform with hundreds of randomly vibrating nylon rods on which volunteers stood barefoot with eyes closed and arms at their sides. When the rods were tuned so the participants said they could no longer feel their shaking, Collins and his colleagues found that the 16 senior citizens, with an average age of 72, swayed much less. In fact, they performed as well as the young volunteers, average age 23, on solid ground. When the vibrations were perceptible, no benefits were seen.

Collins's team has already developed half-inch-thick vibrating gel insoles, and when subjects stood on prototypes, they swayed even less than they did on the platform. "Within a couple of years one could have commercially viable insoles ready," Collins hopes, thereby marking the first everyday application of stochastic resonance.

Charles Choi is based in New York City.

Warming Up America

CONVENTIONAL FUELS STILL REIGN IN HOME HEATING **BY RODGER DOYLE**

Central heating, available in the U.S. since the early 19th century, became popular only after the Civil War. Typically, coal-burning furnaces fueled the early systems. The furnaces warped and cracked, causing gases to escape, and had to be stoked frequently. It took years and countless small improvements, but by the mid-1920s the systems had become reliable and, with the emergence of oil-fired furnaces, more convenient.

Natural gas, which became widely available with the building of a pipeline infrastructure after World War II, had developed into the leading fuel by 1960. Its acceptance resulted in part from its versatility—unlike oil, it can power appliances such as clothes washers and dryers, ovens, ranges and outdoor grills. Because it comes primarily from U.S. and Canadian fields, natural gas is also less vulnerable than oil is to war and embargo. Oil remains the predominant fuel in a few areas, such as New England, where natural-gas pipelines have not yet thoroughly penetrated. Oil users in many regions have the advantage of being able to lock in the price of a season's supply and, in contrast to most gas users, can easily change their supplier.

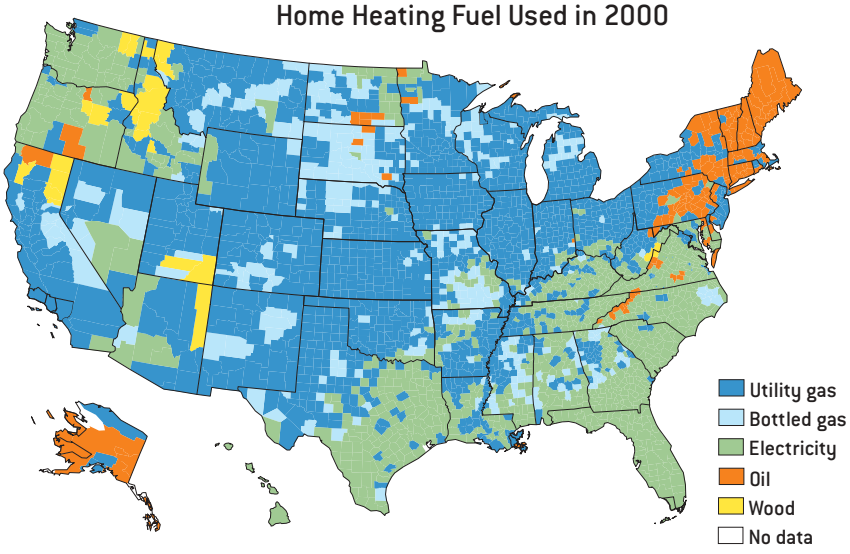
Electric heating, which appeared in less than 1 percent of homes in 1950, now dominates most areas with mild winters and cheap electricity, including the South and the Northwest. Its popularity, at least in the South, was spurred by the low cost of adding electric heating to new houses built with air-conditioning. In the Northeast and Midwest, electricity has not been a popular fuel because of its high cost for cold-weather heating and because it delivers heat at 90 to 95 degrees Fahrenheit, compared with 120 to 140 degrees F for gas and oil, which many in cold climates find preferable. In some areas, such as California, electric heating has not progressed because of building code restrictions. Bottled gas, which is somewhat more expensive than utility gas, is the fuel of choice in rural areas not served by utility pipelines. Wood, the dominant fuel throughout the U.S. economy until the eighth decade of the 19th century, is the leading heating fuel in just a few rural counties.

Home heating, which accounts for less than 7 percent of all energy consumed in the U.S., has had a commendable efficiency record: from 1978 to 1997, the amount of fuel consumed for this purpose declined 44 percent despite a 33 percent increase in the number of housing units and an increase in house size. This improvement came about thanks to better insulation and more efficient equipment following the energy crisis of the 1970s. The U.S. Department of Energy, however, forecasts that energy used in home heating will rise 14 percent over the next two decades. That upsurge is small considering an expected 21 percent increase in the number of houses and the trend toward larger houses. Total energy consumption in the economy, including transportation, manufacturing and commerce, is projected to rise by 31 percent.

Natural gas and electricity will probably dominate the home heating market for the next two decades. Solar heating never took off because of cost and limited winter sunlight in most areas; in 2000 only 47,000 homes relied on it. Fuel cells for home heating are unlikely to be competitively priced until 2010 at the earliest.

Rodger Doyle can be reached at rdoyle2@adelphia.net

Home Heating Fuel Used in 2000



SOURCE: 2000 U.S. Census

HEATING UP THE HOME

Sixty Years of Home Heating Fuel in the U.S., by Percent of Use

Fuel type	2000	1980	1960	1940
Utility gas	51.2	53.1	43.1	11.3
Electricity	30.4	18.4	1.8	—
Oil	9.0	18.2	32.4	10.0
Bottled gas	6.5	5.6	5.1	—
Coal	0.14	0.6	12.2	54.7
Wood	1.7	3.2	4.2	22.8
Solar	0.04	—	—	—
Other fuels	0.4	0.2	0.4	0.4
No fuel	0.7	0.7	0.9	0.8

SOURCE: U.S. Census Bureau. Oil includes kerosene; coal includes coke; bottled gas includes propane and petroleum gas in tanks or in liquid form.

RODGER DOYLE

DATA POINTS: SHORT-STAFFED

The chronic shortage of hospital nurses is affecting U.S. health care. A May 2002 report found increased stays and infection rates among patients when nursing help is scarce. A more recent survey of 168 Pennsylvania hospitals finds that patient deaths become more likely.

Daily amount of nursing care per patient: 11.4 hours

Percent of nurses saying they feel burned out: 43.2

Percent dissatisfied with current job: 41.5

Percent who plan to quit current job within one year: 20

Average number of patients per nurse: 4 to 8

If workload increases by one patient per nurse ...

Percent increase in patient mortality in 30 days: 7

Percent increase in the odds of a nurse burning out: 23

Percent increase in job dissatisfaction: 15

Estimated number of deaths over the 20 months of the survey, if the patient-per-nurse ratio doubled: 1,000

SOURCES: Journal of the American Medical Association, *October 23–30, 2002; workload hours per patient from* New England Journal of Medicine, *May 30, 2002.*

FALLING STAR could be mistaken for an exploding warhead.

NEAR-EARTH OBJECTS

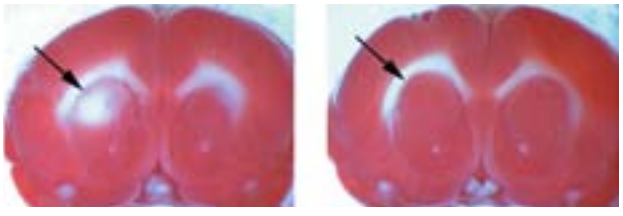
Nuclear Close Call?

An asteroid 15 to 30 miles wide could easily wipe out most of the human race. So might a space rock only 15 to 30 feet in diameter, if trigger-happy nations mistake its fall for a nuclear first strike. That long-standing worry was echoed by U.S. Air Force Brigadier General Simon P. Worden during testimony before a congressional subcommittee last October. He revealed that such a meteoroid burned up over the Mediterranean Sea on June 6, 2002, just as tensions between nuclear powers India and Pakistan were at their highest. U.S. early-warning satellites detected the flash from the rock's entry, which generated an explosion comparable to the Hiroshima burst. Had the meteor entered the atmosphere at the same latitude a few hours earlier, Worden stated, then it could have fallen near the Pakistan-India border and been mistaken for a nuclear detonation. Scientists analyzing U.S. federal satellite data reveal in the November 21, 2002, *Nature* that some 300 three- to 30-foot-wide meteoroids exploded in the upper atmosphere in the past eight years and that once a year a meteoroid burst with the force of five kilotons.

—Charles Choi

NEUROSCIENCE

Sonic Boon



HUNTINGTON-LIKE LESION (*white circular patch*) develops less markedly in a rat's brain if a preventive molecule is administered (*right*).

Sonic the Hedgehog is not just a video game. The gene named after the game triggers a molecular pathway that determines which cells emerge in the central nervous system during embryonic development. Curis, a Cambridge, Mass., biotechnology company, has devised a strategy for creating drugs that mimic the activity of *sonic hedgehog*. It has crafted small molecules that cross the blood-brain barrier in adult mice and activate the signaling pathway to either defend against damage or restore neural function in an area of the brain that has been altered to mimic Parkinson's disease. The molecules also protected nerve cells in models of stroke and Huntington's disease. Any drug development effort to combat neurodegenerative disorders will have to examine carefully whether such pivotal signaling molecules adversely affect untargeted cell populations. Curis presented its results at the Society for Neuroscience meeting last November and in the *Journal of Biology*. It is also one of two research groups that have developed synthetic small molecules to block the *sonic hedgehog* pathway for potential anticancer treatments.

—Gary Stix

P. PARVIAINEN *Photo Researchers, Inc. (top)*; CURIS *(bottom)*; ILLUSTRATION BY MATT COLLINS

PALEONTOLOGY

Flipper Flip-Flop

During the Mesozoic era, long before whales came to preside over the ocean realm, marine reptiles called plesiosaurs were the giants that patrolled the seas. Paleontologists have long sought to understand how these enigmatic beasts, which looked like an ungainly cross between a giraffe and a turtle, captured their prey. Previous work focusing on neck length suggested that the shorter-necked, large-headed plesiosaurs, a group called the pliosauromorphs, were built for high-speed pursuits on the open ocean. The longer-necked, small-headed plesiosauromorphs, on the other hand, were deemed better suited to ambush hunting.

F. Robin O'Keefe of the New York College of Osteopathic Medicine studied the geometry of plesiosaur flippers. Pliosauromorphs, he determined, had low-aspect-ratio flippers optimized for maneuverability and attack—like the short, stout wings of falcons and fighter planes—good for chasing fleet-finned quarry. But plesiosauromorphs had high-aspect-ratio flippers, comparable to the longer, thinner wings of seagulls and bomber planes—fliers built for efficiency and range. Thus, O'Keefe argues that rather than lurking, plesiosauromorphs probably cruised leisurely over long distances in search of smaller, less elusive prey. He presented his findings at the Society of Vertebrate Paleontology meeting held in October 2002. *—Kate Wong*

PLESIOSAURS reigned over the Mesozoic seas.

RICK SARDINHA

MATERIALS SCIENCE

A Stretch for Strong Copper

The harder a material is, the less ductile it tends to be. That trade-off holds true for copper: when composed of tiny crystal domains (or grains) less than 100 nanometers in size, it becomes stronger than the usual coarser-grained copper. Unfortunately, nanocrystalline copper is also usually brittle, which makes practical application difficult. Researchers at Johns Hopkins University have found a way to incorporate both desirable qualities in pure copper. The scientists first cool the metal down with liquid nitrogen and then roll it to a millimeter thickness, thereby breaking up the crystal structure. Careful heat-treating then produces an ultrafine grain structure whose many boundaries make the metal strong, yet it permits about a quarter of the grains to grow coarse, which imparts ductility. The strong but pliable copper, described in the October 31, 2002, *Nature,* could find use in microelectromechanical and biomedical devices. *—Steven Ashley*

PHYSICS

Ice That Sinks

Imagine ice cubes that, when dropped into a glass of water, sink like stones instead of bobbing up to the top. High pressures and temperatures near –200 degrees Celsius can form such ice, which is 25 percent denser than liquid water (ordinary ice is about 8 percent less dense than water). Scientists in the U.K. and Austria used neutron beams to determine that, unlike normal ice, in which molecules line up in crystalline arrays, this very dense ice is amorphous, just like glass and most of the frozen water in the universe. The discovery is the fifth form of amorphous ice (there are 13 kinds of crystalline water ice). A better understanding of how the molecules lock into these structures may explain the behavior of disordered systems in general and of water in life-bearing systems in extreme cold. It might even support a hypothesis that a second form of liquid H_2O exists. The researchers describe their findings in the November 11, 2002, *Physical Review Letters.* *—Charles Choi*

BRIEF POINTS

- **Downloads in a blink: fiber-optic systems commonly rely on lithium niobate crystals to encode electrical signals as light pulses; a new and potentially cheaper polymer sandwich replacement for the crystal can boost signal speed by a factor of 20.**

  Science, November 15, 2002

- **A meta-analysis found that homocysteine levels in the blood are not so strongly correlated with heart attacks as was once thought. A better indicator may be ill will: being hostile predicted future heart disease better than high cholesterol, cigarette smoking or body-mass index.**

  Journal of the American Medical Association, October 23, 2002; *Health Psychology,* November 2002

- **If adults who received childhood smallpox shots (last regularly given in the U.S. in 1972) have residual immunity, then targeted vaccination after a smallpox outbreak could be just as effective as mass vaccinations in preventing the spread of the virus.**

  Science, November 15, 2002

- **Better to give than to receive: elderly people who helped with housework, child care, errands and other tasks reduced their risk of dying by almost 60 percent compared with those who did not help.**

  Psychological Science (in press); www.umich.edu/~newsinfo/

Type It Anywhere

An alumni reunion leads to technology that could banish undersize keypads

By MIKE MAY

In 1998 two inventors, Nazim Kareemi and Cyrus Bamji, struck up a conversation with an informal gathering of alumni from the Massachusetts Institute of Technology in Santa Clara, Calif. Bamji mentioned his concept for controlling electronic devices from a distance—in essence, a new form of remote control. "This idea was humming in my head for some time," he says, "but it didn't gel."

Kareemi, an electrical engineer who had founded PenWare (now owned by Symbol Technologies), a producer of machines that record signatures electronically, took a pragmatic interest in the problem. His experience in the technology business complemented Bamji's ongoing supply of ideas, making the two an ideal team. For his part, Bamji is a jack-of-all-trades and an expert at most. He earned a collection of degrees, from math to computer science, plus a doctoral degree in electrical engineering and computer science, from M.I.T. Then he worked as an architect of electronic devices and systems at Cadence Design Systems in San Jose, Calif.

The two men followed up on their original discussion by starting to think about developing a low-cost gadget that could make a three-dimensional map of its surroundings. After pondering that problem for half a year, they decided that an ideal application would be a virtual keyboard: an image of "q," "w," "e," "r," "t," "y" and the other keys projected on a desktop, where someone could press down fingers. The sequence of keystrokes would be recorded by a nearby personal electronic device or a cellular phone equipped to send electronic mail. The apparatus would register which key had been pressed by using a three-dimensional depth map, which provides information about where a particular key is located.

VIRTUAL KEYBOARD projected onto a tabletop allows a user of a personal electronic device to type faster and more comfortably than with input devices furnished by the manufacturer.

This invention was conceived early in 1999, but financial backing for their brainchild did not come readily. "We presented the keyboard idea to a couple of venture capitalists," Bamji says. "My recollection is that they merely smiled." Yet Kareemi and Bamji believed in their invention, and by April they and an engineer colleague, Abbas Rafii, launched a company called Canesta, based in San Jose, Calif. (The company name is an acronym made from the given names of the founders, plus a few added letters to give it a ring.) They funded the company themselves for a year and then, in 2000, went after their initial round of venture capital and raised $3 million. By that fall they had gone as far as to concoct a working version of the keyboard.

To devise a way for electronics to see in three dimensions, the team wanted to

CANESTA

avoid mistakes made by others who had pursued similar technologies. Earlier researchers who had attempted to create 3-D images had relied on dual cameras and compared images pixel by pixel, a method that demands considerable computer processing. "We took a step back," Bamji explains, "and tried to have a more holistic approach. We needed a 3-D sensor to get away from problems with interpreting light from dark."

Just such a sensing apparatus was incorporated in a product, the Integrated Canesta Keyboard, and introduced in September at a mobile and wireless conference. The product became one of several virtual keyboards that are entering the market.

The electronic guts of this keyboard lie in the Canesta Keyboard Perception Chipset, which includes three parts: a pattern projector, an infrared light source and a sensor. The pattern projector uses a small laser, only nine millimeters on each side, to produce what looks essentially like an ordinary keyboard on a desk. The light gets projected so close to the surface that the user's fingers do not even block it until they touch the desktop. The cylindrical infrared source, a mere 6.5 millimeters in diameter, sends out a beam of infrared light, which bounces off objects and returns to an infrared sensor, an array that can be as small as 100-by-20 light-sensing pixels and that takes up as much room as a pea. When the infrared light is turned on, a timer starts at each pixel, and it stops when the light returns. The time gets converted to distance—how far the light traveled before it hit something—such as a finger touching a virtual key on a tabletop. The sensing mechanism is radar with light.

The collection of distances from the array of pixels provides a 3-D map of the area scanned. Moreover, this device can survey its surroundings more than 50 times every second. Like the pattern projector, the infrared light stays close to the surface. The sensor's view can get blocked if a user hits two keys at once that are exactly in line from the sensor. That happens rarely. But if it does, the keyboard's software makes the shift key "sticky," so even if it gets blocked by a finger on the E, the keyboard will interpret it as the two keys hit together.

The Canesta Keyboard Perception Chipset, according to Kareemi, will cost only tens of dollars, much less than the roughly $80 that current compact keyboards cost for PDAs, and it is now available in sample quantities to companies that will put the chips in their products. The chips are expected to be incorporated in cell phones, PDAs and other electronic products beginning in the first half of 2003.

The first users found it somewhat disconcerting to type without tactile feedback from the virtual keys. So the inventors added click sounds when someone taps a virtual key on the Integrated Canesta Keyboard. As a result, Kareemi says, "it takes 10 to 15 minutes to get used to the virtual keyboard, and then you can type very fast." To find out how fast, Kareemi and his colleagues gathered a group of 20 people who regularly use cell phones, computers and PDAs. On average this group scribbled out 14 words a minute on devices that get input from a stylus, increased that to 25 words a minute on thumb keypads and climbed to 45 words a minute on Canesta's keyboard. The same group, though, pounded out an average of 65 words a minute on an ordinary keyboard. So it seems to take users some time to get used to typing in a virtual world.

Twenty people who regularly use electronic devices typed faster on the Canesta keyboard than with a thumb keypad but slower than on an ordinary keyboard.

The applications, however, go far beyond keyboards. Instead of watching only a person's hands, a Canesta 3-D map could observe an entire person. If the technology were added to a kung fu video game, for instance, a user could stand in the middle of a room and kick and chop as a figure on the screen mimicked the movements. It could also be built into a car to see if other drivers got too close or if a child were in the front passenger seat and the airbag needed to be turned off. Kareemi says, "It could even see if someone was sitting with their legs on the dashboard, and then it could set off the airbag differently in an accident. We can do that easily." Canesta is currently discussing these technologies with three leading automakers, but their names remain secret.

Kareemi and his colleagues have already been granted one patent, and 29 more have been filed. The big question that remains is whether PDA and cellphone users are willing to embrace the typing of e-mails, memos and addresses on the bare surface of a diner lunch counter or an office desktop. SA

Fair Use and Abuse

Get set for an overdue national debate about consumer rights in the digital age By GARY STIX

The Big Red Shearling toy bone allows dog owners to record a short message for their pet. Tinkle Toonz Musical Potty introduces a child to the "magical, musical land of potty training." Both are items on Fritz's Hit List, Princeton University computer scientist Edward W. Felten's Web-based collection of electronic oddities that would be affected by legislation proposed by Democratic Senator Ernest "Fritz" Hollings of South Carolina. Under the bill, the most innocent chip-driven toy would be classified as a "digital media device," Felten contends, and thereby require government-sanctioned copy-protection technology.

The Hollings proposal—the Consumer Broadband and Digital Television Promotion Act—was intended to give entertainment companies assurance that movies, music and books would be safe for distribution over broadband Internet connections or via digital television. Fortunately, the outlook for the initiative got noticeably worse with the GOP victory this past November. The Republicans may favor a less interventionist stance than requiring copy protection in talking dog bones. But the forces supporting the Hollings measure—the movie and record industries, in particular—still place unauthorized copying high on their agenda.

The bill was only one of a number that were introduced last year to bolster existing safeguards for digital works against copyright infringement. The spate of proposed legislation builds on a foundation of anti-piracy measures, such as those incorporated into the Digital Millennium Copyright Act, passed in 1998, and the No Electronic Theft Act, enacted in 1997, both of which amend the U.S. Copyright Act.

The entertainment industry should not feel free just yet to harass users and makers of musical potties. Toward the end of the 2002 congressional year, Representative Rick Boucher of Virginia and Representative Zoe Lofgren of California, along with co-sponsors, introduced separate bills designed to delineate fair use for consumers of digital content. Both the Boucher and Lofgren bills look to amend existing law to allow circumvention of protection measures if a specific use does not infringe copyright. Moreover, the Lofgren bill would let consumers perform limited duplications of legally owned works and transfer them to other media.

The divisions that pit the entertainment industry against fair-use advocates should lay the groundwork for a roiling intellectual-property debate this year. Enough momentum exists for some of these opposing bills to be reintroduced in the new Congress. But, for once, consumers, with the support of information technology and consumer electronics companies, will be well represented. In addition to the efforts of Boucher and Lofgren, grassroots support has emerged. Digitalconsumer.org formed last year to combat new protectionist legislative proposals and to advocate alteration of the DMCA to promote digital fair use. The group has called for guarantees for activities such as copying a CD to a portable MP3 player and making backup copies, which are illegal under the DMCA, if copy protection is violated.

The DMCA has not only undercut fair use but also stifled scientific investigations. Felten and his colleagues faced the threat of litigation under the DMCA when they were about to present a paper on breaking a copy-protection scheme, just one of several instances in which the law has dampened computer-security research (see the Electronic Frontier Foundation's white paper, "Unintended Consequences": www.eff.org/IP/DMCA/20020503_dmca_consequences.html). The legal system should try to achieve a balance between the rights of owners and users of copyrighted works. An incisive debate is urgently needed to restore that balance. SA

JOHN McFAUL

Digits and Fidgets

Is the universe fine-tuned for life? By MICHAEL SHERMER

There was a young fellow from Trinity
Who took the square root of infinity.
But the number of digits
Gave him the fidgets;
He dropped Math and took up Divinity.

In the limerick above, physicist George Gamow dealt with the paradox of a finite being contemplating infinity by passing the buck to theologians.

In an attempt to prove that the universe was intelligently designed, religion has lately been fidgeting with the fine-tuning digits of the cosmos. The John Templeton Foundation even grants cash prizes for such "progress in religion." Last year mathematical physicist and Anglican priest John C. Polkinghorne, recognized because he "has invigorated the search for interface between science and religion," was given $1 million for his "treatment of theology as a natural science." In 2000 physicist Freeman Dyson took home a $945,000 prize for such works as his 1979 book, *Disturbing the Universe,* in which he writes: "As we look out into the universe and identify the many accidents of physics and astronomy that have worked together to our benefit, it almost seems as if the Universe must in some sense have known that we were coming."

We may live in a multiverse in which our universe is only one of many universes.

Mathematical physicist Paul Davies also won a Templeton prize. In his 1999 book, *The Fifth Miracle,* he makes these observations about the fine-tuned nature of the cosmos: "If life follows from [primordial] soup with causal dependability, the laws of nature encode a hidden subtext, a cosmic imperative, which tells them: 'Make life!' And, through life, its by-products: mind, knowledge, understanding. It means that the laws of the universe have engineered their own comprehension. This is a breathtaking vision of nature, magnificent and uplifting in its majestic sweep. I hope it is correct. It would be wonderful if it were correct."

Indeed, it would be wonderful. But not any more wonderful than if it were not correct. Even atheist Stephen W. Hawking sounded like a supporter of intelligent design when he wrote: "And why is the universe so close to the dividing line between collapsing again and expanding indefinitely?… If the rate of expansion one second after the Big Bang had been less by one part in 10^{10}, the universe would have collapsed after a few million years. If it had been greater by one part in 10^{10}, the universe would have been essentially empty after a few million years. In neither case would it have lasted long enough for life to develop. Thus one either has to appeal to the anthropic principle or find some physical explanation of why the universe is the way it is."

In its current version, the anthropic principle posits that we live in a multiverse in which our universe is only one of many universes, all with different laws of nature. Those universes whose parameters are most likely to give rise to life occasionally generate complex life with brains big enough to achieve consciousness and to conceive of such concepts as God and cosmology and to ask such questions as Why? Another explanation can be found in the properties of self-organization and emergence. Water is an emergent property of a particular arrangement of hydrogen and oxygen molecules, just as consciousness is a self-organized emergent property of billions of neurons. The evolution of complex life is an emergent property of simple life: prokaryote cells self-organized into eukaryote cells, which self-organized into multicellular organisms, which self-organized into … and here we are.

Self-organization and emergence arise out of complex adaptive systems that grow and learn as they change. As a complex adaptive system, the cosmos may be one giant autocatalytic (self-driving) feedback loop that generates such emergent properties as life. We can think of self-organization as an emergent property and emergence as a form of self-organization. Complexity is so simple it can be put on a bumper sticker: LIFE HAPPENS.

If life on earth is unique or at least exceptionally rare (and in either case certainly not inevitable), how special is our fleeting, mayfly-like existence? And how important it is that we make the most of our lives and our loves; how critical it is that we work to preserve not only our own species but all species and the biosphere itself. Whether the universe is teeming with life or we are alone, whether our existence is strongly necessitated by the laws of nature or highly contingent and accidental, whether there is more to come or this is all there is, we are faced with a worldview that is breathtaking and majestic in its sweep across time and space. SA

Michael Shermer is publisher of Skeptic *magazine (www.skeptic.com) and the author of* In Darwin's Shadow.

BRAD HINES

Science to Save the World

Economist Jeffrey D. Sachs thinks the science and technology of resource-rich nations can abolish poverty, sickness and other woes of the developing world By DAVID APPELL

In a borrowed office on the 16th floor of United Nations Building 2, Jeffrey D. Sachs is on the telephone when I arrive. Although he began working in New York City only eight weeks ago, he seems right at home. He calls the city a unique base of operations. "I think New York is one of the few places in the world where one could find the breadth, the scale and the depth of expertise that you need to be able to address this," he comments.

FLYNN LARSEN

JEFFREY D. SACHS: SELLING SCIENCE

- Director, Earth Institute at Columbia University; a special adviser to U.N. Secretary-General Kofi A. Annan; chairman of the World Health Organization Commission on Macroeconomics and Health
- Early interest in economics sprang from the tension between capitalism and socialism. "Economics answers the most fundamental questions."
- Tireless world traveler: "The only person I know who goes to India just for the day," says his assistant, Gordon McCord.

By "this," he is referring to sustainable development—and how science and technology can be brought to bear on poverty, AIDS, tropical diseases, climate change and other issues confronting the globe.

Sachs is director of Columbia University's Earth Institute, a collection of about 1,000 scholars across eight institutions. He is also a special adviser to Secretary-General Kofi A. Annan on the Millennium Development Goals, eight key development objectives endorsed by more than 160 world leaders in 2000, and was recently chair of the World Health Organization Commission on Macroeconomics and Health. His curriculum vitae runs 26 pages.

In a pressed white shirt, red tie and blue suit, with J.F.K.-like hair, the 48-year-old Sachs is quick, complete and polished. In the two months after he left Harvard University for Columbia in July, he has been to Columbia's Biosphere 2 Center in Arizona, the Barcelona AIDS conference, Cambodia, the Tibetan plateau, and then to the Johannesburg Summit on sustainability before heading back to New York for the beginning of the semester.

His extensive travels have led him to realize the importance of geography, he informs me as we wait for his first appointment on a brilliant September morning. "It isn't possible to do good economic development thinking without understanding the physical environment, deeply, in which economic development is supposed to take place," he says. He complains that this "physical framing" is hardly considered by the World Bank and International Monetary Fund, nor is it taught to graduate students in economics.

As a result, "the physical scientists inherently feel that public policy somehow passes them by," Sachs remarks. "They feel politicians neglect a lot of the important messages or don't understand the risks, say, of anthropogenic climate change or of biodiversity depletion." Yet he has often encountered a resistance among social scientists, who believe everything is at root a political problem.

When the president of Bulgaria, Georgi Purvanov, arrives, the meeting is a getting-to-know-you, with Purvanov asking through an interpreter for Sachs's help. "[European Union] membership and entry to NATO will be the framework to make Bulgaria advance the fastest," Sachs tells Purvanov. He also recommends that Bulgaria invest in education, science, and technology and points to Ireland's growth in information technology and financial services as a good model. He urges Purvanov to flatter corporate CEOs for their business.

One of Sachs's first international triumphs was as an economic adviser to the government of Bolivia from 1986 to 1990, when he helped to bring down that country's inflation rate from 40,000 percent a year to 10. But his role as leading economic adviser to Russia in 1992 and 1993 has drawn criticism: advice such as the elimination of price controls and of subsidies to unprofitable state enterprises had proved successful in eastern European governments but was fruitless in Russia's tumultuous transition to capitalism.

The meeting with Purvanov ends with thanks all around, and immediately Sachs is before the bright light of a Bulgarian television crew. His assistant, Gordon McCord, worries that we have 30 minutes to get to someplace that is 45 minutes away. Sachs ends his television interview, and we race six blocks through the U.N. security zone to a waiting town car.

"The rich are already rich enough to be able to end poverty."

Once inside, Sachs jumps on his cell phone, talking to a reporter from the *Nation* about cross-border commercial bank lending. Traffic is a mess and has our driver swearing. We pull up to the Crowne Plaza hotel near LaGuardia Airport 45 minutes late; hanging up, he comments that his life is "pretty much to the wall every day."

Prominent in Sachs's frequent op-ed pieces is the inadequacy of foreign aid in light of the tremendous problems affecting the developing world—the genesis of which, he says, was the American use of foreign aid as a tactical tool during the cold war. The strategy, he thinks, remains in play.

"So far the United States remains committed to gimmickry rather than real solutions. In the short term the U.S. is courting a worldwide backlash of anti-Americanism" that nontravelers don't recognize. And he sees the U.S. eventually suffering from its failure to address the collapses of governments, failed economies, mass refugee movements, the spread of disease and terrorist activities arising from such conditions—not to mention the longer-term risks of climate change, biodiversity loss and the depletion of vital biological resources.

Over the next 45 minutes Sachs presents his views of the Indian economy, off the cuff, to about 75 participants at the Global Organization of People of Indian Origin conference. Again he demonstrates his mastery of economic details, holding forth on India's business environment, its recent rain-deficient monsoon season and especially its "profound underinvestment" in health care: only about $2 per person per year.

Sachs is not a fan of unfettered capitalism. "I don't believe in free markets for health care and science policy," he says. Long fascinated by the debate between capitalism and socialism, the Detroit native studied economics at Harvard all the way to his Ph.D. in the field; he received tenure there at the age of 28.

Back in the car, Sachs calls Bono, lead singer of the band U2, who has been active in addressing the problems of developing countries. They traveled together last January, when a visit to an AIDS hospital in Malawi left a deep impression on Sachs. The ward was filled with patients, in some cases three to a bed—or huddling under them, out of the way. Sachs has written of "a constant low-level moan and fixed gazes of the emaciated faces," all for the lack of a dollar-a-day's worth of antiretroviral drugs sold elsewhere in the building. The trip, he explains, demonstrated to him "why you have to be there to get it."

Bono is "very impressive and committed," Sachs says. He leaves him a message about an upcoming meeting of philanthropic foundations at investor George Soros's house.

At the Lamont-Doherty Earth Observatory in Palisades, N.Y., Sachs gives a lecture to his troops, the first time many of them have seen him in person. His talk includes a long and impressively detailed aside on the biology and epidemiology of malaria. "Malaria has been the single greatest shaper of wealth and poverty in the world," he informs the group.

"He is the best ally we could have for raising money for malaria research," says Harold Varmus, who is president of Memorial Sloan-Kettering Cancer Center in New York and who served on the WHO commission with Sachs. Varmus relates how Sachs, while writing the WHO report, took a train with his wife on the Silk Road across Asia, e-mailing sections of the report from Mongolian villages. "He is a phenomenon," Varmus adds.

There's no time for a beer afterward—Sachs is catching a plane back to Boston for the weekend, where his wife and son still live until his son graduates from high school this year. "Like every good day," he says, "it ends with a mad dash to some airport."

When we're in a cellular dead zone, I ask Sachs for his broader views. He sees many underlying trends that are very positive, especially the mobilization of science and technology around the world. "The rich are already rich enough to be able to end poverty. But we have the capacity to wreck things," too, he states. "So many of our problems revolve around our capacity to cooperate on a global scale, which we've never done before in the history of the world. We have to do the things we've never done before." At the airport Sachs tips the driver, bids us good-bye, and goes off to do them. SA

David Appell, a frequent contributor, is based in Ogunquit, Me.

Pigments that turn caustic on exposure to light can fight cancer, blindness and heart disease. Their light-induced toxicity may also help explain the origin of vampire tales

New Light on Medicine

Stories of vampires date back thousands of years. Our modern concept stems from Bram Stoker's quirky classic *Dracula* and Hollywood's Bela Lugosi—the romantic, sexually charged, blood-sucking outcast with a fatal susceptibility to sunlight and an abhorrence of garlic and crosses. In contrast, vampires of folklore cut a pathetic figure and were also known as the undead. In searching for some underlying truth in vampire stories, researchers have speculated that the tales may have been inspired by real people who suffered from a rare blood disease, porphyria. And in seeking treatments for this disorder, scientists have stumbled on a new way to attack other, more common serious ills.

Porphyria is actually a collection of related diseases in which pigments called porphyrins accumulate in the skin, bones and teeth. Many porphyrins are benign in the dark but are transformed by sunlight into caustic, flesh-eating toxins. Without treatment, the worst forms of the disease (such as congenital erythropoietic porphyria) can be grotesque, ultimately exacting the kind of hideous disfigurement one might expect of the undead. The victims' ears and nose get eaten away. Their lips

LIGHT-ACTIVATED DRUGS used in photodynamic therapy could treat diseases of the eye, cancers such as those of the esophagus, and coronary artery disease.

By Nick Lane

and gums erode to reveal red, fanglike teeth. Their skin acquires a patchwork of scars, dense pigmentation and deathly pale hues, reflecting underlying anemia. Because anemia can be treated with blood transfusions, some historians speculate that in the dark ages people with porphyria might have tried drinking blood as a folk remedy. Whatever the truth of this claim, those with congenital erythropoietic porphyria would certainly have learned not to venture outside during the day. They might have learned to avoid garlic, too, for some chemicals in garlic are thought to exacerbate the symptoms of the disease porphyria, turning a mild attack into an agonizing reaction.

While struggling to find a cure for porphyria, scientists came to realize that porphyrins could be not just a problem but a tool for medicine. If a porphyrin is injected into diseased tissue, such as a cancerous tumor, it can be activated by light to destroy that tissue. The procedure is known as photodynamic therapy, or PDT, and has grown from an improbable treatment for cancer in the 1970s to a sophisticated and effective weapon against a diverse array of malignancies today and, most recently, for macular degeneration and pathologic myopia, common causes of adult blindness. Ongoing research includes pioneering treatments for coronary artery disease, AIDS, autoimmune diseases, transplantation rejection and leukemia.

Molecular Mechanisms

THE SUBSTANCES AT THE HEART of porphyria and photodynamic therapy are among the oldest and most important of all biological molecules, because they orchestrate the two most critical energy-generating processes of life: photosynthesis and oxygen respiration. Porphyrins make up a large family of closely related compounds, a colorful set of evolutionary variations on a theme. All porphyrins have in common a flat ring (composed of carbon and nitrogen) with a central hole, which provides space for a metal ion such as iron or magnesium to bind to it. When aligned correctly in the grip of the porphyrin rings, these metal atoms catalyze the most fundamental energy-generating processes in biology. Chlorophyll, the plant pigment that absorbs the energy of sunlight in photosynthesis, is a porphyrin, as is heme, which is at the heart of the oxygen-transporter protein hemoglobin and of many enzymes vital for life, including cytochrome oxidase (which generates energy by transferring electrons to oxygen in a critical step of cellular respiration).

Porphyria arises because of a flaw in the body's heme-making machinery. The body produces heme and other porphyrins in a series of eight coordinated stages, each catalyzed by a separate enzyme. Iron is added at the end to make heme. In porphyria, one of the steps does not occur, leading to a backlog of the intermediate compounds produced earlier in the sequence. The body has not evolved to dispose of these intermediates efficiently, so it dumps them, often in the skin. The intermediates do not damage the skin directly, but many of them cause trouble indirectly. Metal-free porphyrins (as well as metalloporphyrins containing metals that do not interact with the porphyrin ring) can become excited when they absorb light at certain wavelengths; their electrons jump into higher-energy orbitals. The molecules can then transmit their excitation to other molecules having the right kind of bonds, especially oxygen, to produce reactive singlet oxygen and other highly reactive and destructive molecules known as free radicals. Metal-free porphyrins, in other words, are not the agents, but rather the brokers, of destruction. They catalyze the production of toxic forms of oxygen.

Photosensitive reactions are not necessarily harmful. Their beneficial effects have been known since ancient times. In particular, some seeds and fruits contain photosensitive chemicals (photosensitizers) called psoralens, which indirectly led scientists to experiment with porphyrins. Psoralens have been used to treat skin conditions in Egypt and India for several thousand years. They were first incorporated into modern medicine by Egyptian dermatologist Abdel Monem El Mofty of Cairo University, just over 50 years ago, when he began treating patients with vitiligo (a disease that leaves irregular patches of skin without pigment) and, later, those with psoriasis using purified psoralens and sunlight. When activated by light, psoralens react with DNA in proliferating cells to kill them.

Two American dermatologists, Aaron B. Lerner of Yale University and Thomas B. Fitzpatrick of Harvard University, were struck by the potential of psoralens. In the 1960s they showed that psoralens are activated by ultraviolet (UVA) rays, and the researchers later refined psoralen therapy using an ultraviolet lamp similar to those used in solariums today. Their method became known as PUVA (short for psoralen with UVA) and is now one of the most effective treatments for psoriasis and other skin conditions.

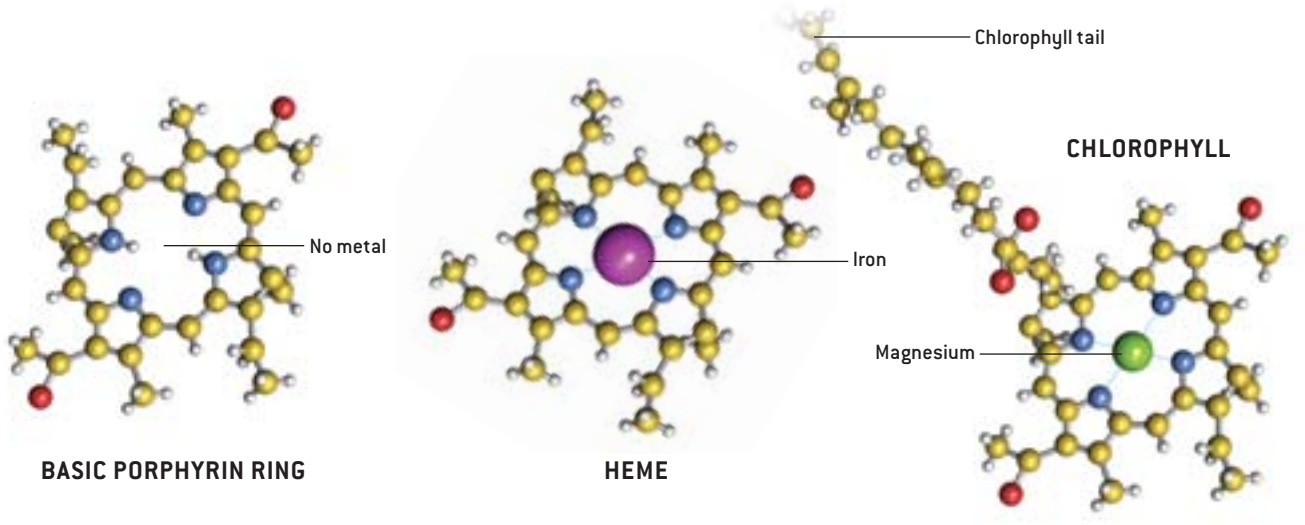
A Way to Kill Cancer Cells?

IN THE EARLY 1970S the success of PUVA impressed Thomas J. Dougherty of the Roswell Park Cancer Institute in Buffalo, N.Y., leading him to wonder if a variant of it could be effective against cancer. Activated psoralens can kill rogue cells to settle inflammation, but in comparison with porphyrins they are not potent photosensitizers. If psoralens could kill individual cells, could porphyrins perhaps devour whole tumors?

Overview/*Light Therapy*

- In photodynamic therapy, light-activated chemicals called porphyrins are used to destroy fast-growing cells and tissue. Doctors could apply the treatment to a variety of ailments, including age-related macular degeneration, tumors and atherosclerotic plaques.
- A few porphyrin drugs are on the market, and several others are undergoing human trials.
- Researchers got the idea for photodynamic therapy from their knowledge of the rare disease porphyria, in which porphyrins accumulate in the skin and certain organs. Unless the disease is managed, victims of the severest type of porphyria can become disfigured, leading some researchers to speculate that they may have inspired medieval vampire legends.

PHOTOGRAPH BY SAM OGDEN (*preceding page*)



PORPHYRINS all have in common a flat ring, mainly composed of carbon and nitrogen, and a central hole where a metal ion can sit. The basic ring (*far left*) becomes caustic when exposed to light; molecules useful for photodynamic therapy also share this trait. Nontoxic examples include heme (a component of the oxygen transporter hemoglobin) and the chlorophyll that converts light to energy in plants.

His idea was the beginning of true photodynamic therapy, in which photosensitizers catalyze the production of oxygen free radicals. It was built on earlier work, which revealed two medically useful properties of the porphyrins: they accumulate selectively in cancer cells and are activated by red light, which penetrates more deeply into biological tissues than do shorter wavelengths, such as blue light or UVA.

Dougherty injected a mixture of porphyrins into the bloodstream of mice with mammary tumors. He then waited a couple days for the porphyrins to build up in the tumors before shining red light on them. His early setup was primitive and passed the light from an old slide projector through a 35mm slide colored red. His results were nonetheless spectacular. The light activated the porphyrins within the tumor, which transferred their energy to oxygen in cells to damage the surrounding tissues. In almost every case, the tumors blackened and died after the light treatment. There were no signs of recurrence.

Dougherty and his colleagues published their data in 1975 in the *Journal of the National Cancer Institute,* with the brave title "Photoradiation Therapy II: Cure of Animal Tumors with Hematoporphyrin and Light." Over the next few years they refined their technique by using a low-power laser to focus red light onto the tumors. They went on to treat more than 100 patients in this way, including people with cancers of the breast, lung, prostate and skin. Their outcomes were gratifying, with a "complete or partial response" in 111 of 113 tumors.

Sadly, though, cancer is not so easily beaten. As more physicians started trying their hand with PDT, some serious drawbacks began to emerge. The affinity of porphyrins for tumors turned out to be a bit of an illusion—porphyrins are taken up by any rapidly proliferating tissue, including the skin, leading to photosensitivity. Although Dougherty's original patients were no doubt careful to avoid the sun, nearly 40 percent of them reported burns and skin rashes in the weeks after PDT.

Potency was another issue. The early porphyrin preparations were mixtures, and they were seldom strong enough to kill the entire tumor. Some porphyrins are not efficient at passing energy to oxygen; others are activated only by light that cannot penetrate more than a few millimeters into the tumor. Some biological pigments normally present in tissues, such as hemoglobin and melanin, also absorb light and in doing so can prevent a porphyrin from being activated. Even the porphyrin itself can cause this problem if it accumulates to such high levels that it absorbs all the light in the superficial layers of the tumor, thus preventing penetration into the deeper layers.

Many of these difficulties could not be resolved without the help of specialists from other disciplines. Chemists were needed to create new, synthetic porphyrins, ones that had greater selectivity for tumors and greater potency and that would be activated by wavelengths of light able to reach farther into tissues and tumors. (For each porphyrin, light activation and absorption occur only at particular wavelengths, so the trick is to design a porphyrin that has its absorption maximum at a wavelength that penetrates into biological tissues.) Physicists were needed to design sources that could produce light of particular wavelengths to activate the new porphyrins or that could be attached to fine endoscopes and catheters or even implanted in tissues. Pharmacologists were needed to devise ways of reducing the time that porphyrins spent circulating in the bloodstream, thereby restricting photosensitive side effects. Finally, clinicians were needed to design trials that could prove an effect and determine the best treatment regimens.

The ideal drug would be not only potent and highly selective for tumors but also broken down quickly into harmless compounds and excreted from the body. The first commercial preparation, porfimer sodium (Photofrin), was approved by the U.S. Food and Drug Administration for the treatment of various cancers. Although it has been helpful against certain cancers (in-

THE AUTHOR

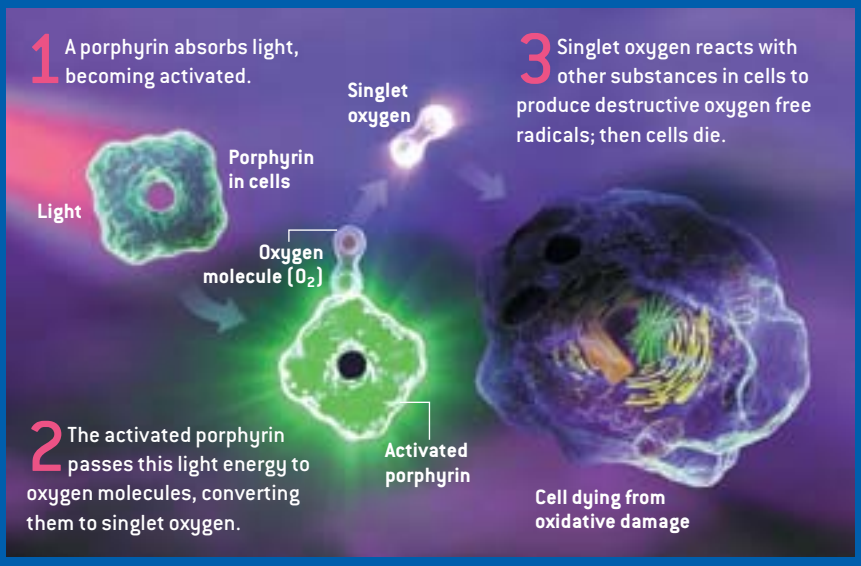
NICK LANE studied biochemistry at Imperial College, University of London. His doctoral research, at the Royal Free Hospital, concentrated on oxygen free radicals and metabolic function in organ transplants. Lane is an honorary senior research fellow at University College London and strategic director at Adelphi Medi Cine, a medical multimedia company based in London. His book, *Oxygen: The Molecule That Made the World,* will be published in the U.S. by Oxford University Press in the spring of 2003.

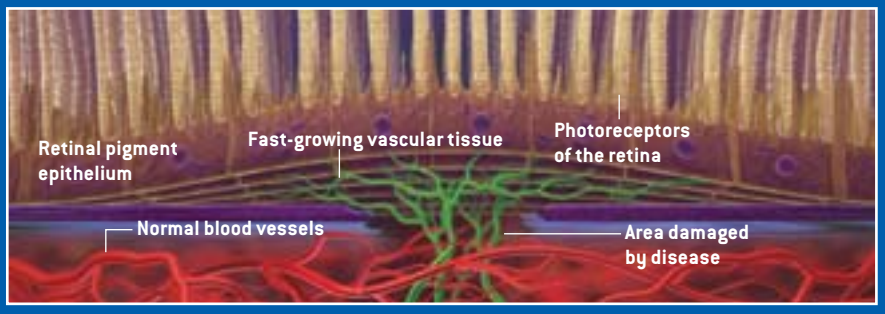
SLIM FILMS

HOW PHOTODYNAMIC THERAPY WORKS

DOCTORS WHO ADMINISTER photodynamic therapy deliver photosensitive chemicals called porphyrins intravenously. These chemicals then collect in rapidly proliferating cells and, when exposed to light, initiate a cascade of molecular reactions that can destroy those cells and the tissues they compose. Some targets for the therapy include abnormal blood vessels in the retinas of people with age-related macular degeneration (the leading cause of adult blindness), cancerous tumors and atherosclerotic plaques in coronary arteries.

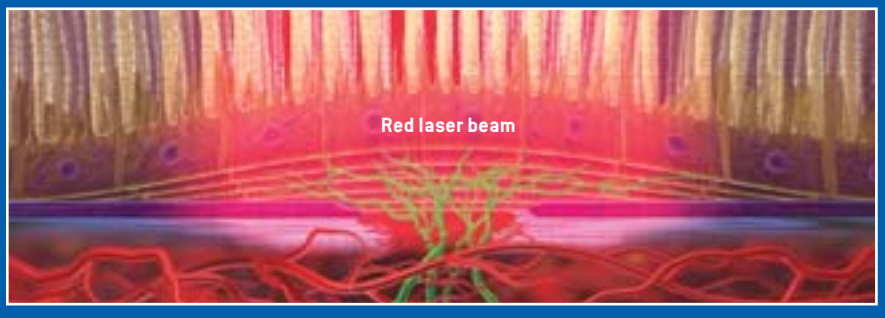
. . . AT THE MOLECULAR LEVEL



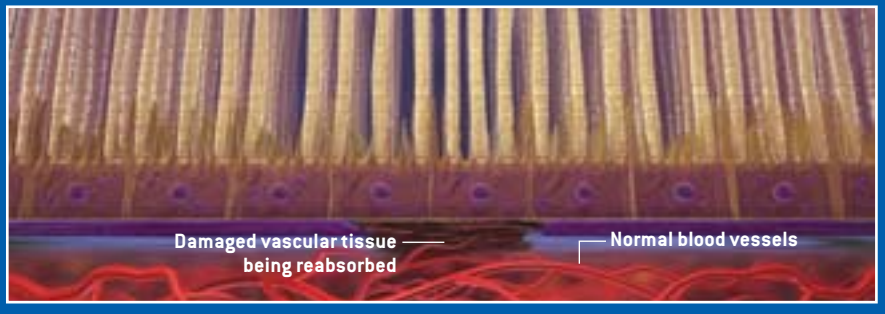


. . . IN THE EYE

1 To treat macular degeneration, a porphyrin (*green*) is injected into a patient's arm. It takes just 15 minutes for the porphyrin to accumulate in abnormal blood vessels under the macula, which is the central part of the retina and responsible for color vision.



2 A red laser light activates the porphyrin, which leads to the destruction of the vascular tissue.

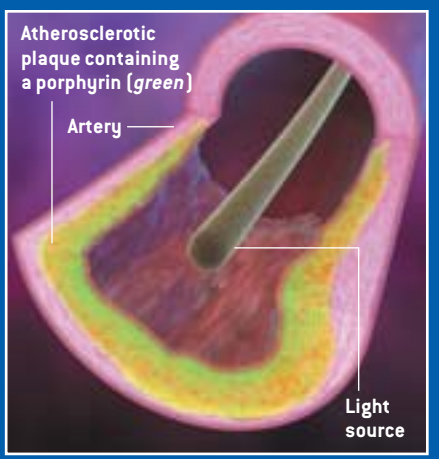


3 After therapy halts damage to the retina, the treated vascular tissue is reabsorbed by the body, and the overlying photoreceptors may settle back into place. Because vessel growth could recur, the patient may require several additional treatments.

... DEEP IN THE BODY

Even long wavelengths of visible light cannot penetrate very far into tissue, so photodynamic therapies for diseased tissue deep within the body require an internal light-delivery system.

1 Here, in an experimental therapy, an optical fiber has been threaded into an artery in which a porphyrin has accumulated in atherosclerotic plaques.



2 The fiber produces red light, activating the porphyrin.

3 Over the course of a few days, the porphyrin destroys unwanted plaques.

HYBRID MEDICAL ANIMATION

cluding esophageal, bladder, head and neck, and skin cancers and some stages of lung cancer), it has not been the breakthrough that had been hoped for and cannot yet be considered an integral part of cancer therapy. Surprisingly, though, the first photosensitizing drug to fulfill most of the stringent criteria for potency and efficacy without causing photosensitivity, verteporfin (Visudyne), was approved in April 2000 by the FDA not to treat of cancer at all but to prevent blindness. As the theories converged with reality, researchers came to realize that PDT can do far more than destroy tumors.

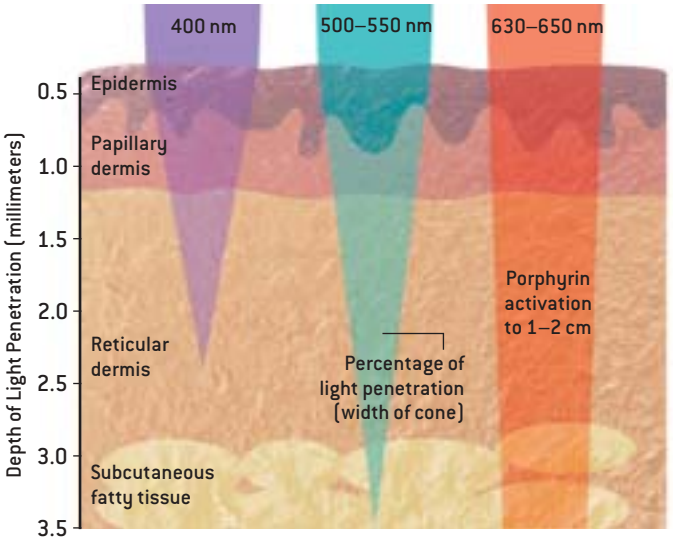
Battling Blindness

ONE THING IT COULD DO, for instance, was combat age-related macular degeneration (AMD), the most common cause of legal blindness in our maturing Western population [see "The Challenge of Macular Degeneration," by Hui Sun and Jeremy Nathans; SCIENTIFIC AMERICAN, October 2001]. Most people who acquire AMD have a benign form and do not lose their sight, but about a tenth have a much more aggressive type called wet AMD. In this case, abnormal, leaking blood vessels, like miniature knots of varicose veins, grow underneath the retina and ultimately damage the sharp central vision required for reading and driving. As the disease progresses, central vision is obliterated, making it impossible to recognize people's faces or the details of objects.

Most attempts to hinder this grimly inexorable process have failed. Dietary antioxidants may be able to delay the onset of the disorder but have little effect on the progression of established disease. Until recently, the only treatment proved to slow the progression of wet AMD was a technique called laser photocoagulation. The procedure involves applying a thermal laser to the blood vessels to fuse them and thus halt their growth. Unfortunately, the laser also burns the normal retina and so destroys a small region to prevent later loss of vision in the rest of the eye. Whether this is worth it depends on the area of the retina that needs to be treated. For most people diagnosed with wet AMD, the area is located below the critical central part of vision or is already too large to benefit from laser coagulation.

Against this depressing backdrop, researchers at Harvard and at the biotechnology firm QLT, Inc., in Vancouver, B.C., reasoned that PDT might halt the growth of these blood vessels and delay or even prevent blindness. If porphyrins could accumulate in any rapidly proliferating tissue—the very problem in cancer—then perhaps they could also accumulate in the blood vessels growing under the retina. Verteporfin, a novel synthetic porphyrin, seemed promising because it had a good track record in preclinical, animal studies at QLT and at the University of British Columbia in the late 1980s and early 1990s.

Verteporfin accumulates in abnormal retinal vessels remarkably quickly: within 15 minutes of injection into an arm vein. When activated by red laser light, verteporfin seals off the vessels, sparing the overlying retina. Any blood vessels that grow back can be nipped in the bud by further treatments. Two major clinical trials, headed by Neil M. Bressler of the Wilmer Eye Institute at Johns Hopkins University, confirmed that PDT



EACH WAVELENGTH OF LIGHT reaches a different depth in tissues, and any given porphyrin absorbs light at specific wavelengths. A porphyrin activated by deeper-penetrating light might be best for treating an internal tumor. In contrast to porphyrins, the psoralens used in PUVA treatments for psoriasis are activated by near-ultraviolet light (400 nanometers), which barely penetrates the skin.

can be given six or seven times over a three-year period without damaging a healthy retina. For people with the most aggressive form of AMD (with mostly "classic" lesions), verteporfin halved the risk of moderate or serious vision loss over a two-year period. The effect is sustained over at least three or four years: patients who are not treated lose as much vision in three months as those treated with verteporfin lose in three years. The treatment also worked, though not as well, for people with less aggressive types of AMD and for those with related diseases such as pathologic myopia and ocular histoplasmosis syndrome. Only a small proportion of patients suffered from sunburns or other adverse reactions, rarely more than 24 hours after the procedure was done.

Some participants in the trials gained little benefit from PDT. For many of them, the disease may have already progressed too far. A reanalysis of clinical data presented by Bressler in April 2002 at the International Congress of Ophthalmology in Sydney, Australia, showed that smaller lesions respond much better to treatment than older, larger ones, implying that early detection and treatment may optimize the benefits of PDT.

Other Treatment Avenues

THE SUCCESS OF OPHTHALMIC PDT has inspired research activity in other fields but also reveals the drawbacks of the treatment. In particular, even red light penetrates no more than a few centimeters into biological tissues [*see illustration above*]. This limitation threatens the utility of PDT in internal medicine—its significance might seem to be skin deep. There are ways of turning PDT inward, however. One ingenious idea is called photoangioplasty, which is now being used to treat coronary artery disease.

Coronary angioplasty is a minimally invasive procedure for treating arteries affected by atherosclerosis. It uses a tiny balloon to open arteries, so that atherosclerotic plaques do not occlude the entire vessel. Photoangioplasty could sidestep many of the problems of conventional angioplasty, notably the restenosis (renarrowing) of treated arteries. The procedure involves injecting a porphyrin into the bloodstream, waiting for it to build up in the damaged arterial walls and then illuminating the artery from the inside, using a tiny light source attached to the end of a catheter. The light activates the porphyrins in the plaques, destroying the abnormal tissues while sparing the normal walls of the artery. The results of a small human trial testing the safety of the synthetic porphyrin motexafin lutetium were presented in March 2002 by Jeffrey J. Popma of Brigham and Women's Hospital at the annual meeting of the American College of Cardiology. Although it is still early in the testing process, the findings fuel hopes for the future: the procedure was safe, and its success at preventing restenosis increased as the dosage increased.

Accumulation of porphyrins in active and proliferating cells raises the possibility of treating other conditions in which abnormal cell activation or proliferation plays a role—among them, infectious diseases. Attempts to treat infections with the pigments had long been frustrated by a limited effect on gram-negative bacteria, which have a complex cell wall that obstructs the uptake of porphyrins into these organisms. One solution, developed by Michael R. Hamblin and his colleagues at Harvard, involved attaching a polymer—usually polylysine, a repetitive chain of the amino acid lysine—to the porphyrin. The polymer disrupts the lipid structure of the bacterial cell wall, enabling the porphyrins to gain entry to the cell. Once inside, they can be activated by light to kill the bacteria. In recent studies of animals with oral infections and infected wounds, the altered porphyrin showed potent antimicrobial activity against a broad spectrum of gram-negative and gram-positive bacteria. As antibiotic resistance becomes more intractable, targeted antimicrobial PDT could become a useful weapon in the medical arsenal.

Several other, related photodynamic methods hinge on the finding that activated immune cells take up greater amounts of photosensitizing drugs than do quiescent immune cells and red blood cells, sparing the quiet cells from irreversible damage. In most infections, nobody would wish to destroy activated immune cells: they are, after all, responsible for the body's riposte to the infection. In these cases, targeting immune cells would be equivalent to "friendly fire" and would give the infection free rein to pillage the body.

In AIDS, however, the reverse is true. The AIDS virus, HIV, infects the immune cells themselves. Targeting infected immune cells would then be more like eliminating double agents. In the laboratory, HIV-infected immune cells take up porphyrins, thereby becoming vulnerable to light treatment. In patients, the light could be applied either by withdrawing blood, illuminating it and transfusing it back into the body (extracorporeal phototherapy) or by shining red light onto the skin, in what is called transdermal phototherapy. In the transdermal approach, light would eliminate activated immune cells in the circulation as

HYBRID MEDICAL ANIMATION

Photodynamic Therapies

THE LIGHT-ACTIVATED drugs listed below are a sampling of those on the market or in development.

DRUG	TARGET	MAKER	STATUS
Levulan (5-aminolevulinic acid)	Acne and actinic keratosis (a precancerous skin disorder), Barrett's esophagus (a precancerous condition)	DUSA PHARMACEUTICALS Toronto	On the market for actinic keratosis; Phase II trials (relatively small studies in humans) have been completed for Barrett's esophagus
Photofrin (porfimer sodium)	Cancers of the esophagus and lung, high-grade dysplasia from Barrett's esophagus	AXCAN SCANDIPHARM Birmingham, Ala.	On the market for esophageal cancer and nonsmall cell lung cancer; awaiting FDA decision on high-grade dysplasia
Visudyne (verteporfin)	Age-related macular degeneration, pathologic myopia and ocular histoplasmosis (eye disorders)	QLT, INC., and NOVARTIS OPHTHALMICS Vancouver, B.C., and Duluth, Ga.	On the market
Metvix (methylaminolevulinic acid)	Actinic keratosis, basal cell skin cancer and squamous cell skin cancer	PHOTOCURE Oslo, Norway	Awaiting final FDA approval for actinic keratosis; in Phase III trials (large studies of efficacy) for skin cancers
PhotoPoint SnET2 (tin ethyl etiopurpurin)	Age-related macular degeneration	MIRAVANT MEDICAL TECHNOLOGIES Santa Barbara, Calif.	Phase III trials have been completed
verteporfin	Basal cell cancer, androgenetic alopecia (male pattern baldness) and prostatic hyperplasia (enlarged prostate)	QLT, INC.	In Phase III trials for basal cell cancer; as QLT0074, in Phase I trials (tests of safety in small numbers of patients) for other conditions
PhotoPoint MV9411 (contains indium)	Plaque psoriasis	MIRAVANT MEDICAL TECHNOLOGIES	In Phase II trials
Antrin (motexafin lutetium)	Diseased arteries	PHARMACYCLICS Sunnyvale, Calif.	Phase II trials for peripheral artery disease and Phase I trials for coronary artery disease have been completed
Lutrin (motexafin lutetium)	Cancerous tumors	PHARMACYCLICS	In Phase I trials for prostate cancer and cervical intraepithelial neoplasia

they passed through the skin. Whether the technique will be potent enough to eliminate diseased immune cells in HIV-infected patients remains an open question.

Autoimmune diseases, rejection of organ transplants, and leukemias are also all linked by the common thread of activated and proliferating immune cells. In autoimmune diseases, components of our own body erroneously activate immune cells. These activated clones then proliferate in an effort to destroy the perceived threat—say, the myelin sheath in multiple sclerosis or the collagen in rheumatoid arthritis. When organs are implanted, activated immune cells may multiply to reject the foreign tissue—the transplanted organ or even the body tissues of the new host, in the case of bone marrow transplants. In leukemia, immune cells and their precursors in the bone marrow produce large numbers of nonfunctional cells. In each instance, PDT could potentially eliminate the unwanted immune cells, while preserving the quiescent cells, to maintain a normal immune response to infection. As in HIV infection, the procedure might work either extracorporeally or transdermally. Much of this research is in late-stage preclinical or early clinical trials. For all the cleverness in exploring possible medical applications, though, we can only hope that more extensive clinical studies will bear fruit. SA

MORE TO EXPLORE

The Colours of Life: An Introduction to the Chemistry of Porphyrins and Related Compounds. L. R. Milgrom. Oxford University Press, 1997.

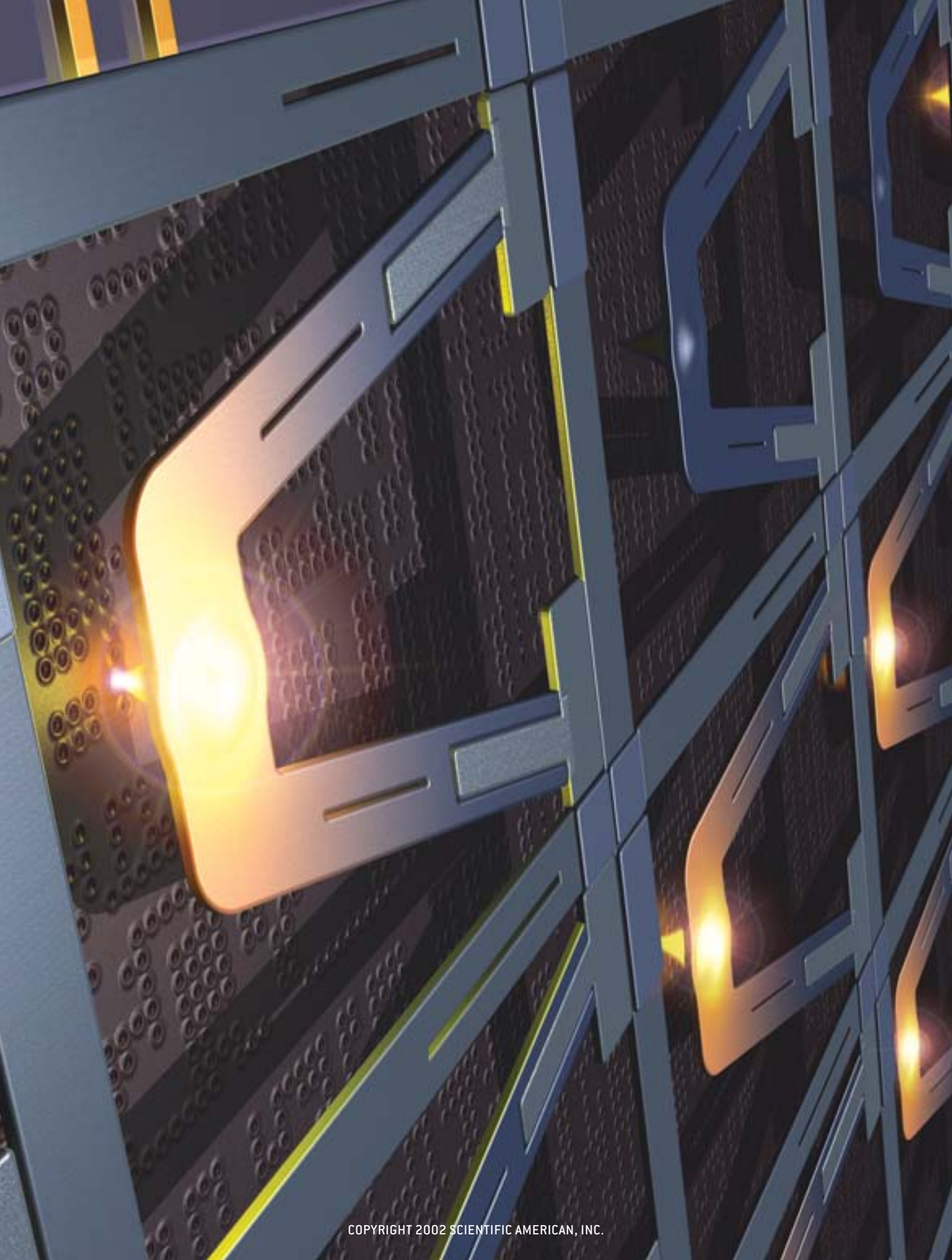
Lethal Weapon. P. Moore in *New Scientist,* Vol. 158, No. 2130, pages 40–43; April 18, 1998.

Verteporfin Therapy for Subfoveal Choroidal Neovascularization in Age-Related Macular Degeneration: Three-Year Results of an Open-Label Extension of 2 Randomized Clinical Trials. TAP Report No. 5. M. S. Blumenkranz et al. in *Archives of Ophthalmology,* Vol. 120, No. 10, pages 1307–1317; October 2002.

Oxygen: The Molecule That Made the World. Nick Lane. Oxford University Press (in press).

An overview of the nature of and treatments for porphyria can be found at **www.sciam.com/explore_directory.cfm**

COMPILED BY TARIQ MALIK



THE NANODRIVE PROJECT

INVENTING A NANOTECHNOLOGY DEVICE FOR MASS PRODUCTION AND CONSUMER USE IS TRICKIER THAN IT SOUNDS

By Peter Vettiger and Gerd Binnig

MAKING TRACKS: Arrays of cantilever-mounted tips inscribe millions of digital bits on a plastic surface in an exceedingly small space.

Many engineers have had the thrill of designing a novel product that then enters mass production and pops up all over the world. We hope—in fact, we would lay better than 50–50 odds on it—that within three years we will experience the rarer pleasure of having launched an entirely new class of machine.

Nanotechnology is much discussed these days as an emerging frontier, a realm in which machines operate at scales of billionths of a meter. Research on microelectromechanical systems (MEMS)—devices that have microscopic moving parts made using the techniques of computer chip manufacture—has sim-

ilarly produced a lot of hype and yet relatively few commercial products. But as we can attest, having spent six years so far on one of the first focused projects to create a nanomechanical device suitable for mass production, at such tiny scales, engineering is inextricably melded with scientific research. Unexpected obstacles appear on the road from a proof-of-principle experiment to a working prototype and then on to a product that succeeds in the marketplace.

Here at IBM we call our project Millipede. If we stay on track, by 2005 or so you will be able to buy a postage stamp–size memory card for your digital camera or portable MP3 player. It will hold not just a few dozen megabytes of video or audio, as typical flash memory cards do, but several gigabytes—sufficient to store an entire CD collection of music or several feature films. You will be able to erase and rewrite data to the card. It will be quite fast and will require moderate amounts of power. You might call it a nanodrive.

That initial application may seem interesting but hardly earth-shaking. After all, flash memory cards with a gigabyte of capacity are already on the market. The impressive part is that Millipede stores digital data in a completely different way from magnetic hard disks, optical compact discs and transistor-based memory chips. After decades of spectacular progress, those mature technologies have entered the home stretch; imposing physical limitations loom before them.

The first nanomechanical drives, in contrast, will barely scratch the surface of their potential. Decades of refinements will lie ahead. In principle, the digital bits that future generations of Millipede-like devices will write, read and erase could continue to shrink until they are individual molecules or even atoms. As the moving parts of the nanodrives get smaller, they will work faster and use power more efficiently. The first products to use Millipede technology will most likely be high-capacity data storage cards for cameras, mobile phones and other portable devices. The nanodrive cards will function in much the same way as the flash memory cards in these devices today but will offer several-gigabyte capacity for lower cost. The same technology might also be a tremendous boon to materials science research, biotechnology or even applications that are not currently foreseeable.

It was this long-term promise that got us so excited half a dozen years ago. Along the way, we learned that sometimes the only way around a barrier is a serendipitous discovery. Fortunately, besides unexpected obstacles, there are also unexpected gifts. It seems there often is a kind of reward from nature if one dares enter new areas. On the other hand, sometimes nature is not so kind, and you must overcome the difficulty yourself. We have worked hard on such problems but not too hard. If at one stage we had no idea how to address an issue, perhaps a year later we found an answer. Good intuition is required in such cases, in which you expect the problem can be solved, although you do not yet know how.

A Crazy Idea

IN A WAY, Millipede got its start on a soccer field. The two of us played on the soccer team of the IBM Zurich Research Laboratory, where we work. We were introduced by another teammate, Heinrich Rohrer. Rohrer had started at the Zurich lab in 1963, the same year as one of us (Vettiger); he had collaborated with the other one (Binnig) on the invention in 1981 of the scanning tunneling microscope (STM), a technology that led to the long-sought ability to see and manipulate individual atoms.

In 1996 we were both looking for a new project in a considerably changed environment. The early 1990s had been a tough time for IBM, and the company had sold off its laser science effort, the technology part of which was managed by Vettiger. Binnig had closed his satellite lab in Munich and moved back to Zurich. Together with Rohrer, we started brainstorming ways to apply STM or other scanning probe techniques, specifically atomic force microscopy (AFM), to the world beyond science.

AFM, invented by Binnig and developed jointly with Christoph Gerber of the Zurich lab and Calvin F. Quate of Stanford University, is the most widely used local probe technique. Like STM, AFM took a radically new approach to microscopy. Rather than magnifying objects by using lenses to guide beams of light or by bouncing electrons off the object, an AFM slowly drags or taps a minuscule cantilever over an object's surface. Perched on the end of the cantilever is a sharp tip tapered to a width of less than 20 nanometers—a few hundred atoms. As the cantilever tip passes over the dips and rises in the surface (either in contact with or in extreme proximity to it), a computer translates the deflection of the lever into an image, revealing, in the best cases, each passing atom.

While Binnig was making the first images of individual silicon atoms in the mid-1980s, he inadvertently kept bumping the tip into the surface, leaving little dents in the silicon. The possibility of using an STM or AFM as an atomic-scale data storage device was obvious: make a dent for a 1, no dent for a 0. But the difficulties were clear, too. The tip has to follow the contours of the medium mechanically, so it must scan very slowly compared with the high-speed rotation of a hard-disk platter or the nanosecond switching time of transistors.

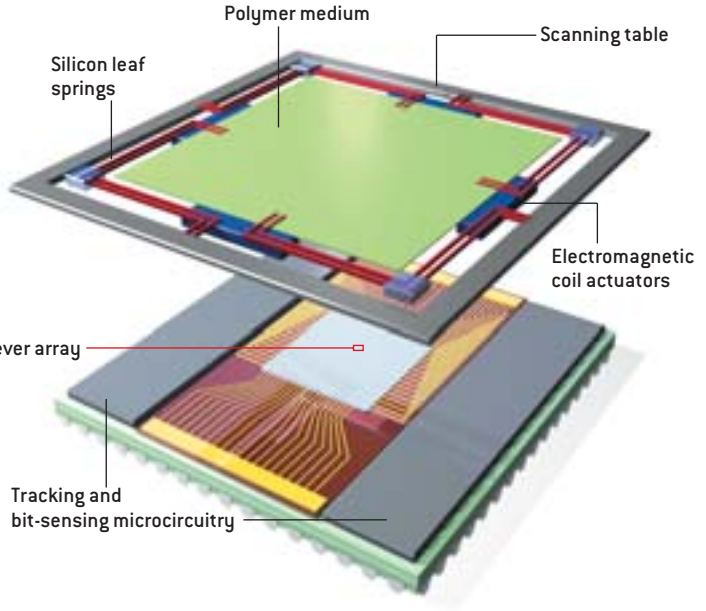
Overview/*Millipede Project*

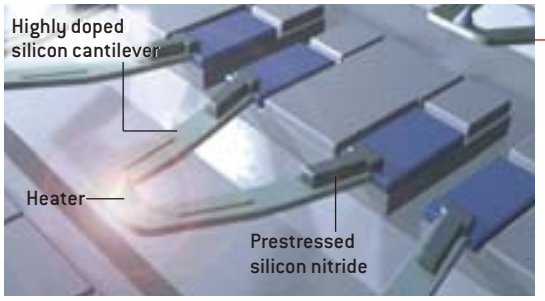
- Today's digital storage devices are approaching physical limits that will block additional capacity. The capabilities of the Millipede "nanodrive"—a micromechanical device with components in the nano-size range—could take off where current technologies will end.
- Millipede uses grids of tiny cantilevers to read, write and erase data on a polymer media. The cantilever tips poke depressions into the plastic to make digital 1's; the absence of a dent is a digital 0.
- The first Millipede products, most likely postage stamp–size memory cards for portable electronic devices, should be available within three years.

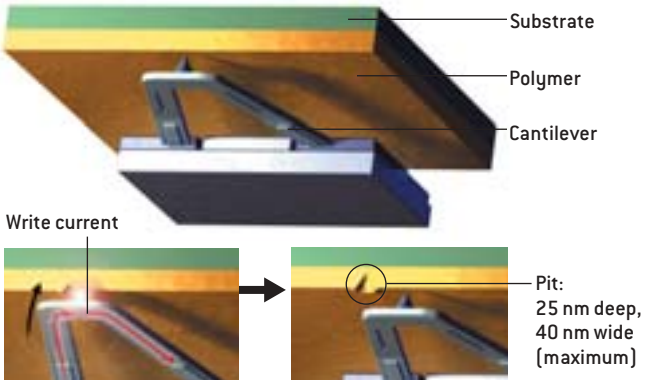
SLIM FILMS (*preceding pages*)

HOW THE NANODRIVE WORKS

THE MILLIPEDE NANODRIVE prototype operates like a tiny phonograph, using the sharp tips of minuscule silicon cantilevers to read data inscribed in a polymer medium. An array of 4,096 levers, laid out in rows with their tips pointing upward, is linked to control microcircuitry that converts information encoded in the analog pits into streams of digital bits. The polymer is suspended on a scanning table by silicon leaf springs, which permits tiny magnets (*not shown*) and electromagnetic coils to pan the storage medium across a plane while it is held level over the tips. The tips contact the plastic because the levers flex upward by less than a micron.

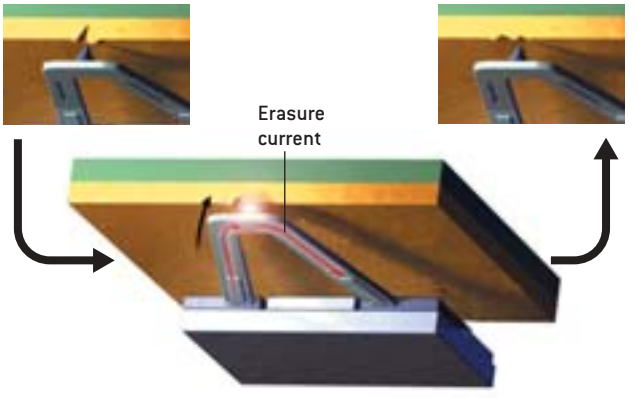






WRITING A BIT

Using heat and mechanical force, tips create conical pits in linear tracks that represent series of digital 1's. To produce a pit, electric current travels through the lever, heating a doped region of silicon at the end to 400 degrees Celsius, which allows the prestressed lever structure to flex into the polymer. The absence of a pit is a 0.

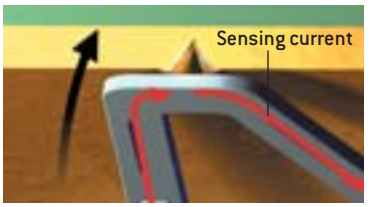


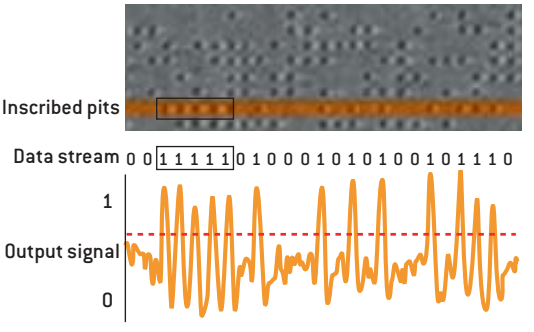
ERASING A BIT

The latest Millipede prototype erases an existing bit by heating the tip to 400 degrees C and then forming another pit just adjacent to the previously inscribed pit, thus filling it in (*shown*). An alternative erasure method involves inserting the hot tip into a pit, which causes the plastic to spring back to its original flat shape.

READING A BIT

To read data, the tips are first heated to about 300 degrees C. When a scanning tip encounters and enters a pit (*below*), it transfers heat to the plastic. Its temperature and electrical resistance thus fall, but the latter by only a tiny amount, about one part in a few thousand. A digital signal processor converts this output signal, or its absence, into a data stream (*far right*).





BRYAN CHRISTIE DESIGN

Other pros and cons soon became apparent. Because of the extremely small mass of the cantilevers, AFM operation with the tip in direct contact with the medium is much faster than that of an STM or a noncontact AFM, though still not as fast as magnetic storage. On the other hand, tips of a contact AFM wear quickly when used to scan metal surfaces. And—perhaps most important—once the tip has made a dent, there was no obvious way to "erase" it.

A group led by Dan Rugar at the IBM Almaden Research Center in San Jose, Calif., had tried shooting laser pulses at the tip to heat it; that would in turn soften the plastic so the tip could dent it. The group was able to create compact disc–like recordings that stored data more densely than even today's digital video discs (DVDs) do. It also performed extensive wear tests with very promising results. But the system was too slow, and it still lacked a technique to erase and rewrite data.

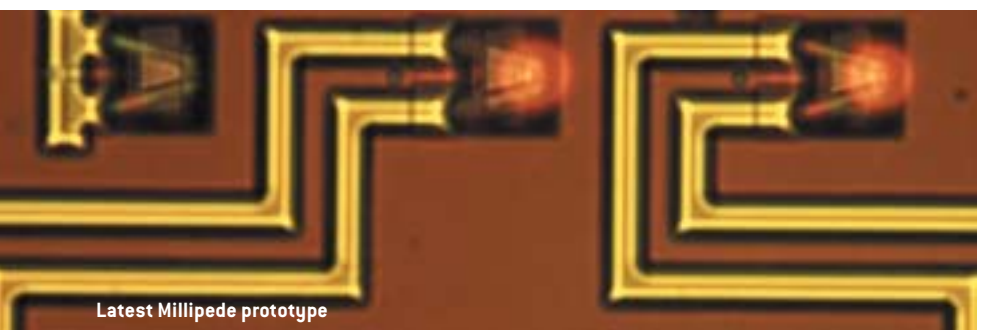
Our team sketched out a design that we thought could supply these missing ingredients. Rather than using just one cantilever, why not exploit chipmakers' ability to construct thousands or even millions of identical microscopic parts on a thumbnail-size slice of silicon? Working together in parallel—like the legs of a millipede—an army of slow AFM tips could read or write data quite rapidly.

Here more imagination was required to envision a chance for success than to come up with the idea itself. Although operating a single AFM is sometimes difficult, we were confident that a massively parallel device incorporating many tips would have a realistic chance of functioning reliably.

As a start, we needed at least one way to erase, be it elegant or not. Alternatives, we thought, might pop up later. We developed a scheme of erasing large fields of bits. We heated them above the temperature at which the polymer starts to flow, in much the same way as the surface of wax gets smooth when warmed by a heat gun. Although the technique worked nicely, it was somewhat complicated because, before erasing a field, all the data that were to be retained had to be transferred into another field. (Later on, as we'll explain, nature presented a much better method.)

With these rough concepts in mind, we started our journey into an interdisciplinary project. With the pair of us working in one team, we bridged two IBM departments, physics and devices. (They were eventually merged into a single science and technology department.) We were also joined by Evangelos Eleftheriou and his team, from our laboratory's communication systems department. Today several other groups from within IBM Research and from universities collaborate with us.

When different cultures meet, misunderstandings cannot be prevented, at least not in the beginning. We, however, experienced a huge benefit from mixing disparate viewpoints.



Latest Millipede prototype

99 Percent Perspiration ...

WE WERE NOT MEMS experts, and researchers in the MEMS and scanning probe technology communities had so far dismissed our project as harebrained. Although others, such as Quate's group at Stanford, were working at that time on STM- or AFM-based data storage, ours was the only project committed to large-scale integration of many probes. We hoped to achieve a certain vindication by presenting a working prototype in January 1998 at the IEEE 11th International Workshop on Micro-Electro-Mechanical Systems in Heidelberg, Germany. Instead we had a nearly working prototype to show. We presented a five-by-five array of tips in an area of 25 square millimeters.

It was able to demonstrate parallel imaging, but parallel writing failed. We had overlooked a niggling but critical technical detail: the wires leading to the heaters were metallic and too thin to handle the current passing through them. They immediately blew like overloaded fuses because of the phenomenon of electromigration in metal films. Electromigration was well described in the literature, and we should have known about it. This was not our only mistake, but in our group mistakes can be admitted and quickly corrected.

Despite the setbacks, our lab's managers sensed real progress. They allowed us to double the size of our team to eight. We had learned from the 25-tip array that the aluminum wiring had to be replaced—which we did with highly doped silicon cantilevers. We also found that it was possible to level out the tip array below the storage medium with high precision in a relatively large area, which made us confident enough to move to a bigger array right away.

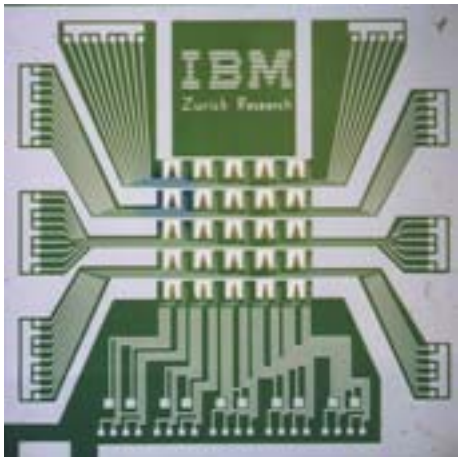
Vettiger recognized one serious problem in May 1998 as he was giving an invited talk at the IBM Almaden lab. He was describing how the cantilevers would be arranged in regular rows and columns,

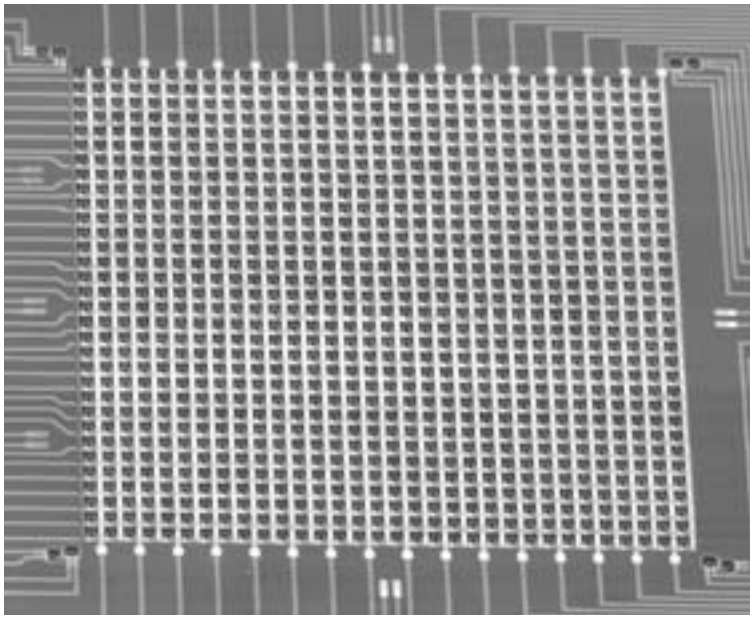
Researchers in the MEMS and scanning probe technology communities had dismissed our project as harebrained.

IBM

MORE TIPS IN A SMALLER SPACE

PROTOTYPE EVOLUTION: Whereas the first-generation Millipede chip contained an array of 25 cantilevers on a square that was five millimeters on a side (*below*), the succeeding prototype (*right*) incorporated 1,024 cantilevers in a smaller, three-millimeter square.





all of them connected to a grid of electrical wires. But as he was explaining how this system would work, he suddenly realized that it wouldn't. Nothing would stop the electric current from going everywhere at once; there would thus be no way to reliably send a signal to an individual cantilever.

The uncontrolled flow of current is actually a well-known phenomenon when units in an array have to be addressed through columns and rows. A common solution is to attach a transistor switch to each unit. But putting transistors on the same chip at the tips was not an option; the tips must be sharpened under intense heat that would destroy tiny transistors. Back at the lab, we tried all kinds of tricks to improve control of the current flow—none of which pleased Vettiger. The bigger the array, the more serious this flaw became. A quick calculation and simulation by Urs Dürig of our team showed that for an array of 1,000 units, addressing single cantilevers for writing would still be possible; reading the small signals of individual levers, however, would fail.

Vettiger slept poorly that night, fretting. The team was just about to complete the chip design for a 1,024-tip array. Vettiger told them to wait. For days the team agonized over the problem, until at last Vettiger and Michel Despont saw a practical answer: place a Schottky diode (an electrical one-way street) next to each cantilever. This highly nonlinear device would block the undesired current from flowing into all the other cantilevers. We reworked the design and soon had a 32-by-32-tip array, our second prototype.

This prototype proved that many of our ideas would work. All 1,024 cantilevers in the array came out intact and bent up by just the right amount so that they applied the correct amount of force when mated to a soft polymer medium called PMMA, which is mounted on a separate chip called a scanning table. Copper electromagnetic coils placed behind the scanning table were able to keep the PMMA surface from tilting too much as it panned left, right, back and forward atop the cantilever tips. (A new media scanner designed by Mark Lantz and Hugo Rothuizen has since reduced vibration sensitivity, which was then a problem.) Each 50-micron-long cantilever had a little resistor at its end. An electrical pulse sent through the tip heated it to around 400 degrees Celsius for a few microseconds.

The initial results with our second prototype were encouraging. More than 80 percent of the 1,024 levers worked properly on first pass, and there was only one narrow "dark" zone crossing the center of the storage field, resulting from a twisting of the chip when it was mounted. Not in our wildest dreams did we expect such success at this early stage of the project.

THE AUTHORS

PETER VETTIGER and *GERD BINNIG* have collaborated extensively to refine technologies for the Millipede nanodrive concept. Vettiger has had a long-standing career as a technologist specializing in microfabrication and nanofabrication. He joined the IBM Zurich Research Laboratory in 1963 and graduated in 1965 with a degree in communications technology and electronics engineering from the Zurich University of Applied Sciences. His academic career culminated in an honorary Ph.D. awarded in 2001 by the University of Basel. Binnig completed his Ph.D. in physics in 1978 at the Johann Wolfgang Goethe University in Frankfurt, Germany, and joined the Zurich lab that same year. His awards for outstanding scientific achievements include the 1986 Nobel Prize for Physics, which he received together with Heinrich Rohrer for the invention of the scanning tunneling microscope.

IBM

From R to D

IN THE FIVE-BY-FIVE DEVICE, each lever had at its base a piezoresistive sensor that converted mechanical strain to a change in resistance, allowing the system to detect when the tip had dropped into a pit—a digital 1. We began exploring approaches to detect pits more definitively. We ran tests with Schottky diodes integrated into the cantilevers, hoping that the strain would modify their resistance. Somehow the diodes did not have the expected properties. We nonetheless observed a strong signal when a bit was sensed. After some head-scratching, we found the surprising reason. It turned out to be a thermal phenomenon. If the cantilever is preheated to about 300 degrees C, not quite hot enough to make a dent, its electrical resistance drops significantly whenever the tip falls into a pit [*see illustration on page 49*]. We never would have thought to use a thermal effect to measure a motion, deflection or position. On macro scales, doing so would be too slow and unreliable because of convection—the circulatory motion that occurs in a fluid medium, in this case air, as heat is transferred between two objects of different temperatures. On the micro scale, however, turbulence does not exist, and hotter and cooler objects reach equilibrium within microseconds.

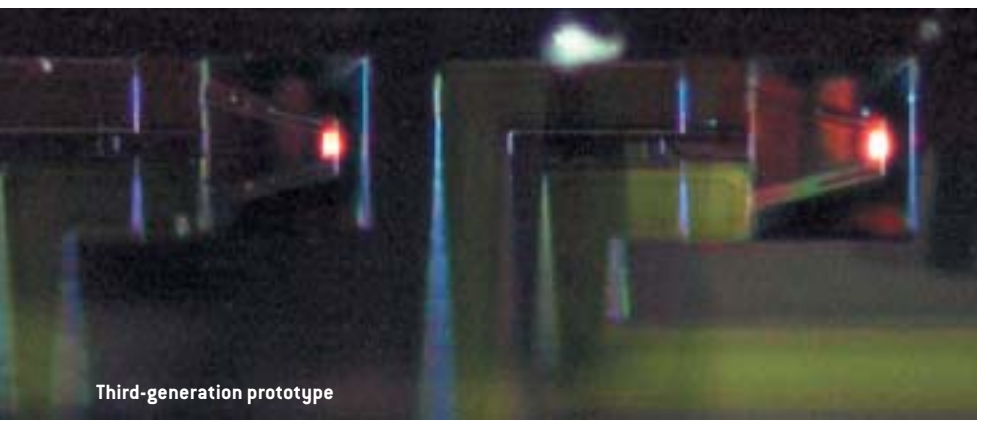
Although this result was unexpected, it was very useful. Now we could use the same heater on each lever for reading bits as well as writing them. Instead of three or four wires per cantilever, only two would be needed.

We presented this second prototype at the 1999 IEEE MEMS conference. This time the other researchers in attendance were more impressed. But what really excited upper managers at IBM were pictures of regular rows of pits that Millipede had written into the polymer. The pits were spaced just 40 nanometers apart—about 30 times the density of the best hard drives then on the market.

Shortly thereafter, in early 2000, the Millipede project changed character. We began focusing more on producing a storage system prototype. The team grew to about a dozen workers. We again brought together two departments, with Eleftheriou and his team joining us. They contributed their unique expertise in recording-channel technology, which they had been applying to magnetic recording very successfully. They began developing the electronic part of a fully functional system prototype—from basic signal processing and error-correction coding to complete system architecture and control.

We had just discovered a way to erase a small area, and in cooperation with Eleftheriou, we could even turn it into a system in which no erasing is required before overwriting. In the new, local erasure method, when the tip temperature is high enough to soften the material, surface tension and the springiness of the polymer cause a pit to pop up again. Instead of annealing a larger field using a heater integrated into the storage substrate—as in the block erasure method described earlier—the tip heats the medium locally. Because of electrostatic forces, a certain loading force on the tip cannot be avoided. So when the tip is heated to a high enough temperature and a new indentation is produced, older bits in close proximity are erased at the same time. If a row of pits is written densely, each newly created bit will eliminate the previous one, and only the last bit in the row will remain. This mechanism can even be used to overwrite old data with new code without knowing what the old one was. In a marriage of our experience in physics with Eleftheriou's recording-channel expertise, we developed a special form of constrained coding for such direct overwriting.

At that point it was clear that the team needed to work on the speed and power efficiency of Millipede. We had to start measuring signal-to-noise ratios, bit error rates and other indicators of how well the nanodrive could record digital data. And we had to choose a size and shape for the nanodrive. The "form factor" can be all-important in the consumer electronics marketplace, specifically in the mobile area, which we had chosen to address first.



Third-generation prototype

Although we scientists no longer consider this a high-risk project, we still rejoice when a new prototype works.

The Road Ahead

IN THE LAST MONTHS of 2002 our group put the final touches on the third-generation prototype, which has 4,096 cantilevers arranged in a 64-by-64 array that measures 6.4 millimeters on a side. Cramming more levers onto a chip is challenging but doable. Today we could fabricate chips with one million levers, and 250 such arrays could be made from a standard 200-millimeter wafer of silicon. The primary task now is to strike the right balance between two desiderata.

IBM

First, our design for a complete nanodrive system—not just the array and the scanning table but also the integrated microelectronics that control them—should be inexpensive enough to be immediately competitive and, especially for handheld devices, operable at low power. But it is critical that the system function dependably despite damage that occurs during years of consumer use.

We have found polymers that work even better than PMMA does. In these plastics, pits appear to be stable for at least three years, and a single spot in the array can be written and erased 100,000 times or more. But we are less sure about how the tips will hold up after making perhaps 100 billion dents over several years of operation. Dürig and Bernd Gotsmann of our team are working closely with colleagues at IBM Almaden to modify existing polymers or develop new ones that meet the requirements for our storage application.

And although human eyes scanning an image of the Millipede medium can easily pick out which blocks in the grid contain pits and which do not, it is no trivial matter to design simple electronic circuitry that does the same job with near-perfect accuracy. Detecting which bits represent 0's and which are 1's is much easier if the pits are all the same depth and are evenly spaced along straight tracks. That means that the scanning table must be made flat, held parallel to the tips and panned with steady speed in linear motion—all to within a few nanometers' tolerance. Just recently, we learned that by suspending the scanning table on thin leaf springs made of silicon, we gain much better control of its movement. Even so, we will add an active feedback system that is very sensitive to the relative position of the two parts to meet such nanoscopic tolerances while the device is jostling around on a jogger's waistband.

Any mechanical system such as Millipede that generates heat has to cope with thermal expansion. If the polymer medium and the silicon cantilevers differ by more than about a single degree C, the alignment of the bits will no longer match that of the tips. A feedback system to compensate for misalignment would add complexity and thus cost. We are not yet certain of the best solution to this problem.

Fortunately, nature has helped again. Millipede and the storage substrate carrying the polymer film are both made from silicon and will therefore expand by the same amount if they are at the same temperature. Additionally, the gap between the tip array and the substrate is so small that the air trapped between them acts as an excellent heat conductor, and a temperature difference between them is hardly achievable.

Because the project has now matured to the point that we can begin the first steps toward product development, our team has been joined by Thomas R. Albrecht, a data storage technologist from IBM Almaden who helped to shepherd IBM's Microdrive to market. Bringing the Microdrive from the lab to the customer was a journey similar to what Millipede may face in the next few years.

For the members of our group, this transition to product development means that we will surrender the Millipede more and more to the hands of others. Stepping back is the most difficult part and, at the same time, the most critical to the success of a project.

Indeed, we cannot yet be certain that the Millipede program will result in a commercial device. Although we scientists no longer consider this a high-risk project, we still rejoice when a new prototype works. If we are lucky, our newest prototypes will reveal problems that our team knows how to fix.

In any case, we are excited that, at a minimum, this nanomechanical technology could allow researchers for the first time to scan a square centimeter of material with near-atomic resolution. So far the project has generated close to 30 relatively basic patents. No one knows whether nanodrives will make it in the market. But they will be a new class of machine that is good for something, and for us that is its own reward. SA

High-Density Memory Projects

IBM'S MILLIPEDE PROJECT is only one of several efforts to bring compact, high-capacity computer memories to market.

COMPANY	DEVICE TECHNOLOGY	MEMORY CAPACITY	COMMERCIALIZATION
Hewlett-Packard *Palo Alto, Calif.*	Thumbnail-size atomic force microscope (AFM) device using electron beams to read and write data onto recording area	At least a gigabyte (GB) at the outset	End of the decade
Hitachi *Tokyo*	AFM-based device; specifics not disclosed	Has not been revealed	Has not been revealed
Nanochip *Oakland, Calif.*	AFM-tipped cantilever arrays that store data on a silicon chip	Half a GB at first; potential for 50 GBs	Expected in 2004
Royal Philips Electronics *Eindhoven, the Netherlands*	Optical system similar to rerecordable CDs using a blue laser to record and read data on a three-centimeter-wide disk	Up to a GB per side, perhaps four GBs in all	Expected in 2004
Seagate Technology *Scotts Valley, Calif.*	Rewritable system using AFM or other method, operating on a centimeter-size chip	As many as 10 GBs on a chip for portables	Expected in 2006 or later

MORE TO EXPLORE

In Touch with Atoms. Gerd Binnig and Heinrich Rohrer in *Reviews of Modern Physics,* Vol. 71, No. 2, pages S324–S330; March 1999.

The "Millipede"—Nanotechnology Entering Data Storage. P. Vettiger, G. Cross, M. Despont, U. Drechsler, U. Dürig, B. Gotsmann, W. Häberle, M. A. Lantz, H. E. Rothuizen, R. Stutz and G. Binnig in *IEEE Transactions on Nanotechnology,* Vol. 1, No. 1, pages 39–55; March 2002.

For more about nanotechnology in IBM Research and elsewhere, see **www.research.ibm.com/pics/nanotech/**

COMPILED BY TARIQ MALIK

An Ancestor to Call Our Own

By Kate Wong

Controversial new fossils could bring scientists closer than ever to the origin of humanity

POITIERS, FRANCE—Michel Brunet removes the cracked, brown skull from its padlocked, foam-lined metal carrying case and carefully places it on the desk in front of me. It is about the size of a coconut, with a slight snout and a thick brow visoring its stony sockets. To my inexpert eye, the face is at once foreign and inscrutably familiar. To Brunet, a paleontologist at the University of Poitiers, it is the visage of the lost relative he has sought for 26 years. "He is the oldest one," the veteran fossil hunter murmurs, "the oldest hominid."

Brunet and his team set the field of paleoanthropology abuzz when they unveiled their find last July. Unearthed from sandstorm-scoured deposits in northern Chad's Djurab Desert, the astonishingly complete cranium—dubbed *Sahelanthropus tchadensis* (and nicknamed Toumaï, which means "hope of life" in the local Goran language)—dates to nearly seven million years ago. It may thus represent the earliest human forebear on record, one who Brunet says "could touch with his finger" the point at which our lineage and the one leading to our closest living relative, the chimpanzee, diverged.

KAZUHIKO SANO

APE OR ANCESTOR? *Sahelanthropus tchadensis,* potentially the oldest hominid on record, forages in a woodland bordering Lake Chad some seven million years ago. Thus far the creature is known only from cranial and dental remains, so its body in this artist's depiction is entirely conjectural.

Less than a century ago simian human precursors from Africa existed only in the minds of an enlightened few. Charles Darwin predicted in 1871 that the earliest ancestors of humans would be found in Africa, where our chimpanzee and gorilla cousins live today. But evidence to support that idea didn't come until more than 50 years later, when anatomist Raymond Dart of the University of the Witwatersrand described a fossil skull from Taung, South Africa, as belonging to an extinct human he called *Australopithecus africanus,* the "southern ape from Africa." His claim met variously with frosty skepticism and outright rejection—the remains were those of a juvenile gorilla, critics countered. The discovery of another South African specimen, now recognized as *A. robustus,* eventually vindicated Dart, but it wasn't until the 1950s that the notion of ancient, apelike human ancestors from Africa gained widespread acceptance.

In the decades that followed, pioneering efforts in East Africa headed by members of the Leakey family, among others, turned up additional fossils. By the late 1970s the australopithecine cast of characters had grown to include *A. boisei, A. aethiopicus* and *A. afarensis* (Lucy and her kind, who lived between 2.9 million and 3.6 million years ago during the Pliocene epoch and gave rise to our own genus, *Homo*). Each was adapted to its own environmental niche, but all were bipedal creatures with thick jaws, large molars and small canines—radically different from the generalized, quadrupedal Miocene apes known from farther back on the family tree. To probe human origins beyond *A. afarensis,* however, was to fall into a gaping hole in the fossil record between 3.6 million and 12 million years ago. Who, researchers wondered, were Lucy's forebears?

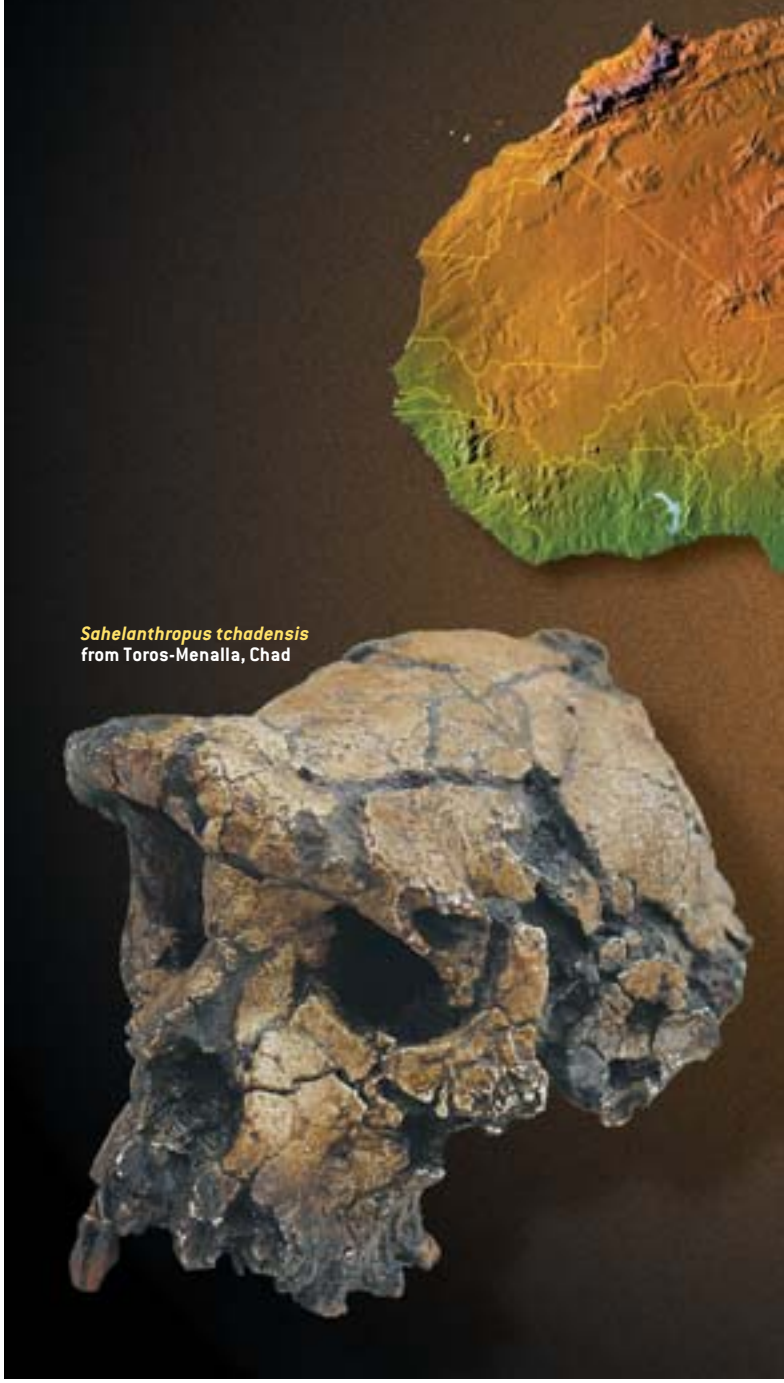
Despite widespread searching, diagnostic fossils of the right age to answer that question eluded workers for nearly two decades. Their luck finally began to change around the mid-1990s, when a team led by Meave Leakey of the National Museums of Kenya announced its discovery of *A. anamensis,* a four-million-year-old species that, with its slightly more archaic characteristics, made a reasonable ancestor for Lucy [see "Early Hominid Fossils from Africa," by Meave Leakey and Alan Walker; Scientific American, June 1997]. At around

Overview/*The Oldest Hominids*

- The typical textbook account of human evolution holds that humans arose from a chimpanzeelike ancestor between roughly five million and six million years ago in East Africa and became bipedal on the savanna. But until recently, hominid fossils more than 4.4 million years old were virtually unknown.
- Newly discovered fossils from Chad, Kenya and Ethiopia may extend the human record back to seven million years ago, revealing the earliest hominids yet.
- These finds cast doubt on conventional paleoanthropological wisdom. But experts disagree over how these creatures are related to humans—if they are related at all.

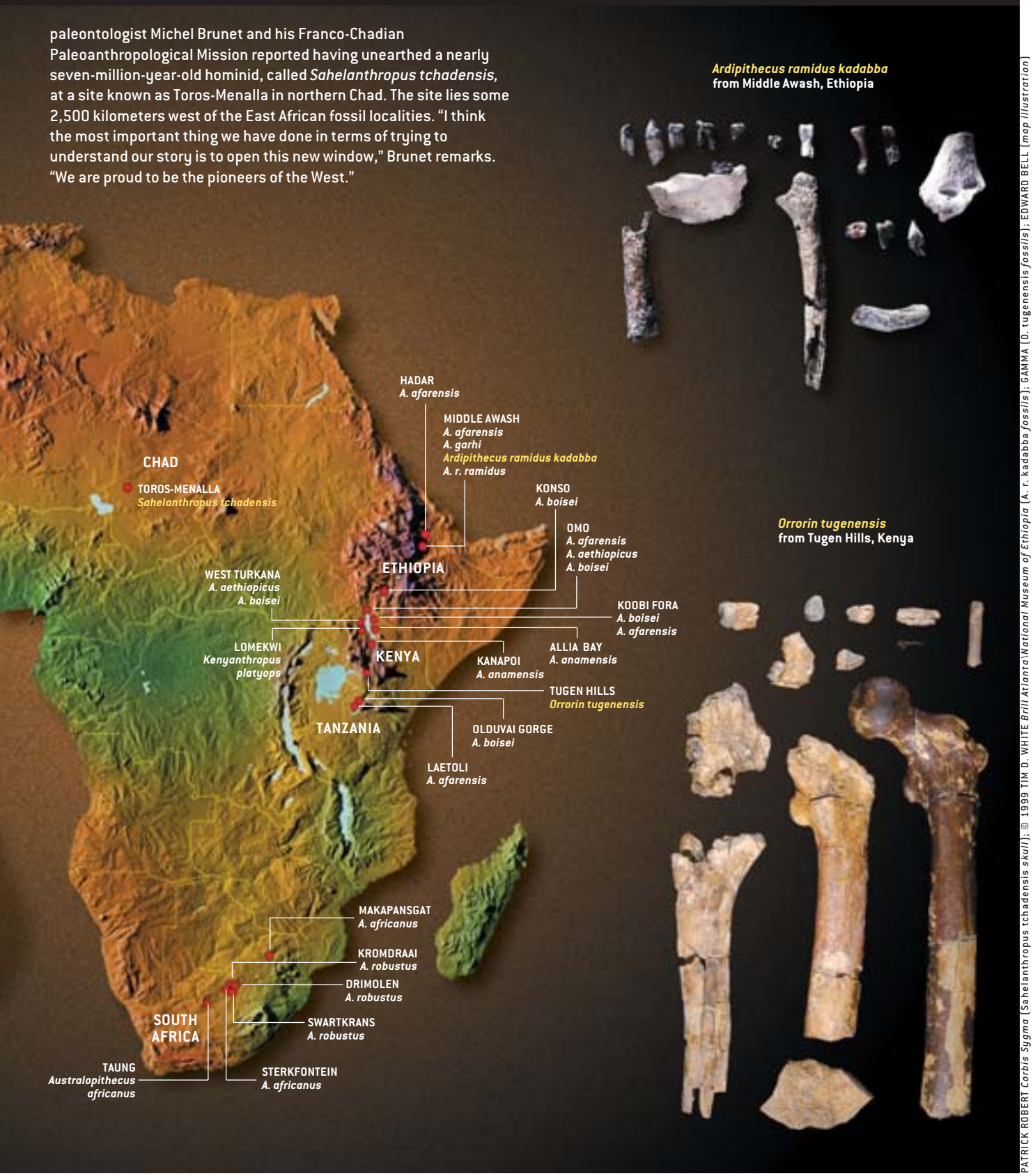
African Roots

RECENT FINDS from Africa could extend in time and space the fossil record of early human ancestors. Just a few years ago, remains more than 4.4 million years old were essentially unknown, and the oldest specimens all came from East Africa. In 2001 paleontologists working in Kenya's Tugen Hills and Ethiopia's Middle Awash region announced that they had discovered hominids dating back to nearly six million years ago (*Orrorin tugenensis* and *Ardipithecus ramidus kadabba,* respectively). Then, last July, University of Poitiers



Sahelanthropus tchadensis **from Toros-Menalla, Chad**

paleontologist Michel Brunet and his Franco-Chadian Paleoanthropological Mission reported having unearthed a nearly seven-million-year-old hominid, called *Sahelanthropus tchadensis,* at a site known as Toros-Menalla in northern Chad. The site lies some 2,500 kilometers west of the East African fossil localities. "I think the most important thing we have done in terms of trying to understand our story is to open this new window," Brunet remarks. "We are proud to be the pioneers of the West."



PATRICK ROBERT *Corbis Sygma* (*Sahelanthropus tchadensis skull*); © 1999 TIM D. WHITE *Brill Atlanta/National Museum of Ethiopia* (*A. r. kadabba fossils*); GAMMA (*O. tugenensis fossils*); EDWARD BELL (*map illustration*)

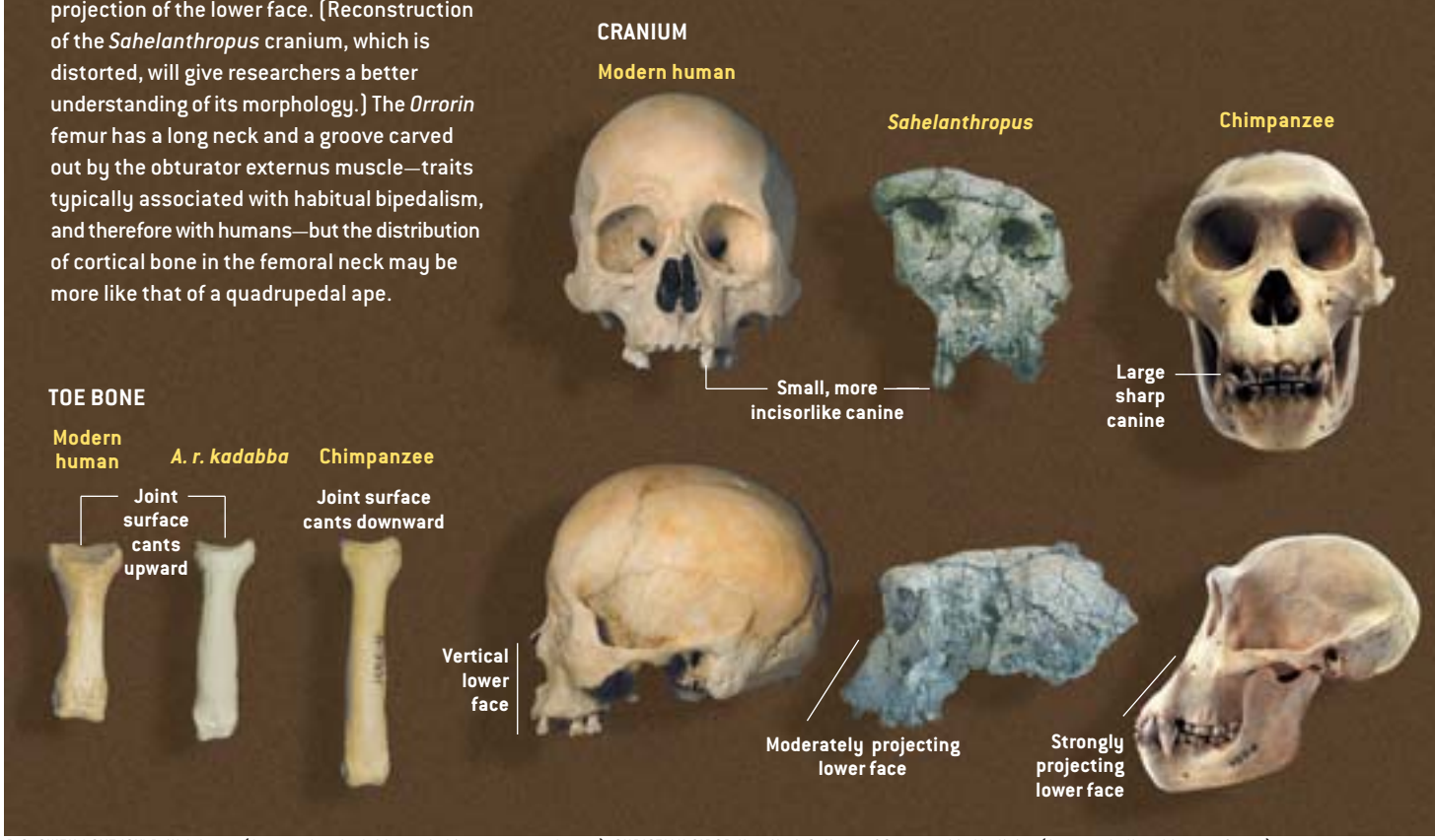
It is the visage of the lost relative he has sought for 26 years. ***"He is the oldest one,"*** *the veteran fossil hunter murmurs,* ***"the oldest hominid."***

the same time, Tim D. White of the University of California at Berkeley and his colleagues described a collection of 4.4-million-year-old fossils from Ethiopia representing an even more primitive hominid, now known as *Ardipithecus ramidus ramidus*. Those findings gave scholars a tantalizing glimpse into Lucy's past. But estimates from some molecular biologists of when the chimp-human split occurred suggested that even older hominids lay waiting to be discovered.

Those predictions have recently been borne out. Over the past few years, researchers have made a string of stunning discoveries—Brunet's among them—that may go a long way toward bridging the remaining gap between humans and their African ape ancestors. These fossils, which range from roughly five million to seven million years old, are upending long-held ideas about when and where our lineage arose and what the last common ancestor of humans and chimpanzees looked like. Not surprisingly, they have also sparked vigorous debate. Indeed, experts are deeply divided over where on the family tree the new species belong and even what constitutes a hominid in the first place.

Anatomy of an Ancestor

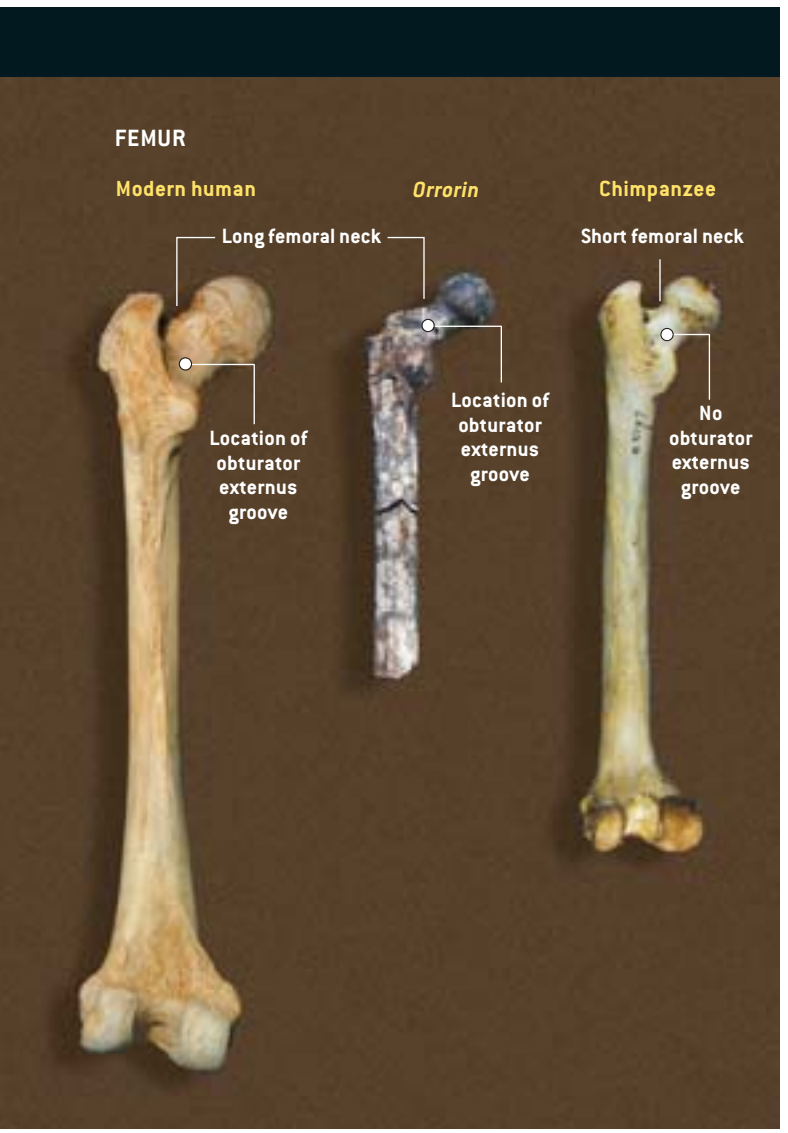
KEY TRAITS link putative hominids *Ardipithecus ramidus kadabba, Orrorin* and *Sahelanthropus* to humans and distinguish them from apes such as chimpanzees. The fossils exhibit primitive apelike characteristics, too, as would be expected of creatures this ancient. For instance, the *A. r. kadabba* toe bone has a humanlike upward tilt to its joint surface, but the bone is long and curves downward like a chimp's does (which somewhat obscures the joint's cant). Likewise, *Sahelanthropus* has a number of apelike traits—its small braincase among them—but is more humanlike in the form of the canines and the projection of the lower face. (Reconstruction of the *Sahelanthropus* cranium, which is distorted, will give researchers a better understanding of its morphology.) The *Orrorin* femur has a long neck and a groove carved out by the obturator externus muscle—traits typically associated with habitual bipedalism, and therefore with humans—but the distribution of cortical bone in the femoral neck may be more like that of a quadrupedal ape.



© C. OWEN LOVEJOY*Brill Atlanta* (*human,* A. r. kadabba *and chimpanzee toe bones*); CHRISTIAN SIDOR *New York College of Osteopathic Medicine* (*human skull and human femur*); MISSION PALÉOANTHROPOLOGIQUE FRANCO-TCHADIENNE (Sahelanthropus *skull*); © 1996 DAVID L. BRILL\DIVISION OF MAMMALS, NATIONAL MUSEUM OF NATURAL HISTORY, SMITHSONIAN INSTITUTION (*chimpanzee skull*); GAMMA (Orrorin *femur*); C. OWEN LOVEJOY *Kent State University* (*chimpanzee femur*)

Standing Tall

THE FIRST HOMINID CLUE to come from beyond the 4.4-million-year mark was announced in the spring of 2001. Paleontologists Martin Pickford and Brigitte Senut of the National Museum of Natural History in Paris found in Kenya's Tugen Hills the six-million-year-old remains of a creature they called *Orrorin tugenensis*. To date the researchers have amassed 19 specimens, including bits of jaw, isolated teeth, finger and arm bones, and some partial upper leg bones, or femurs. According to Pickford and Senut, *Orrorin* exhibits several characteristics that clearly align it with the hominid family—notably those suggesting that, like all later members of our group, it walked on two legs. "The femur is remarkably humanlike," Pickford observes. It has a long femoral neck, which would have placed the shaft at an angle relative to the lower leg (thereby stabilizing the hip), and a groove on the back of that femoral neck, where a muscle known as the obturator externus pressed against the bone during upright walking. In other respects, *Orrorin* was a primitive animal: its canine teeth are large and pointed relative to human canines, and its arm and finger bones retain adaptations for climbing. But the femur characteristics signify to Pickford and Senut that when it was on the ground, *Orrorin* walked like a man.



In fact, they argue, *Orrorin* appears to have had a more humanlike gait than the much younger Lucy did. Breaking with paleoanthropological dogma, the team posits that *Orrorin* gave rise to *Homo* via the proposed genus *Praeanthropus* (which comprises a subset of the fossils currently assigned to *A. afarensis* and *A. anamensis*), leaving Lucy and her kin on an evolutionary sideline. *Ardipithecus*, they believe, was a chimpanzee ancestor.

Not everyone is persuaded by the femur argument. C. Owen Lovejoy of Kent State University counters that published computed tomography scans through *Orrorin*'s femoral neck—which Pickford and Senut say reveal humanlike bone structure—actually show a chimplike distribution of cortical bone, an important indicator of the strain placed on that part of the femur during locomotion. Cross sections of *A. afarensis*'s femoral neck, in contrast, look entirely human, he states. Lovejoy suspects that *Orrorin* was frequently—but not habitually—bipedal and spent a significant amount of time in the trees. That wouldn't exclude it from hominid status, because full-blown bipedalism almost certainly didn't emerge in one fell swoop. Rather *Orrorin* may have simply not yet evolved the full complement of traits required for habitual bipedalism. Viewed that way, *Orrorin* could still be on the ancestral line, albeit further removed from *Homo* than Pickford and Senut would have it.

Better evidence of early routine bipedalism, in Lovejoy's view, surfaced a few months after the *Orrorin* report, when Berkeley graduate student Yohannes Haile-Selassie announced the discovery of slightly younger fossils from Ethiopia's Middle Awash region. Those 5.2-million- to 5.8-million-year-old remains, which have been classified as a subspecies of *Ardipithecus ramidus, A. r. kadabba,* include a complete foot phalanx, or toe bone, bearing a telltale trait. The bone's joint is angled in precisely the way one would expect if *A. r. kadabba* "toed off" as humans do when walking, reports Lovejoy, who has studied the fossil.

Other workers are less impressed by the toe morphology. "To me, it looks for all the world like a chimpanzee foot phalanx," comments David Begun of the University of Toronto, noting from photographs that it is longer, slimmer and more curved than a biped's toe bone should be. Clarification may come when White and his collaborators publish findings on an as yet undescribed partial skeleton of *Ardipithecus,* which White says they hope to do within the next year or two.

Differing anatomical interpretations notwithstanding, if either *Orrorin* or *A. r. kadabba* were a biped, that would not only push the origin of our strange mode of locomotion back by nearly 1.5 million years, it would also lay to rest a popular idea about the conditions under which our striding gait evolved. Received wisdom holds that our ancestors became bipedal on the African savanna, where upright walking may have kept the blistering sun off their backs, given them access to previously out-of-reach foods, or afforded them a better view above the tall

Humanity may have arisen more than a million years earlier than a number of molecular studies had estimated. More important, it may have originated in a different locale.

grass. But paleoecological analyses indicate that *Orrorin* and *Ardipithecus* dwelled in forested habitats, alongside monkeys and other typically woodland creatures. In fact, Giday WoldeGabriel of Los Alamos National Laboratory and his colleagues, who studied the soil chemistry and animal remains at the *A. r. kadabba* site, have noted that early hominids may not have ventured beyond these relatively wet and wooded settings until after 4.4 million years ago.

If so, climate change may not have played as important a role in driving our ancestors from four legs to two as has been thought. For his part, Lovejoy observes that a number of the savanna-based hypotheses focusing on posture were not especially well conceived to begin with. "If your eyes were in your toes, you could stand on your hands all day and look over tall grass, but you'd never evolve into a hand-walker," he jokes. In other words, selection for upright posture alone would not, in his view, have led to bipedal locomotion. The most plausible explanation for the emergence of bipedalism, Lovejoy says, is that it freed the hands and allowed males to collect extra food with which to woo mates. In this model, which he developed in the 1980s, females who chose good providers could devote more energy to child rearing, thereby maximizing their reproductive success.

The Oldest Ancestor?

THE PALEOANTHROPOLOGICAL community was still digesting the implications of the *Orrorin* and *A. r. kadabba* discoveries when Brunet's fossil find from Chad came to light. With *Sahelanthropus* have come new answers—and new questions. Unlike *Orrorin* and *A. r. kadabba,* the *Sahelanthropus* material does not include any postcranial bones, making it impossible at this point to know whether the animal was bipedal, the traditional hallmark of humanness. But Brunet argues that a suite of features in the teeth and skull, which he believes belongs to a male, judging from the massive brow ridge, clearly links this creature to all later hominids. Characteristics of *Sahelanthropus*'s canines are especially important in his assessment. In all modern and fossil apes, and therefore presumably in the last common ancestor of chimps and humans, the large upper canines are honed against the first lower premolars, producing a sharp edge along the back of the canines. This so-called honing canine-premolar complex is pronounced in males, who use their canines to compete with one another for females. Humans lost these fighting teeth, evolving smaller, more incisorlike canines that occlude tip to tip, an arrangement that creates a distinctive wear pattern over time. In their size, shape and wear, the *Sahelanthropus* canines are modified in the human direction, Brunet asserts.

At the same time, *Sahelanthropus* exhibits a number of apelike traits, such as its small braincase and widely spaced eye sockets. This mosaic of primitive and advanced features, Brunet says, suggests a close relationship to the last common ancestor. Thus, he proposes that *Sahelanthropus* is the earliest member of the human lineage and the ancestor of all later hominids, in-

WITNESS/GAMMA

HUNTING FOR HOMINIDS: Michel Brunet (*left*), whose team uncovered *Sahelanthropus,* has combed the sands of the Djurab Desert in Chad for nearly a decade. Martin Pickford and Brigitte Senut (*center*) discovered *Orrorin* in Kenya's Tugen Hills. Tim White (*top right*) and Yohannes Haile-Selassie (*bottom right*) found *Ardipithecus* in the Middle Awash region of Ethiopia.

cluding *Orrorin* and *Ardipithecus*. If Brunet is correct, humanity may have arisen more than a million years earlier than a number of molecular studies had estimated. More important, it may have originated in a different locale than has been posited. According to one model of human origins, put forth in the 1980s by Yves Coppens of the College of France, East Africa was the birthplace of humankind. Coppens, noting that the oldest human fossils came from East Africa, proposed that the continent's Rift Valley—a gash that runs from north to south—split a single ancestral ape species into two populations. The one in the east gave rise to humans; the one in the west spawned today's apes [see "East Side Story: The Origin of Humankind," by Yves Coppens; SCIENTIFIC AMERICAN, May 1994]. Scholars have recognized for some time that the apparent geographic separation might instead be an artifact of the scant fossil record. The discovery of a seven-million-year-old hominid in Chad, some 2,500 kilometers west of the Rift Valley, would deal the theory a fatal blow.

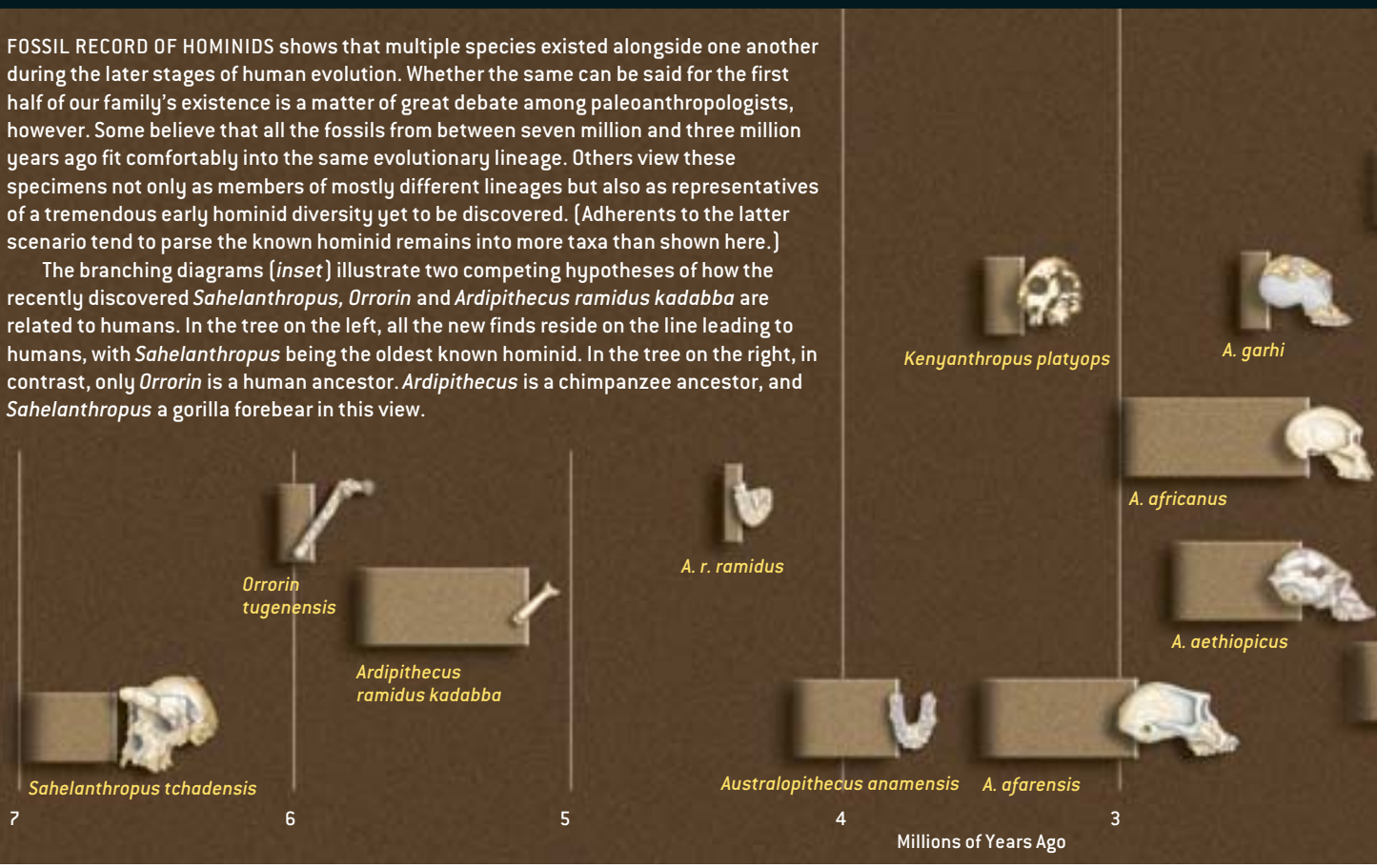
Most surprising of all may be what *Sahelanthropus* reveals about the last common ancestor of humans and chimpanzees. Paleoanthropologists have typically imagined that that creature resembled a chimp in having, among other things, a strongly projecting lower face, thinly enameled molars and large canines. Yet *Sahelanthropus,* for all its generally apelike traits, has only a moderately prognathic face, relatively thick enamel, small canines and a brow ridge larger than that of any living ape. "If *Sahelanthropus* shows us anything, it shows us that the last common ancestor was not a chimpanzee," Berkeley's White remarks. "But why should we have expected otherwise?" Chimpanzees have had just as much time to evolve as humans have had, he points out, and they have become highly specialized, fruit-eating apes.

Brunet's characterization of the Chadian remains as those of a human ancestor has not gone unchallenged, however. "Why *Sahelanthropus* is necessarily a hominid is not particularly clear," comments Carol V. Ward of the University of Missouri. She and others are skeptical that the canines are as humanlike as Brunet claims. Along similar lines, in a letter published last October in the journal *Nature,* in which Brunet's team initially reported its findings, University of Michigan paleoanthropologist Milford H. Wolpoff, along with *Orrorin* discoverers Pickford and Senut, countered that *Sahelanthropus* was an ape rather than a hominid. The massive brow and certain features on the base and rear of *Sahelanthropus*'s skull, they observed, call to mind the anatomy of a quadrupedal ape with a difficult-to-chew diet, whereas the small canine suggests that it was a female of such a species, not a male human ancestor. Lacking proof that *Sahelanthropus* was bipedal, so their reasoning goes, Brunet doesn't have a leg to stand on. (Pickford and Senut further argue that the animal was specifically a go-

Hominids in Time

FOSSIL RECORD OF HOMINIDS shows that multiple species existed alongside one another during the later stages of human evolution. Whether the same can be said for the first half of our family's existence is a matter of great debate among paleoanthropologists, however. Some believe that all the fossils from between seven million and three million years ago fit comfortably into the same evolutionary lineage. Others view these specimens not only as members of mostly different lineages but also as representatives of a tremendous early hominid diversity yet to be discovered. (Adherents to the latter scenario tend to parse the known hominid remains into more taxa than shown here.)

The branching diagrams (*inset*) illustrate two competing hypotheses of how the recently discovered *Sahelanthropus, Orrorin* and *Ardipithecus ramidus kadabba* are related to humans. In the tree on the left, all the new finds reside on the line leading to humans, with *Sahelanthropus* being the oldest known hominid. In the tree on the right, in contrast, only *Orrorin* is a human ancestor. *Ardipithecus* is a chimpanzee ancestor, and *Sahelanthropus* a gorilla forebear in this view.



rilla ancestor.) In a barbed response, Brunet likened his detractors to those Dart encountered in 1925, retorting that *Sahelanthropus*'s apelike traits are simply primitive holdovers from its own ape predecessor and therefore uninformative with regard to its relationship to humans.

The conflicting views partly reflect the fact that researchers disagree over what makes the human lineage unique. "We have trouble defining hominids," acknowledges Roberto Macchiarelli, also at the University of Poitiers. Traditionally paleoanthropologists have regarded bipedalism as the characteristic that first set human ancestors apart from other apes. But subtler changes—the metamorphosis of the canine, for instance—may have preceded that shift.

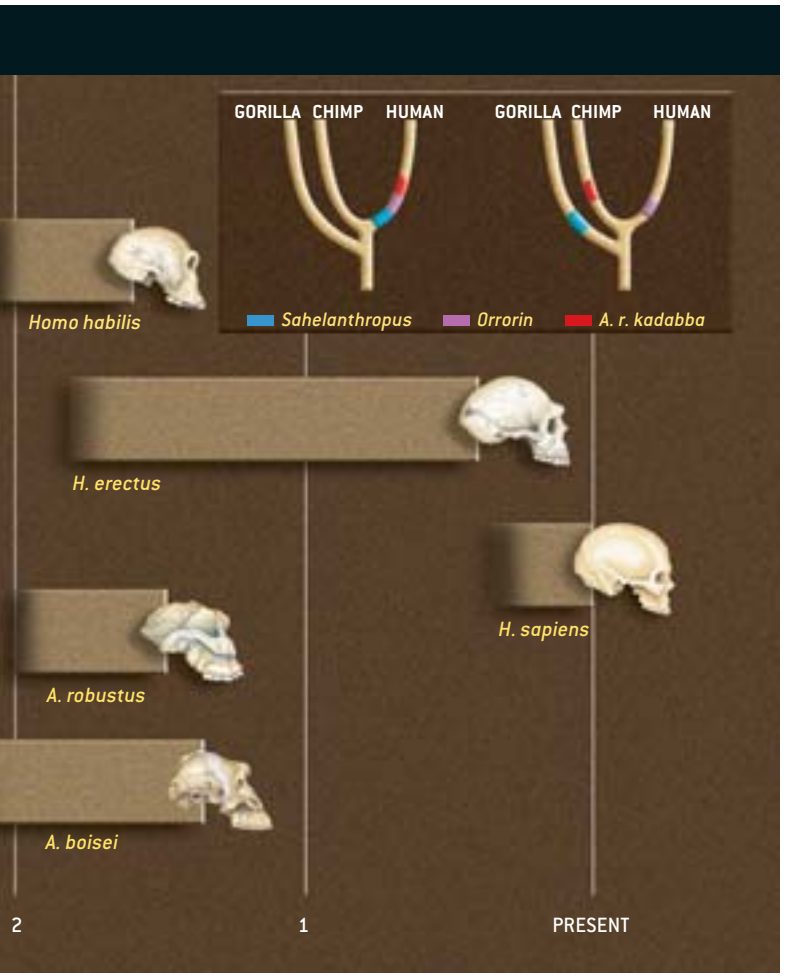
To understand how animals are related to one another, evolutionary biologists employ a method called cladistics, in which organisms are grouped according to shared, newly evolved traits. In short, creatures that have these derived characteristics in common are deemed more closely related to one another than they are to those that exhibit only primitive traits inherited from a more distant common ancestor. The first occurrence in the fossil record of a shared, newly acquired trait serves as a baseline indicator of the biological division of an ancestral species into two daughter species—in this case, the point at which chimps and humans diverged from their common ancestor—and that trait is considered the defining characteristic of the group.

Thus, cladistically "what a hominid is from the point of view of skeletal morphology is summarized by those characters preserved in the skeleton that are present in populations that directly succeeded the genetic splitting event between chimps and humans," explains William H. Kimbel of Arizona State University. With only an impoverished fossil record to work from, paleontologists can't know for certain what those traits were. But the two leading candidates for the title of seminal hominid characteristic, Kimbel says, are bipedalism and the transformation of the canine. The problem researchers now face in trying to suss out what the initial changes were and which, if any, of the new putative hominids sits at the base of the human clade is that so far *Orrorin, A. r. kadabba* and *Sahelanthropus* are represented by mostly different bony elements, making comparisons among them difficult.

How Many Hominids?

MEANWHILE THE ARRIVAL of three new taxa to the table has intensified debate over just how diverse early hominids were. Experts concur that between three million and 1.5 million years ago, multiple hominid species existed alongside one

ILLUSTRATIONS BY PATRICIA J. WYNNE AND CORNELIA BLIK



another at least occasionally. Now some scholars argue that this rash of discoveries demonstrates that human evolution was a complex affair from the outset. Toronto's Begun—who believes that the Miocene ape ancestors of modern African apes and humans spent their evolutionarily formative years in Europe and western Asia before reentering Africa—observes that *Sahelanthropus* bears exactly the kind of motley features that one would expect to see in an animal that was part of an adaptive radiation of apes moving into a new milieu. "It would not surprise me if there were 10 or 15 genera of things that are more closely related to *Homo* than to chimps," he says. Likewise, in a commentary that accompanied the report by Brunet and his team in *Nature*, Bernard Wood of George Washington University wondered whether *Sahelanthropus* might hail from the African ape equivalent of Canada's famed Burgess Shale, which has yielded myriad invertebrate fossils from the Cambrian period, when the major modern animal groups exploded into existence. Viewed that way, the human evolutionary tree would look more like an unkempt bush, with some, if not all, of the new discoveries occupying terminal twigs instead of coveted spots on the meandering line that led to humans.

Other workers caution against inferring the existence of multiple, coeval hominids on the basis of what has yet been found. "That's *X-Files* paleontology," White quips. He and Brunet both note that between seven million and four million years ago, only one hominid species is known to have existed at any given time. "Where's the bush?" Brunet demands. Even at humanity's peak diversity, two million years ago, White says, there were only three taxa sharing the landscape. "That ain't the Cambrian explosion," he remarks dryly. Rather, White suggests, there is no evidence that the base of the family tree is anything other than a trunk. He thinks that the new finds might all represent snapshots of the *Ardipithecus* lineage through time, with *Sahelanthropus* being the earliest hominid and with *Orrorin* and *A. r. kadabba* representing its lineal descendants. (In this configuration, *Sahelanthropus* and *Orrorin* would become species of *Ardipithecus*.)

Investigators agree that more fossils are needed to elucidate how *Orrorin*, *A. r. kadabba* and *Sahelanthropus* are related to one another and to ourselves, but obtaining a higher-resolution picture of the roots of humankind won't be easy. "We're going to have a lot of trouble diagnosing the very earliest members of our clade the closer we get to that last common ancestor," Missouri's Ward predicts. Nevertheless, "it's really important to sort out what the starting point was," she observes. "Why the human lineage began is the question we're trying to answer, and these new finds in some ways may hold the key to answering that question—or getting closer than we've ever gotten before."

It may be that future paleoanthropologists will reach a point at which identifying an even earlier hominid will be well nigh impossible. But it's unlikely that this will keep them from trying. Indeed, it would seem that the search for the first hominids is just heating up. "The *Sahelanthropus* cranium is a messenger [indicating] that in central Africa there is a desert full of fossils of the right age to answer key questions about the genesis of our clade," White reflects. For his part, Brunet, who for more than a quarter of a century has doggedly pursued his vision through political unrest, sweltering heat and the blinding sting of an unrelenting desert wind, says that ongoing work in Chad will keep his team busy for years to come. "This is the beginning of the story," he promises, "just the beginning." As I sit in Brunet's office contemplating the seven-million-year-old skull of *Sahelanthropus*, the fossil hunter's quest doesn't seem quite so unimaginable. Many of us spend the better part of a lifetime searching for ourselves. SA

Kate Wong is a writer and editor for ScientificAmerican.com

MORE TO EXPLORE

Late Miocene Hominids from the Middle Awash, Ethiopia. Yohannes Haile-Selassie in *Nature,* Vol. 412, pages 178–181; July 12, 2001.

Extinct Humans. Ian Tattersall and Jeffrey H. Schwartz. Westview Press, 2001.

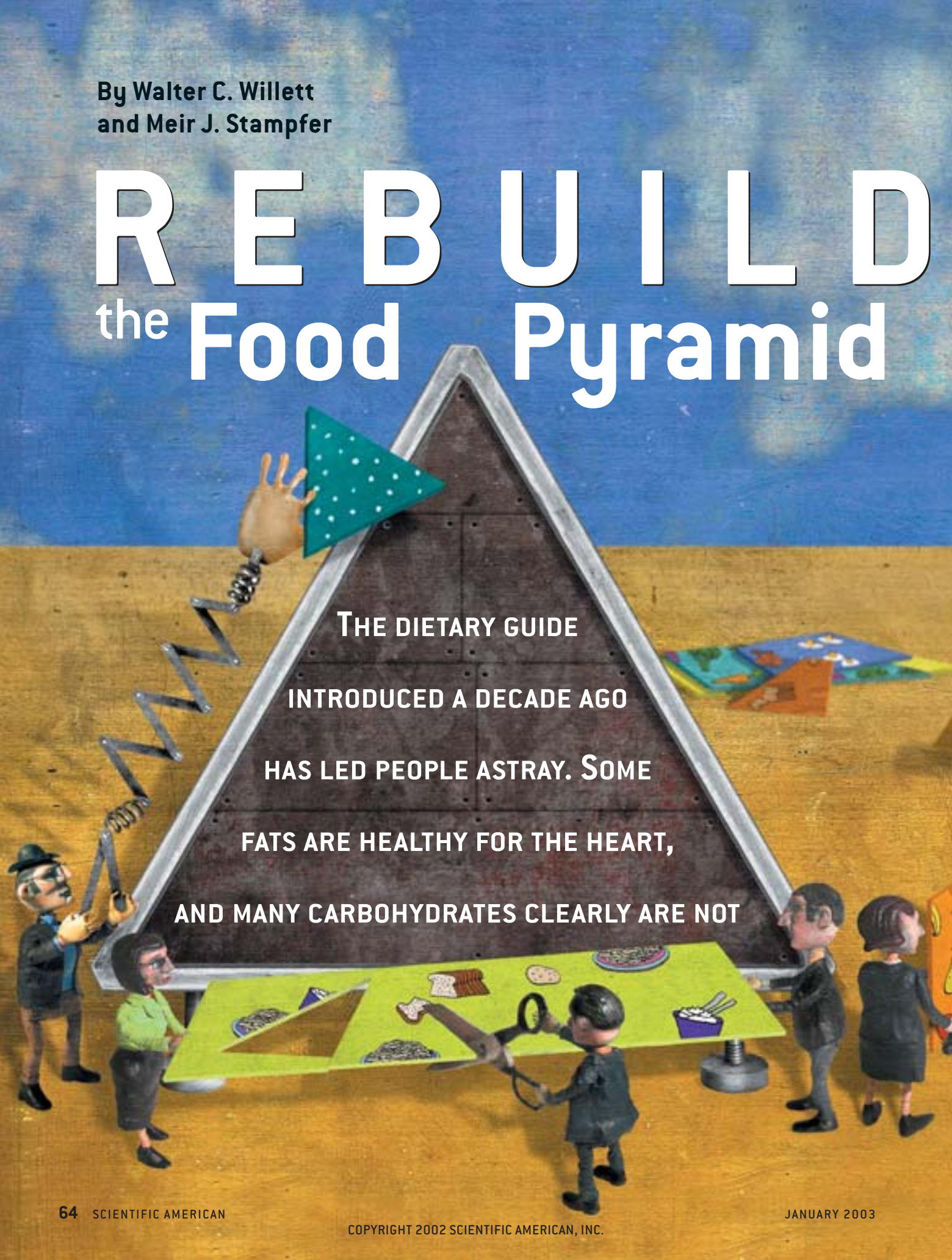
Bipedalism in *Orrorin tugenensis* Revealed by Its Femora. Martin Pickford, Brigitte Senut, Dominique Gommercy and Jacques Treil in *Comptes Rendus: Palevol,* Vol. 1, No. 1, pages 1–13; 2002.

A New Hominid from the Upper Miocene of Chad, Central Africa. Michel Brunet, Franck Guy, David Pilbeam, Hassane Taisso Mackaye et al. in *Nature,* Vol. 418, pages 145–151; July 11, 2002.

The Primate Fossil Record. Edited by Walter C. Hartwig. Cambridge University Press, 2002.

By Walter C. Willett and Meir J. Stampfer

REBUILD the Food Pyramid



THE DIETARY GUIDE INTRODUCED A DECADE AGO HAS LED PEOPLE ASTRAY. SOME FATS ARE HEALTHY FOR THE HEART, AND MANY CARBOHYDRATES CLEARLY ARE NOT

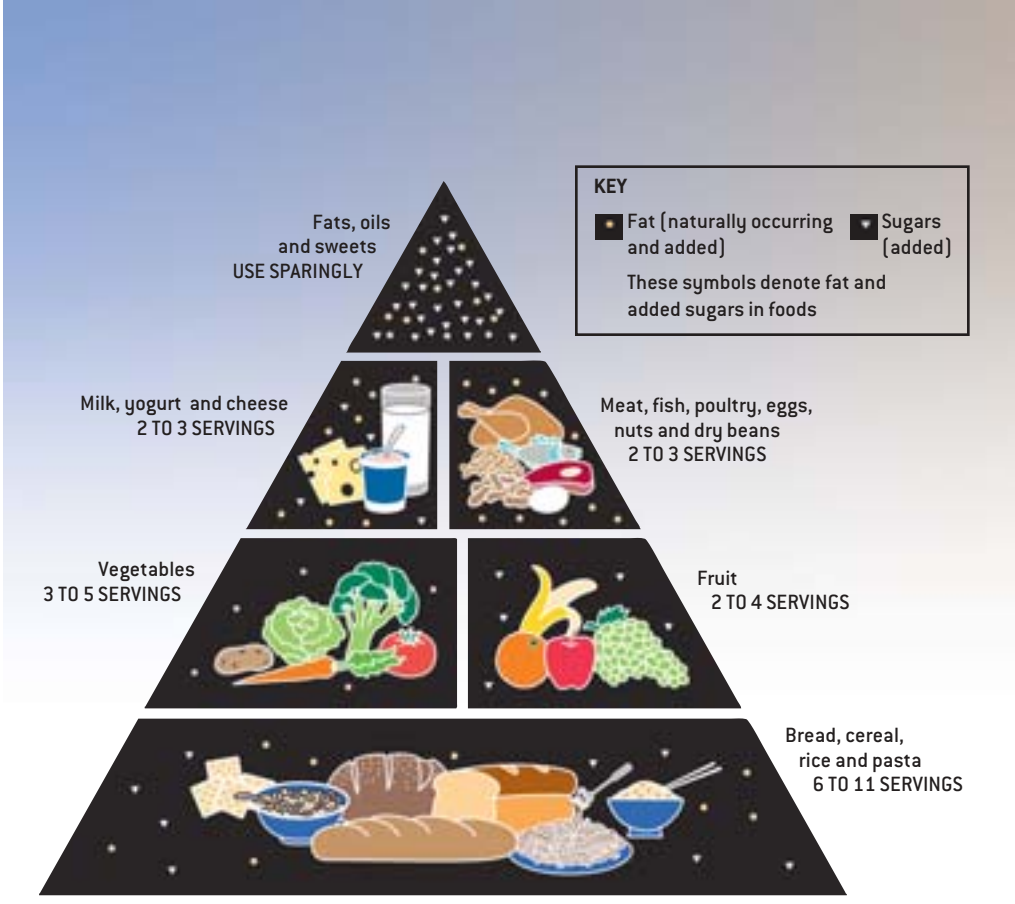
ING

In 1992

the U.S. Department of Agriculture officially released the Food Guide Pyramid, which was intended to help the American public make dietary choices that would maintain good health and reduce the risk of chronic disease. The recommendations embodied in the pyramid soon became well known: people should minimize their consumption of fats and oils but should eat six to 11 servings a day of foods rich in complex carbohydrates—bread, cereal, rice, pasta and so on. The food pyramid also recommended generous amounts of vegetables (including potatoes, another plentiful source of complex carbohydrates), fruit and dairy products, and at least two servings a day from the meat and beans group, which lumped together red meat with poultry, fish, nuts, legumes and eggs.

Even when the pyramid was being developed, though, nutritionists had long known that some types of fat are essential to health and can reduce the risk of cardiovascular disease. Furthermore, scientists had found little evidence that a high intake of carbohydrates is beneficial. Since 1992 more and more research has shown that the USDA pyramid is grossly flawed. By promoting the consumption of all complex carbohydrates and eschewing all fats and oils, the pyramid provides misleading guidance. In short, not all fats are bad for you, and by no means are all complex carbohydrates good for you. The USDA's Center for Nutrition Policy and Promotion is now reassessing the pyramid, but this effort is not expected to be completed until 2004. In the meantime, we have drawn up a new pyramid that better reflects the current understanding of the relation between diet and health. Studies indicate that adherence to the recommendations in the revised pyramid can significantly reduce the risk of cardiovascular disease for both men and women.

How did the original USDA pyramid go so wrong? In part, nutritionists fell victim to a desire to simplify their dietary recommendations. Researchers had known for decades that saturated fat—found in abundance in red meat and dairy products—raises cholesterol levels in the blood. High cholesterol levels, in turn, are associated with a high risk of coronary heart disease (heart attacks and other ailments caused by the blockage of the arteries to the heart). In the 1960s controlled feeding studies, in which the participants eat carefully prescribed diets for several weeks, substantiated that saturated fat increases cholesterol levels. But the studies also showed that polyunsaturated fat—found in vegetable oils and fish—reduces cholesterol. Thus, dietary advice during the 1960s and 1970s emphasized the replacement of saturated fat with polyunsaturated fat, not total fat reduction. (The subsequent doubling of polyunsaturated fat consumption among Americans probably contributed greatly to the halving of coronary heart disease rates in the U.S. during the 1970s and 1980s.)



OLD FOOD PYRAMID

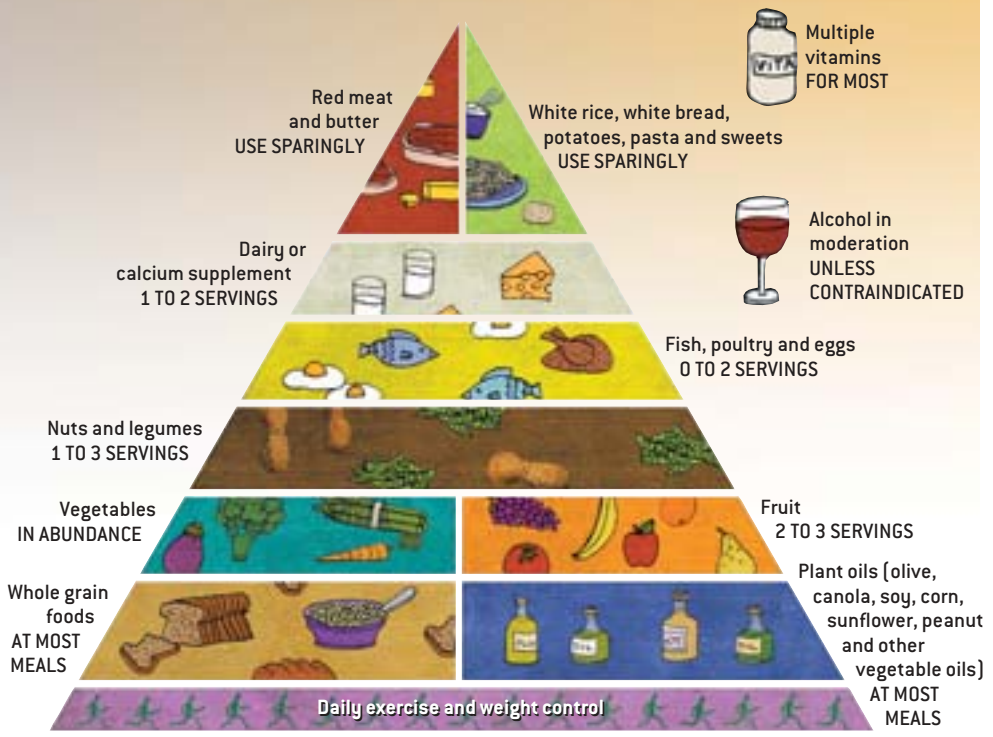
conceived by the U.S. Department of Agriculture was intended to convey the message "Fat is bad" and its corollary "Carbs are good." These sweeping statements are now being questioned.

For information on the amount of food that counts as one serving, visit www.nal.usda.gov:8001/py/pmap.htm

Overview/*The Food Guide Pyramid*

- The U.S. Department of Agriculture's Food Guide Pyramid, introduced in 1992, recommended that people avoid fats but eat plenty of carbohydrate-rich foods such as bread, cereal, rice and pasta. The goal was to reduce the consumption of saturated fat, which raises cholesterol levels.
- Researchers have found that a high intake of refined carbohydrates such as white bread and white rice can wreak havoc on the body's glucose and insulin levels. Replacing these carbohydrates with healthy fats—monounsaturated or polyunsaturated—actually lowers one's risk of heart disease.
- Nutritionists are now proposing a new food pyramid that encourages the consumption of healthy fats and whole grain foods but recommends avoiding refined carbohydrates, butter and red meat.

RICHARD BORGE (*preceding pages*); USDA/DHHS (*old pyramid*)



NEW FOOD PYRAMID

outlined by the authors distinguishes between healthy and unhealthy types of fat and carbohydrates. Fruits and vegetables are still recommended, but the consumption of dairy products should be limited.

The notion that fat in general is to be avoided stems mainly from observations that affluent Western countries have both high intakes of fat and high rates of coronary heart disease. This correlation, however, is limited to saturated fat. Societies in which people eat relatively large portions of monounsaturated and polyunsaturated fat tend to have lower rates of heart disease [*see illustration on next page*]. On the Greek island of Crete, for example, the traditional diet contained much olive oil (a rich source of monounsaturated fat) and fish (a source of polyunsaturated fat). Although fat constituted 40 percent of the calories in this diet, the rate of heart disease for those who followed it was lower than the rate for those who followed the traditional diets of Japan, in which fat made up only 8 to 10 percent of the calories. Furthermore, international comparisons can be misleading: many negative influences on health, such as smoking, physical inactivity and high amounts of body fat, are also correlated with Western affluence.

Unfortunately, many nutritionists decided it would be too difficult to educate the public about these subtleties. Instead they put out a clear, simple message: "Fat is bad." Because saturated fat represents about 40 percent of all fat consumed in the U.S., the rationale of the USDA was that advocating a low-fat diet would naturally reduce the intake of saturated fat. This recommendation was soon reinforced by the food industry, which began selling cookies, chips and other products that were low in fat but often high in sweeteners such as high-fructose corn syrup.

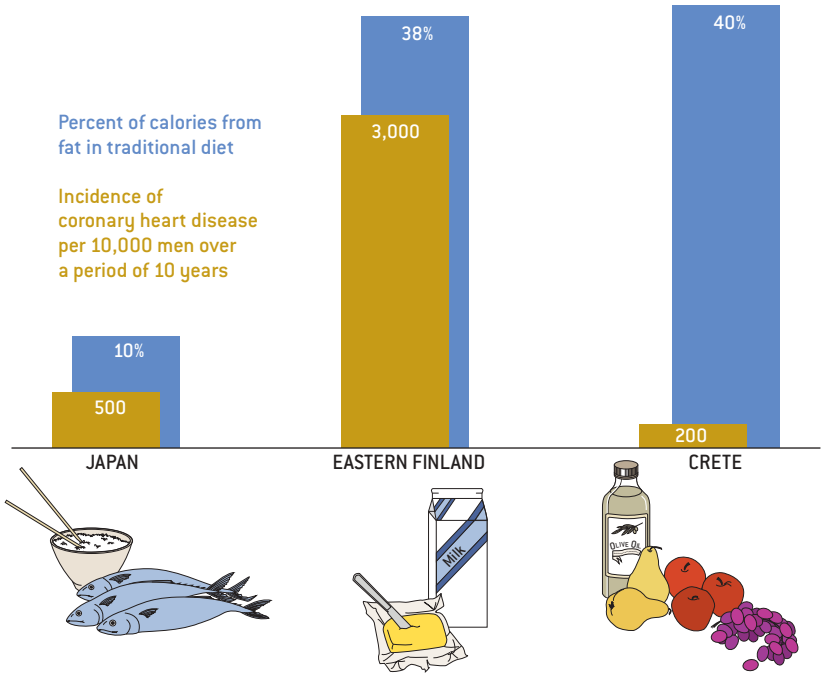
When the food pyramid was being developed, the typical American got about 40 percent of his or her calories from fat, about 15 percent from protein and about 45 percent from carbohydrates. Nutritionists did not want to suggest eating more protein, because many sources of protein (red meat, for example) are also heavy in saturated fat. So the "Fat is bad" mantra led to the corollary "Carbs are good." Dietary guidelines from the American Heart Association and other groups recommended that people get at least half their calories from carbohydrates and no more than 30 percent from fat. This 30 percent limit has become so entrenched among nutritionists that even the sophisticated observer could be forgiven for thinking that many studies must show that individuals with that level of fat intake enjoyed better health than those with higher levels. But no study has demonstrated long-term health benefits that can be directly attributed to a low-fat diet. The 30 percent limit on fat was essentially drawn from thin air.

The wisdom of this direction became even more questionable after researchers found that the two main cholesterol-carrying chemicals—low-density lipoprotein (LDL), popularly known as "bad cholesterol," and high-density lipoprotein (HDL), known as "good cholesterol"—have very different effects on the risk of coronary heart disease. Increasing the ratio of LDL to HDL in the blood raises the risk, whereas decreasing the ratio lowers it. By the early 1990s controlled feeding studies had shown that when a person replaces calories from saturated fat with an equal amount of calories from carbohydrates the levels of LDL and total cholesterol fall, but the level of HDL also falls. Because the ratio of LDL to HDL does not change, there is only a small reduction in the person's risk of heart disease. Moreover, the switch to carbohydrates boosts the blood levels of triglycerides, the component molecules of fat, probably because of effects on the body's endocrine system. High triglyceride levels are also associated with a high risk of heart disease.

The effects are more grievous when a person switches from either monounsaturated or polyunsaturated fat to carbohydrates. LDL levels rise and HDL levels drop, making the cholesterol ratio worse. In contrast, replacing saturated fat with either monounsaturated or polyunsaturated fat improves this ratio and would be expected to reduce heart disease. The only fats that are significantly more deleterious than carbohydrates are the trans-unsaturated fatty acids; these are produced by the partial hydrogenation of liquid vegetable oil, which causes it to solidify.

RICHARD BORGE

Fat and Heart Disease



INTERNATIONAL COMPARISONS reveal that total fat intake is a poor indicator of heart disease risk. What is important is the type of fat consumed. In regions where saturated fats traditionally made up much of the diet (for example, eastern Finland), rates of heart disease were much higher than in areas where monounsaturated fats were prevalent (such as the Greek island of Crete). Crete's Mediterranean diet, based on olive oil, was even better for the heart than the low-fat traditional diet of Japan.

Found in many margarines, baked goods and fried foods, trans fats are uniquely bad for you because they raise LDL and triglycerides while reducing HDL.

The Big Picture

TO EVALUATE FULLY the health effects of diet, though, one must look beyond cholesterol ratios and triglyceride levels. The foods we eat can cause heart disease through many other pathways, including raising blood pressure or boosting the tendency of blood to clot. And other foods can prevent heart disease in surprising ways; for instance, omega-3 fatty acids (found in fish and some plant oils) can reduce the likelihood of ventricular fibrillation, a heart rhythm disturbance that causes sudden death.

The ideal method for assessing all these adverse and beneficial effects would be to conduct large-scale trials in which individuals are randomly assigned to one diet or another and followed for many years. Because of practical constraints and cost, few such studies have been conducted, and most of these have focused on patients who already suffer from heart disease. Though limited, these studies have supported the benefits of replacing saturated fat with polyunsaturated fat, but not with carbohydrates.

The best alternative is to conduct large epidemiological studies in which the diets of many people are periodically assessed and the participants are monitored for the development of heart disease and other conditions. One of the best-known examples of this research is the Nurses' Health Study, which was begun in 1976 to evaluate the effects of oral contraceptives but was soon extended to nutrition as well. Our group at Harvard University has followed nearly 90,000 women in this study who first completed detailed questionnaires on diet in 1980, as well as more than 50,000 men who were enrolled in the Health Professionals Follow-Up Study in 1986.

After adjusting the analysis to account for smoking, physical activity and other recognized risk factors, we found that a participant's risk of heart disease was strongly influenced by the type of dietary fat consumed. Eating trans fat increased the risk substantially, and eating saturated fat increased it slightly. In contrast, eating monounsaturated and polyunsaturated fats decreased the risk—just as the controlled feeding studies predicted. Because these two effects counterbalanced each other, higher overall consumption of fat did not lead to higher rates of coronary heart disease. This finding reinforced a 1989 report by the National Academy of Sciences that concluded that total fat intake alone was not associated with heart disease risk.

But what about illnesses besides coronary heart disease? High rates of breast, colon and prostate cancers in affluent Western countries have led to the belief that the consumption of fat, particularly animal fat, may be a risk factor. But large epidemiological studies have shown little evidence that total fat consumption or intakes of specific types of fat during midlife affect the risks of breast or colon cancer. Some studies have indicated that prostate cancer and the consumption of animal fat may be associated, but reassuringly there is no suggestion that vegetable oils increase any cancer risk. Indeed, some studies have suggested that vegetable oils may slightly reduce such risks. Thus, it is reasonable to make decisions about dietary fat on the basis of its effects on cardiovascular disease, not cancer.

Finally, one must consider the impact of fat consumption on obesity, the most serious nutritional problem in the U.S. Obesity is a major risk factor for several diseases, including type 2 diabetes (also called adult-onset diabetes), coronary heart disease, and cancers of the breast, colon, kidney and esophagus. Many nutritionists believe that eating fat can contribute to weight gain because fat contains more calories per gram than protein or carbohydrates. Also, the process of storing dietary fat in the body may be more efficient than the conversion of carbohydrates to body fat. But recent controlled feeding studies have shown that these considerations are not practically important. The best way to avoid obesity is to limit your total calories, not just the fat calories. So the critical issue is whether the fat composition of a diet can influence one's ability to control caloric intake. In other words, does eating fat leave you

CORNELIA BLIK

more or less hungry than eating protein or carbohydrates? There are various theories about why one diet should be better than another, but few long-term studies have been done. In randomized trials, individuals assigned to low-fat diets tend to lose a few pounds during the first months but then regain the weight. In studies lasting a year or longer, low-fat diets have consistently not led to greater weight loss.

Carbo-Loading

NOW LET'S LOOK at the health effects of carbohydrates. Complex carbohydrates consist of long chains of sugar units such as glucose and fructose; sugars contain only one or two units. Because of concerns that sugars offer nothing but "empty calories"—that is, no vitamins, minerals or other nutrients—complex carbohydrates form the base of the USDA food pyramid. But refined carbohydrates, such as white bread and white rice, can be very quickly broken down to glucose, the primary fuel for the body. The refining process produces an easily absorbed form of starch—which is defined as glucose molecules bound together—and also removes many vitamins and minerals and fiber. Thus, these carbohydrates increase glucose levels in the blood more than whole grains do. (Whole grains have not been milled into fine flour.)

Or consider potatoes. Eating a boiled potato raises blood sugar levels higher than eating the same amount of calories from table sugar. Because potatoes are mostly starch, they can be rapidly metabolized to glucose. In contrast, table sugar (sucrose) is a disaccharide consisting of one molecule of glucose and one molecule of fructose. Fructose takes longer to convert to glucose, hence the slower rise in blood glucose levels.

A rapid increase in blood sugar stimulates a large release of insulin, the hormone that directs glucose to the muscles and liver. As a result, blood sugar plummets, sometimes even going below the baseline. High levels of glucose and insulin can have negative effects on cardiovascular health, raising triglycerides and lowering HDL (the good cholesterol). The precipitous decline in glucose can also lead to more hunger after a carbohydrate-rich meal and thus contribute to overeating and obesity.

In our epidemiological studies, we have found that a high intake of starch from refined grains and potatoes is associated with a high risk of type 2 diabetes and coronary heart disease. Conversely, a greater intake of fiber is related to a lower risk of these illnesses. Interestingly, though, the consumption of fiber did not lower the risk of colon cancer, as had been hypothesized earlier.

Overweight, inactive people can become resistant to insulin's effects and therefore require more of the hormone to regulate their blood sugar. Recent evidence indicates that the adverse metabolic response to carbohydrates is substantially worse among people who already have insulin resistance. This finding may account for the ability of peasant farmers in Asia and elsewhere, who are extremely lean and active, to consume large amounts of refined carbohydrates without experiencing diabetes or heart disease, whereas the same diet in a more sedentary population can have devastating effects.

The best way to avoid obesity is to LIMIT YOUR TOTAL CALORIES, not just the fat calories.

Eat Your Veggies

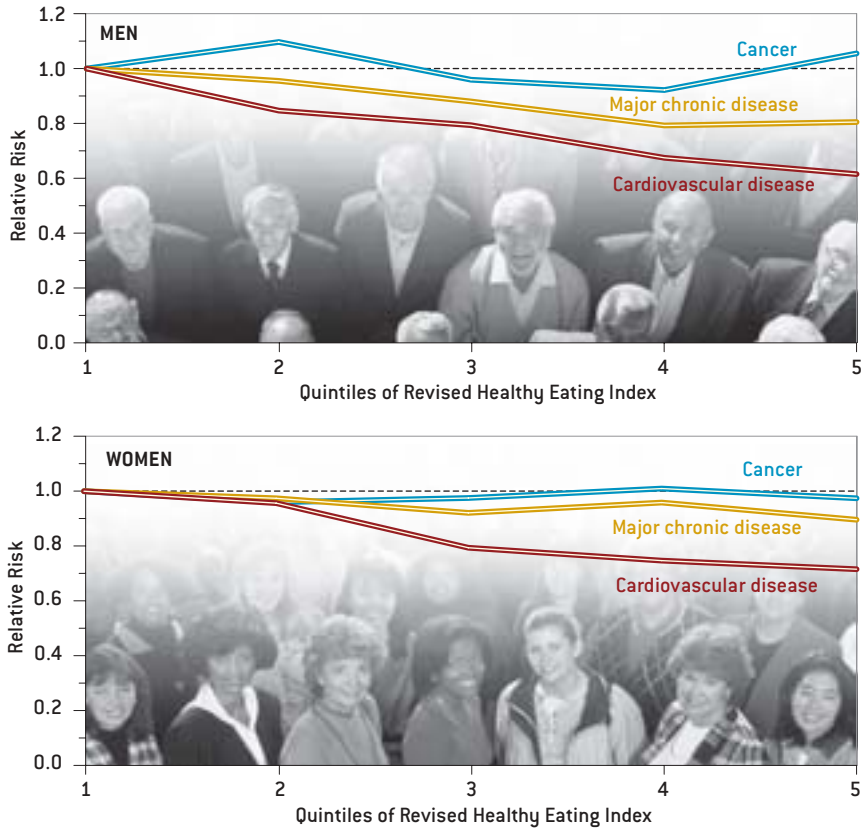
HIGH INTAKE OF FRUITS and vegetables is perhaps the least controversial aspect of the food pyramid. A reduction in cancer risk has been a widely promoted benefit. But most of the evidence for this benefit has come from case-control studies, in which patients with cancer and selected control subjects are asked about their earlier diets. These retrospective studies are susceptible to numerous biases, and recent findings from large prospective studies (including our own) have tended to show little relation between overall fruit and vegetable consumption and cancer incidence. (Specific nutrients in fruits and vegetables may offer benefits, though; for instance, the folic acid in green leafy vegetables may reduce the risk of colon cancer, and the lycopene found in tomatoes may lower the risk of prostate cancer.)

The real value of eating fruits and vegetable may be in reducing the risk of cardiovascular disease. Folic acid and potassium appear to contribute to this effect, which has been seen in several epidemiological studies. Inadequate consumption of folic acid is responsible for higher risks of serious birth defects as well, and low intake of lutein, a pigment in green leafy vegetables, has been associated with greater risks of cataracts and degeneration of the retina. Fruits and vegetables are also the primary source of many vitamins needed for good health. Thus, there are good reasons to consume the recommended five servings a day, even if doing so has little impact on cancer risk. The inclusion of potatoes as a vegetable in the USDA pyra-

THE AUTHORS

WALTER C. WILLETT and *MEIR J. STAMPFER* are professors of epidemiology and nutrition at the Harvard School of Public Health. Willett chairs the school's department of nutrition, and Stampfer heads the department of epidemiology. Willett and Stampfer are also professors of medicine at Harvard Medical School. Both of them practice what they preach by eating well and exercising regularly.

Benefits of the New Pyramid



HEALTH EFFECTS OF THE RECOMMENDATIONS in the revised food pyramid were gauged by studying disease rates among 67,271 women in the Nurses' Health Study and 38,615 men in the Health Professionals Follow-Up Study. Women and men in the fifth quintile (the 20 percent whose diets were closest to the pyramid's recommendations) had significantly lower rates of cardiovascular disease than those in the first quintile (the 20 percent who strayed the most from the pyramid). The dietary recommendations had no significant effect on cancer risk, however.

mid has little justification, however; being mainly starch, potatoes do not confer the benefits seen for other vegetables.

Another flaw in the USDA pyramid is its failure to recognize the important health differences between red meat (beef, pork and lamb) and the other foods in the meat and beans group (poultry, fish, legumes, nuts and eggs). High consumption of red meat has been associated with an increased risk of coronary heart disease, probably because of its high content of saturated fat and cholesterol. Red meat also raises the risk of type 2 diabetes and colon cancer. The elevated risk of colon cancer may be related in part to the carcinogens produced during cooking and the chemicals found in processed meats such as salami and bologna.

Poultry and fish, in contrast, contain less saturated fat and more unsaturated fat than red meat does. Fish is a rich source of the essential omega-3 fatty acids as well. Not surprisingly, studies have shown that people who replace red meat with chicken and fish have a lower risk of coronary heart disease and colon cancer. Eggs are high in cholesterol, but consumption of up to one a day does not appear to have adverse effects on heart disease risk (except among diabetics), probably because the effects of a slightly higher cholesterol level are counterbalanced by other nutritional benefits. Many people have avoided nuts because of their high fat content, but the fat in nuts, including peanuts, is mainly unsaturated, and walnuts in particular are a good source of omega-3 fatty acids. Controlled feeding studies show that nuts improve blood cholesterol ratios, and epidemiological studies indicate that they lower the risk of heart disease and diabetes. Also, people who eat nuts are actually less likely to be obese; perhaps because nuts are more satisfying to the appetite, eating them seems to have the effect of significantly reducing the intake of other foods.

Yet another concern regarding the USDA pyramid is that it promotes overconsumption of dairy products, recommending the equivalent of two or three glasses of milk a day. This advice is usually justified by dairy's calcium content, which is believed to prevent osteoporosis and bone fractures. But the highest rates of fractures are found in countries with high dairy consumption, and large prospective studies have not shown a lower risk of fractures among those who eat plenty of dairy products. Calcium is an essential nutrient, but the requirements for bone health have probably been overstated. What is more, we cannot assume that high dairy consumption is safe: in several studies, men who consumed large amounts of dairy products experienced an increased risk of prostate cancer, and in some studies, women with high intakes had elevated rates of ovarian cancer. Although fat was initially assumed to be the responsible factor, this has not been supported in more detailed analyses. High calcium intake itself seemed most clearly related to the risk of prostate cancer.

More research is needed to determine the health effects of dairy products, but at the moment it seems imprudent to recommend high consumption. Most adults who are following a good overall diet can get the necessary amount of calcium by consuming the equivalent of one glass of milk a day. Under certain circumstances, such as after menopause, people may need more calcium than usual, but it can be obtained at lower cost and without saturated fat or calories by taking a supplement.

A Healthier Pyramid

ALTHOUGH THE USDA'S food pyramid has become an icon of nutrition over the past decade, until recently no studies had evaluated the health of individuals who followed its guidelines. It very likely has some benefits, especially from a high intake of fruits and vegetables. And a de-

KEN FISHER (*men*) AND ED HONOWITZ (*women*) Getty Images; CORNELIA BLIK (*graph*)

crease in total fat intake would tend to reduce the consumption of harmful saturated and trans fats. But the pyramid could also lead people to eat fewer of the healthy unsaturated fats and more refined starches, so the benefits might be negated by the harm.

To evaluate the overall impact, we used the Healthy Eating Index (HEI), a score developed by the USDA to measure adherence to the pyramid and its accompanying dietary guidelines in federal nutrition programs. From the data collected in our large epidemiological studies, we calculated each participant's HEI score and then examined the relation of these scores to subsequent risk of major chronic disease (defined as heart attack, stroke, cancer or nontraumatic death from any cause). When we compared people in the same age groups, women and men with the highest HEI scores did have a lower risk of major chronic disease. But these individuals also smoked less, exercised more and had generally healthier lifestyles than the other participants. After adjusting for these variables, we found that participants with the highest HEI scores did not experience significantly better overall health outcomes. As predicted, the pyramid's harms counterbalanced its benefits.

Because the goal of the pyramid was a worthy one—to encourage healthy dietary choices—we have tried to develop an alternative derived from the best available knowledge. Our revised pyramid [*see illustration on page 67*] emphasizes weight control through exercising daily and avoiding an excessive total intake of calories. This pyramid recommends that the bulk of one's diet should consist of healthy fats (liquid vegetable oils such as olive, canola, soy, corn, sunflower and peanut) and healthy carbohydrates (whole grain foods such as whole wheat bread, oatmeal and brown rice). If both the fats and carbohydrates in your diet are healthy, you probably do not have to worry too much about the percentages of total calories coming from each. Vegetables and fruits should also be eaten in abundance. Moderate amounts of healthy sources of protein (nuts, legumes, fish, poultry and eggs) are encouraged, but dairy consumption should be limited to one to two servings a day. The revised pyramid recommends minimizing the consumption of red meat, butter, refined grains (including white bread, white rice and white pasta), potatoes and sugar.

Trans fat does not appear at all in the pyramid, because it has no place in a healthy diet. A multiple vitamin is suggested for most people, and moderate alcohol consumption can be a worthwhile option (if not contraindicated by specific health conditions or medications). This last recommendation comes with a caveat: drinking no alcohol is clearly better than drinking too much. But more and more studies are showing the benefits of moderate alcohol consumption (in any form: wine, beer or spirits) to the cardiovascular system.

Men and women eating in accordance with THE NEW PYRAMID had a lower risk of major chronic disease.

Can we show that our pyramid is healthier than the USDA's? We created a new Healthy Eating Index that measured how closely a person's diet followed our recommendations. Applying this revised index to our epidemiological studies, we found that men and women who were eating in accordance with the new pyramid had a lower risk of major chronic disease [*see illustration on opposite page*]. This benefit resulted almost entirely from significant reductions in the risk of cardiovascular disease—up to 30 percent for women and 40 percent for men. Following the new pyramid's guidelines did not, however, lower the risk of cancer. Weight control and physical activity, rather than specific food choices, are associated with a reduced risk of many cancers.

Of course, uncertainties still cloud our understanding of the relation between diet and health. More research is needed to examine the role of dairy products, the health effects of specific fruits and vegetables, the risks and benefits of vitamin supplements, and the long-term effects of diet during childhood and early adult life. The interaction of dietary factors with genetic predisposition should also be investigated, although its importance remains to be determined.

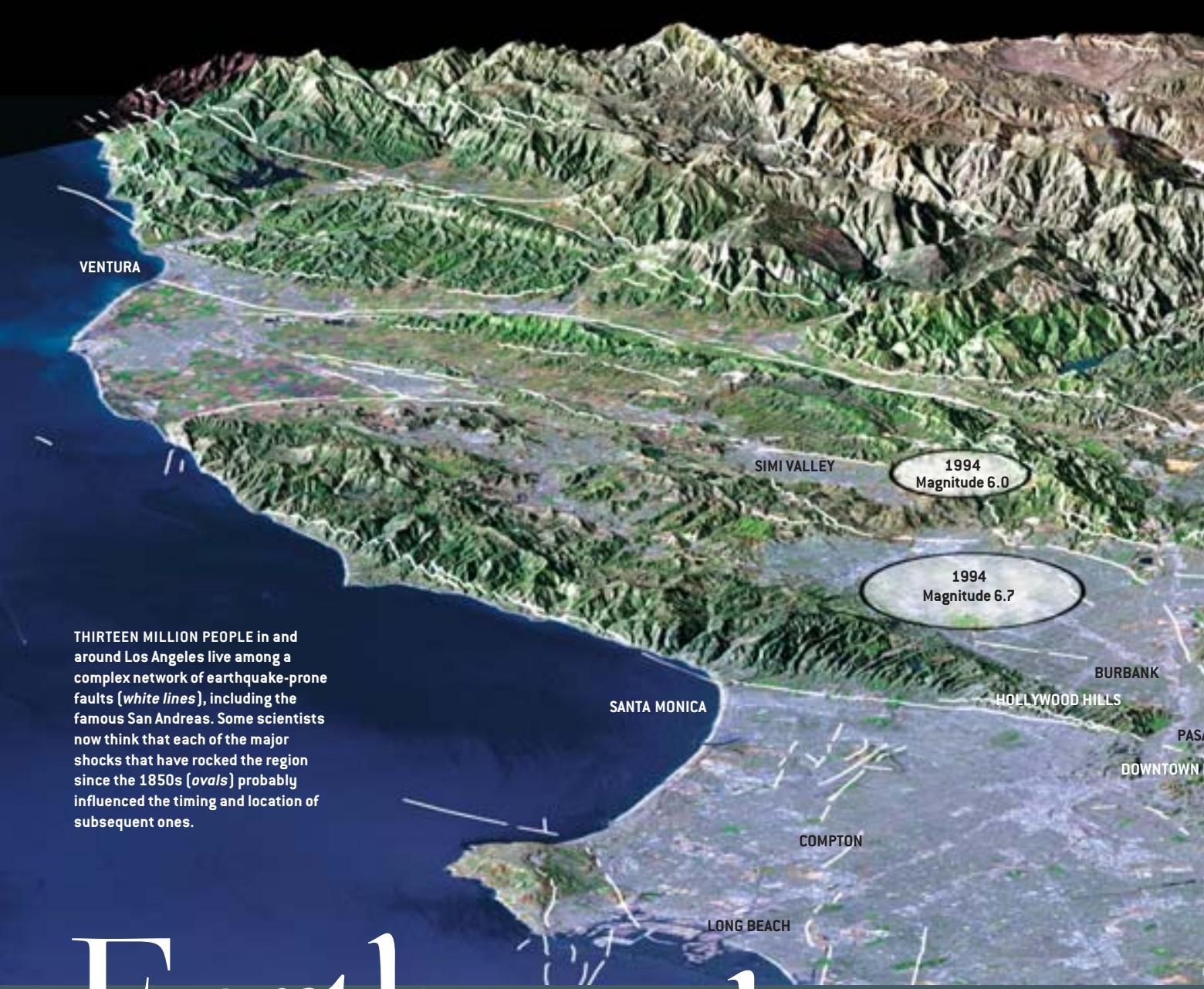
Another challenge will be to ensure that the information about nutrition given to the public is based strictly on scientific evidence. The USDA may not be the best government agency to develop objective nutritional guidelines, because it may be too closely linked to the agricultural industry. The food pyramid should be rebuilt in a setting that is well insulated from political and economic interests. SA

MORE TO EXPLORE

Primary Prevention of Coronary Heart Disease in Women through Diet and Lifestyle. Meir J. Stampfer, Frank B. Hu, JoAnn E. Manson, Eric B. Rimm and Walter C. Willett in *New England Journal of Medicine,* Vol. 343, No. 1, pages 16–22; July 6, 2000.

Eat, Drink, and Be Healthy: The Harvard Medical School Guide to Healthy Eating. Walter C. Willett, P. J. Skerrett and Edward L. Giovannucci. Simon & Schuster, 2001.

Dietary Reference Intakes for Energy, Carbohydrates, Fiber, Fat, Protein and Amino Acids (Macronutrients). Food and Nutrition Board, Institute of Medicine, National Academy of Sciences. National Academies Press, 2002. Available online at **www.nap.edu/books/0309085373/html/**

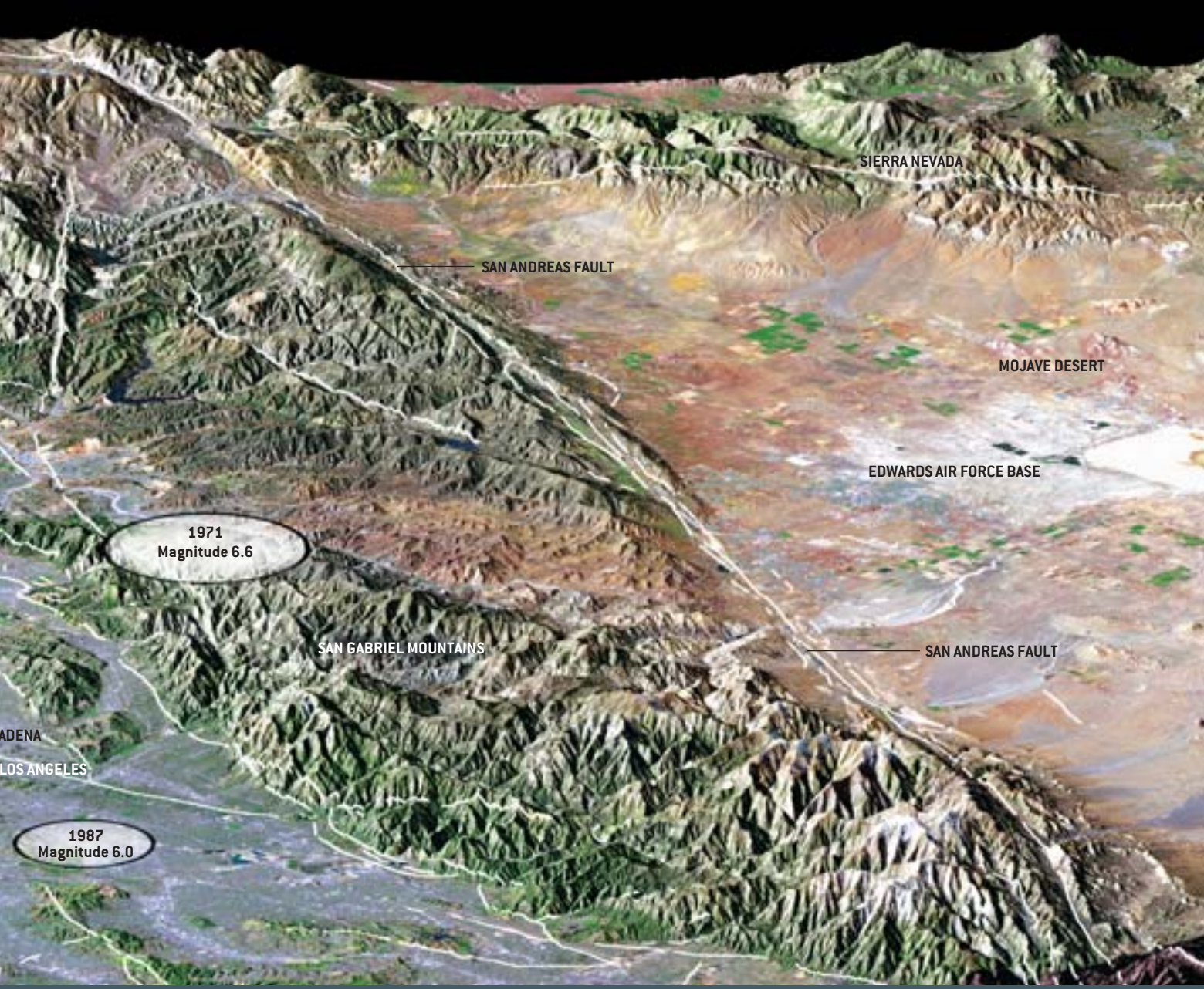


THIRTEEN MILLION PEOPLE in and around Los Angeles live among a complex network of earthquake-prone faults (*white lines*), including the famous San Andreas. Some scientists now think that each of the major shocks that have rocked the region since the 1850s (*ovals*) probably influenced the timing and location of subsequent ones.

Earthquake Conversations

Contrary to prevailing wisdom, large earthquakes can interact in unexpected ways. This exciting discovery could dramatically improve scientists' ability to pinpoint future shocks

By Ross S. Stein



For decades, earthquake experts dreamed of being able to divine the time and place of the world's next disastrous shock. But by the early 1990s the behavior of quake-prone faults had proved so complex that they were forced to conclude that the planet's largest tremors are isolated, random and utterly unpredictable. Most seismologists now assume that once a major earthquake and its expected aftershocks do their damage, the fault will remain quiet until stresses in the earth's crust have time to rebuild, typically over hundreds or thousands of years. A recent discovery—that earthquakes interact in ways never before imagined—is beginning to overturn that assumption.

This insight corroborates the idea that a major shock relieves stress—and thus the likelihood of a second major tremor—in some areas. But it also suggests that the probability of a succeeding earthquake elsewhere along the fault or on a nearby fault can actually jump by as much as a factor of three. To the people who must stand ready to provide emergency services or to those who set prices for insurance premiums, these refined predictions can be critical in determining which of their constituents are most vulnerable.

At the heart of this hypothesis—known as stress triggering—is the realization that faults are unexpectedly responsive to subtle stresses they acquire as neighboring faults shift and shake. Drawing on records of past tremors and novel calculations of fault behavior, my colleagues and I have learned that the stress relieved during an earthquake does not simply dissipate; instead it moves down the fault and concentrates in sites nearby. This jump in stress promotes subsequent tremors. Indeed, studies of about two dozen faults since 1992 have convinced many of us that earthquakes can be triggered even when the stress swells by as little as one eighth the pressure required to inflate a car tire.

Such subtle cause-and-effect relations among large shocks were not thought to exist—and never played into seismic forecasting—until now. As a result, many scientists have been understandably skeptical about embracing this basis for a new approach to forecasting. Nevertheless, the stress-triggering hypothesis has continued to gain credibility through its ability to explain the location and frequency of earthquakes that followed several destructive shocks in California, Japan and Turkey. The hope of furnishing better warnings for such disasters is the primary motivation behind our ongoing quest to interpret these unexpected conversations between earthquakes.

Aftershocks Ignored

CONTRADICTING the nearly universal theory that major earthquakes strike at random was challenging from the start—especially considering that hundreds of scientists searched in vain for more than three decades to find predictable patterns in global earthquake activity, or seismicity. Some investigators looked for changing rates of small tremors or used sensitive instruments to measure the earth's crust as it tilts, stretches and migrates across distances invisible to the naked eye. Others tracked underground movements of gases, fluids and electromagnetic energy or monitored tiny cracks in the rocks to see whether they open or close before large shocks. No matter what the researchers examined, they found little consistency from one major earthquake to another.

Despite such disparities, historical records confirm that about one third of the world's recorded tremors—so-called aftershocks—cluster in space and time. All true aftershocks were thought to hit somewhere along the segment of the fault that slipped during the main shock. Their timing also follows a routine pattern, according to observations first made in 1894 by Japanese seismologist Fusakichi Omori and since developed into a basic principle known as Omori's law. Aftershocks are most abundant immediately after a main shock. Ten days later the rate of aftershocks drops to 10 percent of the initial rate, 100 days later it falls to 1 percent, and so on. This predictable jump and decay in seismicity means that an initial tremor modifies the earth's crust in ways that raise the prospect of succeeding ones, contradicting the view that earthquakes occur randomly in time. But because aftershocks are typically smaller than the most damaging quakes scientists would like to be able to predict, they were long overlooked as a key to unlocking the secrets of seismicity.

Once aftershocks are cast aside, the remaining tremors indeed appear—at least at first glance—to be random. But why ignore the most predictable earthquakes to prove that the rest are without order? My colleagues and I decided to hunt instead for what makes aftershocks so regular. We began our search in one of the world's most seismically active regions—the San Andreas Fault system that runs through California. From local records of earthquakes and aftershocks, we knew that on the day following a magnitude 7.3 event, the chance of another large shock striking within 100 kilometers is nearly 67 percent—20,000 times the likelihood on any other day. Something about the first shock seemed to dramatically increase the odds of subsequent ones, but what?

That big leap in probability explains why no one was initially surprised in June 1992 when a magnitude 6.5 earthquake struck near the southern California town of Big Bear only three hours after a magnitude 7.3 shock occurred 40 kilometers away, near Landers. (Fortunately, both events took place in the sparsely populated desert and left Los Angeles unscathed.) The puzzling contradiction to prevailing wisdom was that the Big Bear shock struck far from the fault that had slipped during Landers's shaking. Big Bear fit the profile of an aftershock in its timing but not in its location. We suspected that its mysterious placement might hold the clue we were looking for.

By mapping the locations of Landers, Big Bear and hundreds of other California earthquakes, my colleagues and I began to notice a remarkable pattern in the distribution not only of true aftershocks but also of other, smaller earthquakes that

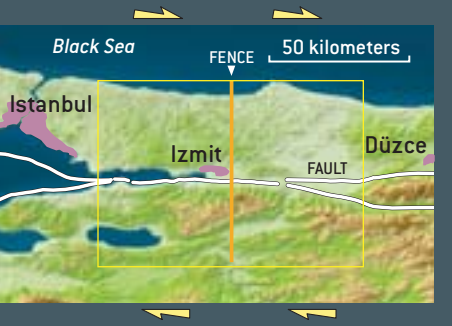
Overview/*Shifting Priorities*

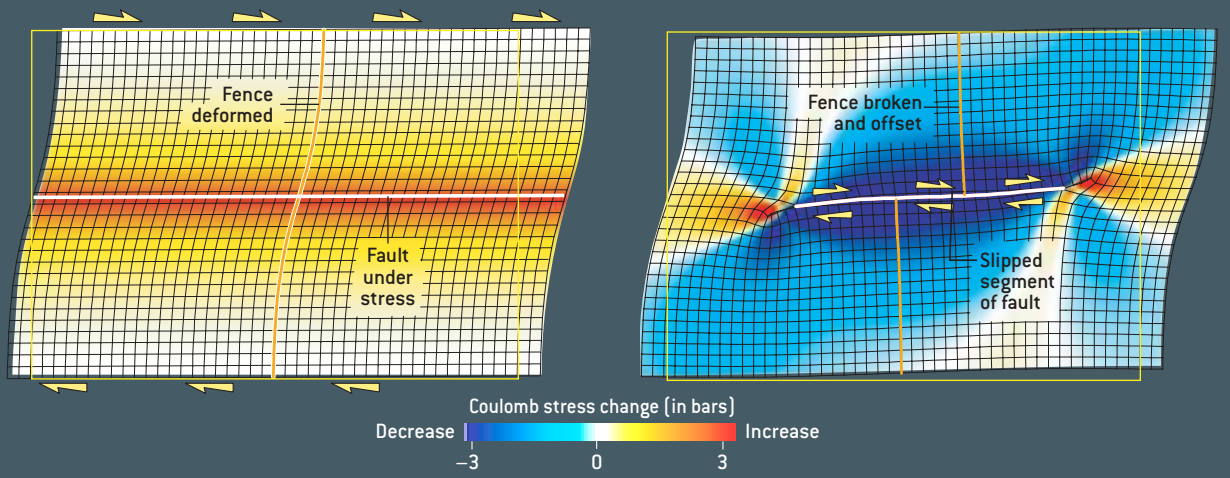
- Scientists used to think that one large earthquake has no notable influence on the timing or location of the next one, but a surprising new discovery is challenging that perspective.
- Earthquake-prone faults are now proving to be unexpectedly responsive to subtle stresses they acquire during shocks that strike along nearby faults.
- All else being equal, the regions in the earth's crust where the stress rises—even by a small amount—will be the sites of the next earthquakes.
- If this hypothesis turns out to be right, its implications could dramatically improve the ability of nations, cities and individuals to evaluate their earthquake vulnerability.

ROBERT E. CRIPPEN *Jet Propulsion Laboratory (preceding pages)*

STRESSED OUT

BUILDUP AND RELEASE of the stress that accumulates slowly as the earth's tectonic plates grind past one another mark the cycle of all great earthquakes. Along Turkey's North Anatolian fault (*white line*), the land north of the fault is moving eastward relative to the land to the south (*yellow arrows*) but gets stuck along the fault. When the stress finally overcomes friction along the fault, the rocks on either side slip past one another violently. A catastrophic example of this phenomenon occurred on August 17, 1999, when a magnitude 7.4 shock took 25,000 lives in and around the city of Izmit. Calculations of stress before and after the Izmit earthquake (*below*) reveal that, after the shock, the so-called Coulomb stress dropped along the segment of the fault that slipped but increased elsewhere. —*R.S.S.*





BEFORE THE EARTHQUAKE

The segment of the North Anatolian fault near Izmit accumulated significant stress (*red*) during the 200 years since its last major stress-relieving shock. An imaginary deformed fence and grid superimposed over the landscape illustrate this high stress. Squares along the fault are stretched into parallelograms (*exaggerated 15,000 times*), with the greatest change in shape, and thus stress, occurring closest to the fault.

AFTER THE EARTHQUAKE

The earthquake relieved stress (*blue*) all along the segment of the fault that slipped. The formerly deformed fence broke and became offset by several meters at the fault, and the grid squares closest to the fault returned to their original shape. High stress is now concentrated beyond both ends of the failed fault segment, where the grid squares are more severely contorted than before the shock struck.

follow a main shock by days, weeks or even years. Like the enigmatic Big Bear event, a vast majority of these subsequent tremors tended to cluster in areas far from the fault that slipped during the earthquake and thus far from where aftershocks are supposed to occur [*see box on page 78*]. If we could determine what controlled this pattern, we reasoned, the same characteristics might also apply to the main shocks themselves. And if that turned out to be true, we might be well on our way to developing a new strategy for forecasting earthquakes.

Triggers and Shadows

WE BEGAN BY LOOKING at changes within the earth's crust after major earthquakes, which release some of the stress that accumulates slowly as the planet's shifting tectonic plates grind past each other. Along the San Andreas Fault, for instance, the plate carrying North America is moving south relative to the one that underlies the Pacific Ocean. As the two sides move in opposite directions, shear stress is exerted parallel to the plane of the fault; as the rocks on opposite sides of the fault press against each other, they exert a second stress, perpendicular to the fault plane. When the shear stress exceeds the frictional resistance on the fault or when the stress pressing the two sides of the fault together is eased, the rocks on either side will slip past each other suddenly, releasing tremendous energy in the form of an earthquake. Both components of stress, which when added together are called Coulomb stress, diminish along the segment of the fault that slips. But because that stress cannot simply disappear, we knew it must be redistributed to other points along the same fault or to other faults nearby. We also suspected that this increase in Coulomb stress could be sufficient to trigger earthquakes at those new locations.

Geophysicists had been calculating Coulomb stresses for years, but scientists had never used them to explain seismicity. Their reasoning was simple: they assumed that the changes

SERKAN BOZKURT *U.S. Geological Survey*

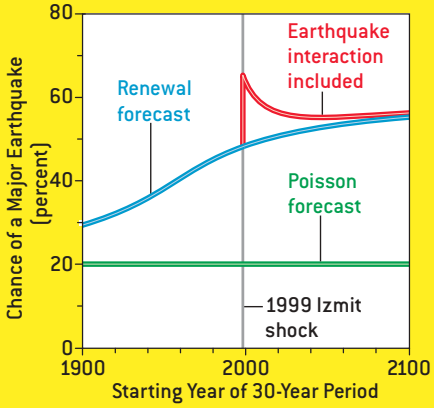
Forecasting under Stress

So many features of earthquake behavior are still unknown that scientists have directed their limited insight to playing the odds

HOW PEOPLE PERCEIVE the threat of an earthquake in their part of the world depends in great part on what kind of warnings are presented to them. Most of today's seismic forecasts assume that one earthquake is unrelated to the next. Every fault segment is viewed as having an average time period between tremors of a given size—the larger the shock, the greater the period, for example—but the specific timing of the shocks is believed to be random. The best feature of this method, known as a Poisson probability, is that a forecast can be made without knowing when the last significant earthquake occurred. Seismologists can simply infer the typical time period between major shocks based on geologic records of much older tremors along that segment of the fault. This conservative strategy yields odds that do not change with time.

In contrast, a more refined type of forecast called the renewal probability predicts that the chances of a damaging shock climb as more time passes since the last one struck. These growing odds are based on the assumption that stress along a fault increases gradually in the wake of a major earthquake. My colleagues and I build the probabilities associated with earthquake interactions on top of this second traditional technique by including the effects of stress changes imparted by nearby earthquakes. Comparing the three types of forecasts for Turkey's North Anatolian fault near Istanbul illustrates their differences, which are most notable immediately after a major shock.

In the years leading up to the catastrophic Izmit earthquake of August 1999, the renewal probability of a shock of magnitude 7 or greater on the four faults within 50 kilometers of Istanbul had been rising slowly since the last large earthquake struck each of them, between 100 and 500 years ago. According to this type of forecast, the August shock created a sharp drop in the likelihood of a second major tremor in the immediate vicinity of Izmit, because the faults there were thought to have relaxed. But the quake caused no change in the 48 percent chance of severe shaking 100 kilometers to the west, in Istanbul, sometime in the next 30 years. Those odds will continue to grow slowly with time—unlike the Poisson probability, which will remain at only 20 percent regardless of other tremors that may occur near the capital city.



PREDICTED ODDS of a major earthquake striking within 50 kilometers of Istanbul can vary dramatically. The odds, which stay the same or rise slowly with time in traditional forecasts (*green* and *blue*), jump significantly when stresses transferred during the 1999 Izmit earthquake are included (*red*).

When the effects of my team's new stress-triggering hypothesis were added to the renewal probability, everything changed. The most dramatic result was that the likelihood of a second quake rocking Istanbul shot up suddenly because some of the stress relieved near Izmit during the 1999 shock moved westward along the fault and concentrated closer to the city. That means the Izmit shock raised the probability of an Istanbul quake in the next 30 years from 48 percent to 62 percent. This so-called interaction probability will continue to decrease over time as the renewal probability climbs. The two forecasts will then converge at about 54 percent in the year 2060—assuming the next major earthquake doesn't occur before then.

—R.S.S.

CRUMPLED BUILDINGS dot Düzce, Turkey, in the wake of a major tremor that struck in November 1999. Some scientists suspect that this disaster was triggered by an earlier shock near Izmit.

CORNELIA BLIK (*graph*); AFB *Corbis* (*photograph*)

were too meager to make a difference. Indeed, the amount of stress transferred is generally quite small—less than 3.0 bars, or at most 10 percent of the total change in stress that faults typically experience during an earthquake. I had my doubts about whether this could ever be enough to trigger a fault to fail. But when Geoffrey King of the Paris Geophysical Institute, Jian Lin of the Woods Hole Oceanographic Institution in Massachusetts and I calculated the areas in southern California where stress had increased after major earthquakes, we were amazed to see that the increases—small though they were—matched clearly with sites where the succeeding tremors had clustered. The implications of this correlation were unmistakable: regions where the stress rises will harbor the majority of subsequent earthquakes, both large and small. We also began to see something equally astonishing: small reductions in stress could inhibit future tremors. On our maps, earthquake activity plummeted in these so-called stress shadows.

Coulomb stress analysis nicely explained the locations of certain earthquakes in the past, but a more important test would be to see whether we could use this new technique to forecast the sites of *future* earthquakes reliably. Six years ago I joined geophysicist James H. Dieterich of the U.S. Geological Survey (USGS) and geologist Aykut A. Barka of Istanbul Technical University to assess Turkey's North Anatolian fault, among the world's most heavily populated fault zones. Based on our calculations of where Coulomb stress had risen as a result of past earthquakes, we estimated that there was a 12 percent chance that a magnitude 7 shock or larger would strike the segment of the fault near the city of Izmit sometime between 1997 and 2027. That may seem like fairly low odds, but in comparison, all but one other segment of the 1,000-kilometer-long fault had odds of only 1 to 2 percent.

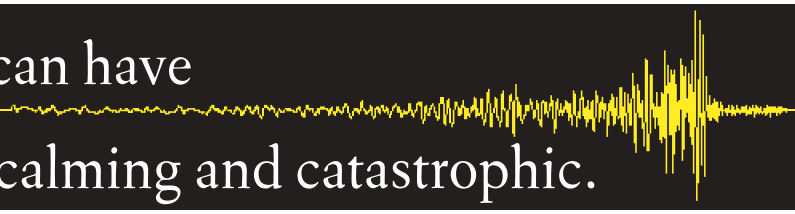
We did not have to wait long for confirmation. In August 1999 a magnitude 7.4 quake devastated Izmit, killing 25,000 people and destroying more than $6.5 billion worth of property. But this earthquake was merely the most recent in a falling-domino-style sequence of 12 major shocks that had struck the North Anatolian fault since 1939. In a particularly brutal five-year period, fully 700 kilometers of the fault slipped in a deadly westward progression of four shocks. We suspected that stress transferred beyond the end of each rupture triggered the successive earthquake, including Izmit's.

In November 1999 the 13th domino fell. Some of the Coulomb stress that had shifted away from the fault segment near Izmit triggered a magnitude 7.1 earthquake near the town of Düzce, about 100 kilometers to the east. Fortunately, Barka had calculated the stress increase resulting from the Izmit shock and had published it in the journal *Science* two months earlier. Barka's announcement had emboldened engineers to close school buildings in Düzce that were lightly damaged by the first shock despite pleas by school officials who said that students had nowhere else to gather for classes. Some of these buildings were flattened by the November shock.

If subsequent calculations by Parsons of the USGS, Shinji Toda of Japan's Active Fault Research Center, Barka, Dieterich and me are correct, that may not be the last of the Izmit quake's aftermath. The stress transferred during that shock has also raised the probability of strong shaking in the nearby capital, Istanbul, sometime this year from 1.9 percent to 4.2 percent. Over the next 30 years we estimate those odds to be 62 percent; if we assumed large shocks occur randomly, the odds would be just 20 percent [*see box on opposite page*].

The stress-triggering hypothesis offers some comfort alongside such gloom and doom. When certain regions are put on high alert for earthquakes, the danger inevitably drops in others. In Turkey the regions of reduced concern happen to be sparsely populated relative to Istanbul. But occasionally the opposite is true. One of the most dramatic examples is the relative lack of seismicity that the San Francisco Bay Area, now home to five million people, has experienced since the great magnitude 7.9 earthquake of 1906. A 1998 analysis by my USGS colleagues Ruth A. Harris and Robert W. Simpson demonstrated that the stress shadows of the 1906 shock fell across several parallel strands of the San Andreas Fault in the Bay Area, while the stress increases occurred well to the north and south. This could explain why the rate of damaging shocks in the Bay Area dropped by an order of magnitude compared with the 75 years preceding 1906. Seismicity in the Bay Area is calculated to slowly emerge from this shadow as stress rebuilds on the faults;

Even tiny stress changes can have momentous effects, both calming and catastrophic.



UNIVERSITY OF SOUTHERN CALIFORNIA AND CONSORTIUM OF ORGANIZATIONS FOR STRONG-MOTION OBSERVATION SYSTEMS (*reversed seismogram of Whittier Narrows, Calif., earthquake*)

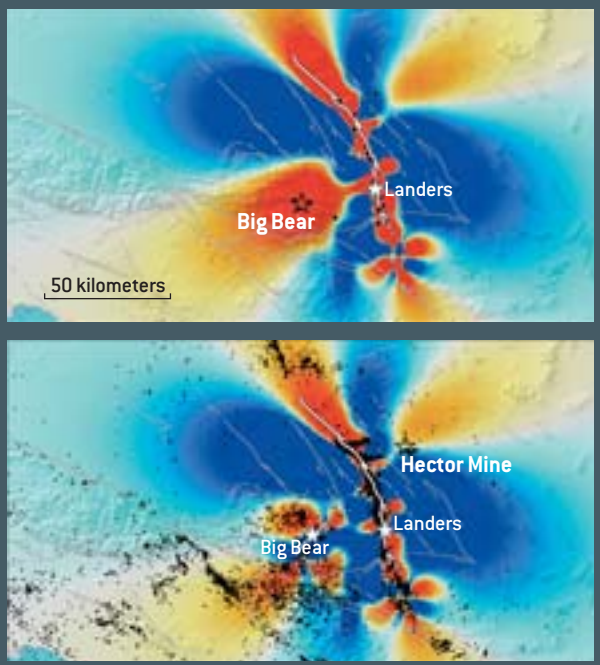
THE AUTHOR

ROSS S. STEIN is a geophysicist with the U.S. Geological Survey's Earthquake Hazards Team in Menlo Park, Calif. He joined the survey in 1981 after earning a Ph.D. from Stanford University in 1980 and serving as a postdoctoral fellow at Columbia University. Stein's research, which has been devoted to improving scientists' ability to assess earthquake hazards, has been funded by U.S. agencies such as the Office of Foreign Disaster Assistance and by private companies, including the European insurance company Swiss Re. For the work outlined in this article, Stein received the Eugene M. Shoemaker Distinguished Achievement Award of the USGS in 2000. He also presented the results in his Frontiers of Geophysics Lecture at the annual meeting of the American Geophysical Union in 2001. Stein has appeared in several TV documentaries, including *Great Quakes: Turkey* (Learning Channel, 2001).

EARTHQUAKE CLUSTERS

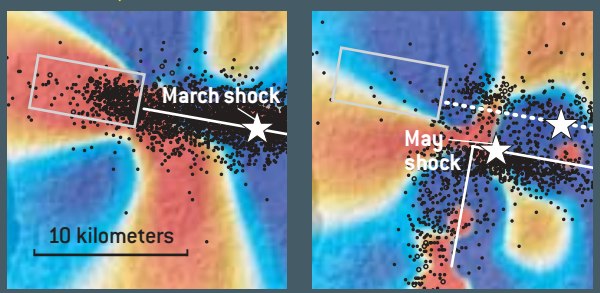
PLACES WHERE STRESS JUMPS (*red*) after major earthquakes (*filled stars*) tend to be the sites of subsequent tremors, both large (*open stars*) and small (*black dots*). Conversely, few tremors occur where the stress plummets (*blue*), regardless of the location of nearby faults (*white lines*). —*R.S.S.*

SOUTHERN CALIFORNIA, U.S.



MAGNITUDE 7.3 SHOCK in the southern California desert near Landers in 1992 increased the expected rate of earthquakes to the southwest, where the magnitude 6.5 Big Bear shock struck three hours later (*top*). Stresses imparted by the combination of the Landers and Big Bear events coincided with the regions where the vast majority of tremors occurred over the next seven years, culminating with the magnitude 7.1 Hector Mine quake in 1999 (*bottom*).

KAGOSHIMA, JAPAN



TWIN EARTHQUAKES can turn the rate of earthquake occurrence, or seismicity, up and down in the same spot. In March 1997 a magnitude 6.5 tremor increased stress and seismicity to the west of the ruptured fault (*above left*). Seismicity in that area then dropped along with stress (*above right*) following a magnitude 6.3 shock that struck 48 days later three kilometers to the south.

the collapsed highways and other damage wrought by the 1989 Loma Prieta shock may be a harbinger of this reawakening.

Bolstering the Hypothesis

EXAMINATIONS OF the earthquakes in Turkey and in southern California fortified our assertion that even tiny stress changes can have momentous effects, both calming and catastrophic. But despite the growing number of examples we had to support this idea, one key point was difficult to explain: roughly one quarter of the earthquakes we examined occurred in areas where stress had *decreased*. It was easy for our more skeptical colleagues to argue that no seismicity should occur in these shadow zones, because the main shock would have relieved at least some stress and thus pushed those segments of the fault further from failure. We now have an answer. Seismicity never shuts off completely in the shadow zones, nor does it turn on completely in the trigger zones. Instead the *rate* of seismicity—the number of earthquakes per unit of time—merely drops in the shadows or climbs in the trigger zones relative to the preceding rate in that area.

We owe this persuasive extension of stress triggering to a theory proposed by Dieterich in 1994. Known as rate/state friction, it jettisons the comfortable concept of friction as a property that can only vary between two values—high friction when the material is stationary and lower friction when it is sliding. Rather, faults can become stickier or more slippery as the rate of movement along the fault changes and as the history of motion, or the state, evolves. These conclusions grew out of lab experiments in which Dieterich's team sawed a miniature fault into a Volkswagen-size slab of granite and triggered tiny earthquakes.

When earthquake behavior is calculated with friction as a variable rather than a fixed value, it becomes clear that Omori's law is a fundamental property not just of so-called aftershocks but of *all* earthquakes. The law's prediction that the rate of shocks will first jump and then diminish with time explains why a region does not forever retain the higher rate of seismicity that results from an increase in stress. But that is only half the story. Dieterich's theory reveals a characteristic of the seismicity that Omori's law misses entirely. In areas where a main shock relieves stress, the rate of seismicity immediately plunges but will slowly return to preshock values in a predictable manner. These points may seem subtle, but rate/state friction allowed us for the first time to make predictions of how jumps or declines in seismicity would change over time. When calculating Coulomb stresses alone, we could define the general location of new earthquakes but not their timing.

Our emerging ideas about both the place and the time of stress-triggered earthquakes were further confirmed by a global study conducted early last year. Parsons considered the more than 100 earthquakes of magnitude 7 or greater that have occurred worldwide in the past 25 years and then examined all subsequent shocks of at least magnitude 5 within 250 kilometers of each magnitude 7 event. Among the more than 2,000 shocks in this inventory, 61 percent occurred at sites where a preceding shock increased the stress, even by a small amount.

SERKAN BOZKURT *USGS*, KEITH RICHARDS-DINGER *U.S. Navy* AND SHINJI TODA *Japan's Active Fault Research Center*

Few of these triggered shocks were close enough to the main earthquake to be considered an aftershock, and in all instances the rate of these triggered tremors decreased in the time period predicted by rate/state friction and Omori's law.

Now that we are regularly incorporating the concept of rate/state friction into our earthquake analyses, we have begun to uncover more sophisticated examples of earthquake interaction than Coulomb stress analyses alone could have illuminated. Until recently, we had explained only relatively simple situations, such as those in California and Turkey, in which a large earthquake spurs seismicity in some areas and makes it sluggish in others. We knew that a more compelling case for the stress-triggering hypothesis would be an example in which successive, similar-size shocks are seen to turn the frequency of earthquakes up and down in the same spot, like a dimmer switch on an electric light.

Toda and I discovered a spectacular example of this phenomenon, which we call toggling seismicity. Early last year we began analyzing an unusual pair of magnitude 6.5 earthquakes that struck Kagoshima, Japan, in 1997. Immediately following the first earthquake, which occurred in March, a sudden burst of seismicity cropped up in a 25-square-kilometer region just beyond the west end of the failed segment of the fault. When we calculated where the initial earthquake transferred stress, we found that it fell within the same zone as the heightened seismicity. We also found that the rate immediately began decaying just as rate/state friction predicted. But when the second shock struck three kilometers to the south only seven weeks later, the region of heightened seismicity experienced a sudden, additional drop of more than 85 percent. In this case, the trigger zone of the first earthquake had fallen into the shadow zone of the second one. In other words, the first quake turned seismicity up, and the second one turned it back down.

A New Generation of Forecasts

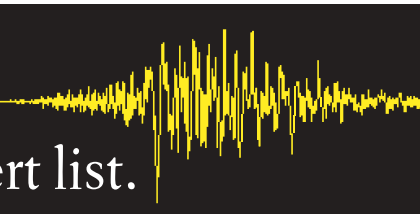
EAVESDROPPING ON the conversations between earthquakes has revealed, if nothing else, that seismicity is highly interactive. And although phenomena other than stress transfer may influence these interactions, my colleagues and I believe that enough evidence exists to warrant an overhaul of traditional probabilistic earthquake forecasts. By refining the likelihood of dangerous tremors to reflect subtle jumps and declines in stress, these new assessments will help governments, the insurance industry and the public at large to better evaluate their earthquake risk. Traditional strategies already make some degree of prioritizing possible, driving the strengthening of buildings and other precautions in certain cities or regions at the expense of others. But our analyses have shown that taking stress triggering into account will raise different faults to the top of the high-alert list than using traditional methods alone will. By the same token, a fault deemed dangerous by traditional practice may actually be a much lower risk.

An important caveat is that any type of earthquake forecast is difficult to prove right and almost impossible to prove wrong. Regardless of the factors that are considered, chance plays a tremendous role in whether a large earthquake occurs, just as it does in whether a particular weather pattern produces a rainstorm. The meteorologists' advantage over earthquake scientists is that they have acquired millions more of the key measurements that help improve their predictions. Weather patterns are much easier to measure than stresses inside the earth, after all, and storms are much more frequent than earthquakes.

Refining earthquake prediction must follow the same path, albeit more slowly. That is why my team has moved forward by building an inventory of forecasts for large earthquakes near the shock-prone cities of Istanbul, Landers, San Francisco and Kobe. We are also gearing up to make assessments for Los Angeles and Tokyo, where a major earthquake could wreak trillion-dollar devastation. Two strong shocks along Alaska's Denali fault in the fall of 2002—magnitude 6.7 on October 23 and magnitude 7.9 on November 3—appear to be another stress-triggered sequence. Our calculations suggest that the first shock increased the likelihood of the second by a factor of 100 during the intervening 10 days. We are further testing the theory by forecasting smaller, nonthreatening earthquakes, which are more numerous and thus easier to predict.

In the end, the degree to which any probabilistic forecast will protect people and property is still uncertain. But scientists have plenty of reasons to keep pursuing this dream: several hundred million people live and work along the world's most active fault zones. With that much at stake, stress triggering—or any other phenomenon that has the potential to raise the odds of a damaging earthquake—should not be ignored. SA

Taking stress triggering into account will raise different faults to the top of the high-alert list.



CALIFORNIA STRONG MOTION INSTRUMENTATION PROGRAM AND CONSORTIUM OF ORGANIZATIONS FOR STRONG-MOTION OBSERVATION SYSTEMS (*seismogram of Hector Mine, Calif., earthquake*)

MORE TO EXPLORE

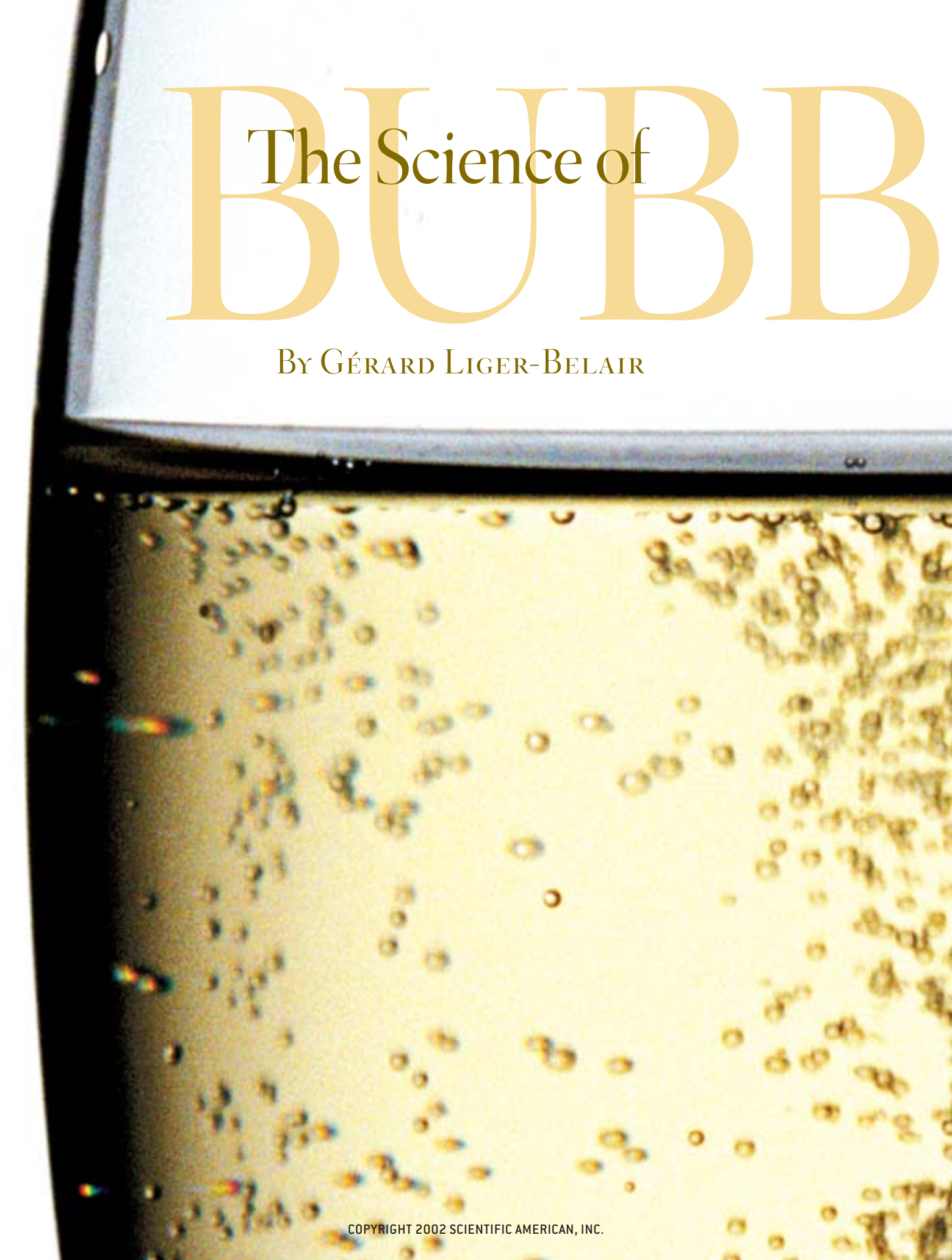
Earthquakes Cannot Be Predicted. Robert J. Geller, David D. Jackson, Yan Y. Kagan and Francesco Mulargia in *Science,* Vol. 275, page 1616; March 14, 1997. **http://scec.ess.ucla.edu/~ykagan/perspective.html**

Heightened Odds of Large Earthquakes Near Istanbul: An Interaction-Based Probability Calculation. Tom Parsons, Shinji Toda, Ross S. Stein, Aykut Barka and James H. Dieterich in *Science,* Vol. 288, pages 661–665, April 28, 2000.

View earthquake animations and download Coulomb 2.2 (Macintosh software and tutorial for calculating earthquake stress changes) at **http://quake.usgs.gov/~ross**

The Science of BUBB

By Gérard Liger-Belair



LY

Scientists study the nose-tickling effervescence of champagne—an alluring and unmistakable aspect of its appeal

Pop open a bottle of champagne and pour yourself a glass. Take a sip. The elegant surface fizz—a boiling fumarole of rising and collapsing bubbles—launches thousands of golden droplets into the air, conveying the wine's enticing flavors and aromas to tongue and nostrils alike. A percussive symphony of diminutive pops accompanies the tasty mouthful, juxtaposing a refreshing carbonated chill and a comforting alcoholic warmth. Such is the enchantment of bubbly, the classic sparkling wine of northeastern France's Champagne district, a libation that has become a fixture at festive celebrations worldwide.

Among the hallmarks of a good champagne are multiple bubble trains rising in lines from the sides of a poured glass like so many tiny hot-air balloons. When they reach the surface, the bubbles form a ring, the so-called *collerette,* at the top of a filled flute. Although no scientific evidence correlates the quality of a champagne with the fineness of its bubbles, people nonetheless often make a connection between the two. Because ensuring the traditional effervescent personality of champagne is big business, it has become important for *champenois* vintners to achieve the perfect *petite* bubble. Therefore, a few years ago several research colleagues from the University of Reims Champagne-Ardenne and Moët & Chandon and I decided to examine the behavior of bubbles in carbonated

ON THE RISE: The short but spectacular ascent of champagne bubbles to the top of a glass embodies a complex physicochemical system that is both more functional and more picturesque than one might expect.

beverages. Our goal was to determine, illustrate and finally better understand the role played by each of the many parameters involved in the bubbling process. The simple but close observation of a glass filled with sparkling wine, beer or soda revealed an unexplored and visually appealing phenomenon. Our initial results concerned the three main phases of a bubble's life—birth, ascent and collapse.

Bubble Genesis

IN CHAMPAGNE, sparkling wines and beers, carbon dioxide is principally responsible for producing gas bubbles, which form when yeast ferments sugars, converting them into alcohol and carbon dioxide molecules. Industrial carbonation is the source of the fizz in soda drinks. After bottling or canning, equilibrium is established between the carbon dioxide dissolved in the liquid and the gas in the space directly under the cork, cap or tab, according to Henry's law. That law states that the amount of gas dissolved in a fluid is proportional to the pressure of the gas with which it is in equilibrium.

When the container is opened, the pressure of the gaseous carbon dioxide above the liquid falls abruptly, breaking the prevailing thermodynamic equilibrium. As a result, the liquid is supersaturated with carbon dioxide molecules. To regain a thermodynamic stability corresponding to atmospheric pressure, carbon dioxide molecules must leave the supersaturated fluid. When the beverage is poured into a glass, two mechanisms enable dissolved carbon dioxide to escape: diffusion through the free surface of the liquid and bubble formation.

To cluster into embryonic bubbles, however, dissolved carbon dioxide molecules must push their way through aggregated liquid molecules, which are strongly linked by van der Waals forces (dipole attraction). Thus, bubble formation is limited by this energy barrier; to overcome it requires supersaturation ratios higher than those that occur in carbonated beverages.

In weakly supersaturated liquids, including champagne, sparkling wines, beers and sodas, bubble formation requires preexisting gas cavities with radii of curvature large enough to overcome the nucleation energy barrier and grow freely. This is the case because the curvature of the bubble interface leads to an excess of pressure inside the gas pocket that is inversely proportional to its radius (according to Laplace's law). The smaller the bubble, the higher the overpressure within it. Below a critical radius, the pressure excess in a gas pocket forbids dissolved carbon dioxide to diffuse into it. In newly opened champagne, the critical radius is submicrometric, around 0.2 micron.

THE AUTHOR

GÉRARD LIGER-BELAIR is associate professor at the University of Reims Champagne-Ardenne in France, where he studies the physical chemistry of bubbles in carbonated beverages. He also consults for the research department of Moët & Chandon. Liger-Belair splits his time between the science of thin films and interfaces and high-speed photography, the results of which have appeared in many exhibitions.

The author would like to thank Philippe Jeandet and Michèle Vignes-Adler for their valuable discussions, the Europôl'Agro Institute and Jean-Claude Colson and his Recherche Oenologie Champagne Université association for their financial support, and Moët & Chandon, Champagne Pommery and Verrerie Cristallerie d'Arques for their collaborative efforts.

EISENHUT & MAYER *Getty Images (preceding pages and this page)*; COURTESY OF GÉRARD LIGER-BELAIR *(insets)*

CHAMPAGNE BUBBLES

LIFE CYCLE OF A BUBBLE: The brief existence of a champagne bubble begins on a tiny mote of cellulose left clinging to the sides of a glass after it is wiped dry (*bottom*). When the sparkling wine is poured, a submicron-diameter gas pocket forms on the cellulose fiber. Dissolved carbon dioxide under pressure enters this small cavity and eventually expands it so much that buoyancy causes it to separate from the nucleation site. During its journey to the surface, the bubble enlarges further as more carbon dioxide makes its way in (*middle*). Simultaneously, aromatic flavor molecules in the beverage attach themselves to the gas/water membrane, a phenomenon that slows the bubble's ascent by boosting its drag. Soon after emerging at the surface, the flavor-carrying gas capsule collapses on itself and shoots a bit of the wine into the air, thus enhancing the champagne's smell and savor (*top*).

SHOW ME THE BUBBLES: Champagne is best served in a long-stemmed flute—a slender glass with a tulip- or cone-shaped bowl containing six ounces or more (*left*), say wine experts. The tall, narrow configuration of the flute extends and highlights the flow of bubbles to the crown, where the restricted open surface area concentrates the aromas carried up with the bubbles and released by their collapse. The slender stemware also prolongs the drink's chill and helps retains its effervescence.

The flute is better suited for drinking sparkling wine than the "champagne" coupe, the short-stemmed, saucer-shaped glasses that were once so fashionable (*see bottom of box on next page*). Legend has it that these glasses were originally modeled on the celebrated breasts of Marie Antoinette, queen of France in the late 18th century. Although it is still popular, the coupe was never designed for drinking champagne and thus does not allow the taster to fully enjoy its qualities, according to wine connoisseurs. Shallow and wide-brimmed, coupe glasses tend to be unstable and too likely to spill, and they fail to keep the wine as cold as tall wine flutes do. Further, coupes do not present the elegant bubble trains to best advantage.

To open a bottle of champagne, hold it with the cork pointed at a 45-degree angle upward, away from you or anyone else. Hold the cork and gently turn the bottle in one direction. Rather than popping off with a bang, the cork should come off with a subdued sigh. A loud pop wastes bubbles; as the saying goes, "The ear's gain is the palate's loss." —*G.L.-B.*

To observe bubble production sites (or "bubble nurseries") in detail, we aimed a high-speed video camera fitted with a microscope objective lens at the bases of hundreds of bubble trains. Contrary to general belief, these nucleation sites are not located on irregularities on the glass surface, which feature length scales far below the critical radii of curvature required for bubble formation. Bubble nurseries, it turns out, arise on impurities *attached* to the glass wall. Most are hollow and roughly cylindrical cellulose fibers that fell out of the air or remained from the process of wiping the glass dry. Because of their geometries, these exogenous particles cannot be wet completely by the beverage and are thus able to entrap gas pockets when a glass is filled [*see bottom illustration at left*].

During bubble formation, dissolved carbon dioxide molecules migrate into the minute gas pockets. Eventually a macroscopic bubble grows, which initially remains rooted to its nucleation site because of capillary forces. Finally the bubble's increasing buoyancy causes it to detach, providing the opportunity for a new bubble to form. The process repeats until bubble production ebbs because of a dearth of dissolved carbon dioxide.

Cyclic bubble production at a nucleation site is characterized by its bubbling frequency—that is, the number of bubbles produced per second, a figure that can be illustrated using a stroboscope. When the strobe's flash frequency equals that of bubble production, the bubble train appears frozen.

Because the kinetics of bubble growth depend also on the dissolved carbon dioxide content, bubble formation frequencies vary from one carbonated beverage to another. In champagne, for example, in which the gas content is approximately three times that of beer, the most active nucleation sites emit up to about 30 bubbles per second, whereas beer bubble nurseries produce up to only about 10 bubbles per second.

Bubble Ascent

AFTER A BUBBLE is released from its nucleation site, it grows as it makes its way to the surface [*see middle illustration at left*]. Bubble enlargement during ascent is caused by a continuous diffusion of dissolved carbon dioxide through the bubble's gas/liquid interface. Buoyancy increases as bubbles expand, causing them to accelerate continuously and separate from one another on the way up.

Beers and sparkling wines are not pure liquids. In addition to alcohol and dissolved carbon dioxide, they contain many other organic compounds that can display surface activity like that of a soap molecule. These surfactants, comprising mostly proteins and glycoproteins, have a water-soluble and a water-insoluble part. Rather than staying dissolved in the bulk liquid, surfactants prefer to gather around the surface of a bubble with their hydrophobic ends aimed into the gas and their hydrophilic ends stuck in the liquid.

The surfactant coating around a bubble becomes crucial to its behavior when growing buoyancy induces detachment and forces the gas pocket to plow its way through the intervening liquid molecules. Adsorbed surfactant molecules stiffen a bubble by forming something like a shield on its surface. According to fluid dynamical theory, a rigid sphere rising through a fluid runs into greater resistance than a more flexible sphere with a surface free of surfactants. In addition, surfactant molecules encountered during ascent gradu-

ally collect on the bubble surface, enlarging its immobile area. Therefore, the hydrodynamic drag experienced by a rising bubble of fixed radius increases progressively; the bubble slows to a minimum velocity when the gas/liquid interface becomes totally contaminated. Strictly speaking, the complete rigidification of the bubble boundary occurs before it is completely covered by surfactants, as recently demonstrated by a French team from Louis Pasteur University in Strasbourg.

The behavior of rising, expanding bubbles is more complex than that of those with fixed radii. In the former instance, a bubble's volumetric expansion during its ascent through the supersaturated liquid causes its surface area to grow, offering greater space for surfactant adsorption. Swelling bubbles are consequently subject to opposing effects. If the dilation rate overcomes the speed with which surfactants stiffen the surface, a bubble "cleans" its interface constantly because the ratio of the surface area covered by surfactants to that free from surface-active agents decreases. If this ratio increases, the bubble surface becomes inexorably contaminated by a surfactant monolayer and grows rigid.

By measuring the drag coefficients of expanding champagne and beer bubbles during their journeys up to the surface and then comparing them with data found in the scientific literature of bubble dynamics, we concluded that beer bubbles act very much like rigid spheres. In contrast, bubbles in champagne, sparkling wines and sodas present a more flexible interface during ascent. This is not overly surprising, as beer contains much higher quantities of surfactant macromolecules (on the order of several hundred milligrams per liter) than does champagne (only a few mg/L). Furthermore, because the gas content of beer is lower, growth rates of beer bubbles are lower than those of champagne bubbles. As a result, the cleaning effect caused by a beer bubble's expansion may be too weak to avoid rigidification of its gas/liquid interface. In champagne, sparkling wines and sodas, bubbles grow too rapidly and there are too few surfactant molecules for them to become stiff.

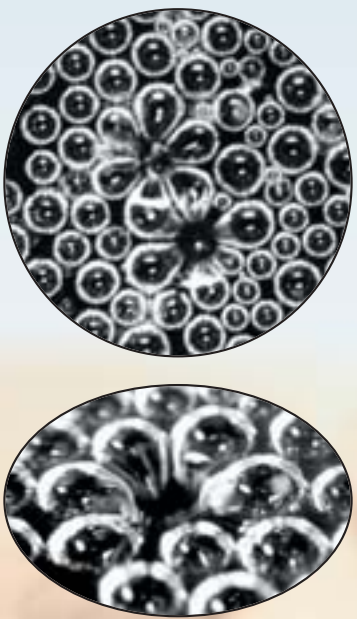
Bubble Collapse

IN THE SEVERAL SECONDS after its birth and release, a bubble travels the few centimeters to the beverage's surface, eventually reaching a diameter of about a millimeter. Like an iceberg, a gas bubble at the top of a drink emerges only slightly from the liquid, with most of its volume remaining below the surface. The emerged part, the bubble cap, is a hemispherical liquid film that gets progressively thinner as a result of drainage around the sides. When a bubble cap reaches a certain critical thickness, it becomes sensitive to vibrations and thermal gradients, whereupon it finally ruptures. Working independently in 1959, two physicists, Geoffrey Ingram Taylor of the University of Cambridge and Fred E. C. Culick of the California Institute of Technology, showed that surface tension causes a hole to appear in the bubble cap and that the hole propagates very quickly. For bubbles of millimeter size, this disintegration takes from 10 to 100 microseconds [*see sequence at right*].

After the bubble cap breaks, a complex hydrodynamic process ensues, causing the collapse of the submerged part of the bubble. For an instant, an open cavity remains in the liquid surface. Then the inrushing sides of the cavity meet and eject a high-speed liquid

DEATH OF A BUBBLE

POP GOES THE BUBBLE: The collapse of a champagne bubble begins nearly as soon as it breaches the surface of the beverage (*sequence at right*). In the second image, the thin liquid film that constitutes the emerged part of the bubble has just ruptured. During this extremely brief event, the shape of the bubble—nearly a millimeter wide—remains intact. The collapse of the cavity gives rise to a high-speed liquid jet that flies above the surface (*third image*). Because of its own velocity, this jet becomes unstable, forming a capillary wave that breaks up the flow into droplets. Hundreds of bubbles are bursting every second after pouring, so the liquid surface is literally spiked with cone-shaped structures, unfortunately too short-lived to be seen with the naked eye.



BUBBLE FLORETS: Because the collapse of surface bubbles is very rapid (less than 100 microseconds), few photographs capture the process. But clusters of champagne bubbles form visually appealing flower-shaped structures (*left*) when they are deformed by the sucking force created by a nearby popped bubble. The bubble caps adjacent to the bubble-free central zones are stretched toward the now empty cavities.

TOSHIO NAKAJIMA *Photonica*; COURTESY OF GÉRARD LIGER-BELAIR (*insets*)

jet above the free surface. Because of its high velocity, this jet becomes unstable, developing a capillary wave (the Rayleigh-Plateau instability) that fragments it into droplets called jet drops. The combined effects of inertia and surface tension give the detaching jet drops a variety of often surprising shapes. Finally they take on a quasi-spherical shape. Because hundreds of bubbles are bursting each second, the beverage surface is spiked with transient conical structures, which are too short-lived to be seen with the unaided eye.

Aroma and Flavor Release

BEYOND AESTHETIC considerations, bubbles bursting at the free surface impart what merchants call "feel" to champagne, sparkling wines, beers and many other beverages. Jet drops are launched at several meters per second up to a few centimeters above the surface, where they come in contact with human sense organs. Nociceptors (pain receptors) in the nose are thus stimulated during tasting, as are touch receptors in the mouth when bubbles burst over the tongue; this bursting also yields a slightly acidic aqueous solution.

In addition to mechanical stimulation, bubbles collapsing at the surface are believed to play a major role in the release of flavors and aromas. The molecular structures of many aromatic compounds in carbonated beverages show surface activity. Hence, bubbles rising and expanding in the bulk liquid serve as traps for aromatic molecules, dragging them along on their way up. These molecules concentrate at the surface. Bursting bubbles are thus thought to spray into the air clouds of tiny droplets containing high concentrations of aromatic molecules, which emphasize the flavors of the beverages. Our future research plans include quantifying this flavor-release effect for each of the numerous aromatic molecules in champagne.

Contrary to what we had first thought, effervescence in carbonated beverages turned out to be a fantastic tool for investigating the physical chemistry of rising, expanding and collapsing bubbles. We hope readers will never look at a glass of champagne in quite the same way. SA

MORE TO EXPLORE

Through a Beer Glass Darkly. Neil Shafer and Richard Zare in *Physics Today,* Vol. 44, pages 48–52; 1991.

Beauty of Another Order: Photography in Science. Ann Thomas. Yale University Press, 1997.

The Secrets of Fizz in Champagne Wines: A Phenomenological Study. Gérard Liger-Belair et al. in *American Journal of Enology and Viticulture,* Vol. 52, pages 88–92; 2001.

Kinetics of Gas Discharging in a Glass of Champagne: The Role of Nucleation Sites. Gérard Liger-Belair et al. in *Langmuir,* Vol. 18, pages 1294–1301; 2002.

Physicochemical Approach to the Effervescence in Champagne Wines. Gérard Liger-Belair in *Annales de Physique,* Vol. 27, No. 4, pages 1–106; 2002.

BALLISTICS

Scratch Match

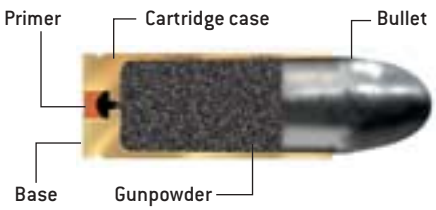
Solving a case like the October sniper shootings around Washington, D.C., hinges heavily on firearms identification. How do examiners match evidence to a gun? If a gun is recovered, a forensic scientist test-fires it to determine the markings it leaves on bullets and cartridge casings. The examiner then compares these under a microscope with the caliber (diameter), rifling pattern (series of grooves), and impressions and striations (microscopic marks left by unique imperfections in a gun's firing pin and barrel) on the bullets and casings found at the crime scene.

If test-fires don't match, or if no gun is found, the examiner will measure the rifling patterns on the recovered ammunition and compare them with the General Rifling Characteristics database to see which gun models might match. The problem in some cases, however, is that "20 to 150 brands of firearms could leave the same rifling pattern," says Scott Doyle, forensic specialist for the Kentucky State Police in Louisville.

If the examiner needs to go to the next level, he can take photographs of the impressions and striations and send them online to the National Integrated Ballistic Information Network (NIBIN), overseen by the Federal Bureau of Alcohol, Tobacco, and Firearms (ATF) and the FBI. Every day local forensic scientists from across the country upload the images of test-fires and crime scene evidence, which NIBIN stores. For each case, it will send back 10 or so of the closest matches, if found, indicating specific guns that might have fired the ammunition. The examiner must then manually contrast the images against his own.

The sniper shootings have raised debate over whether the ATF should turn NIBIN into a national ballistic-fingerprint system. Firearms manufacturers would be required to test-fire every new gun and enter images into NIBIN. The system would help but would not be a panacea. The marks from a gun can change if the barrel rusts over time; criminals can tamper with the barrel, too, says Robert Shem, chief examiner for the Alaska Crime Lab in Anchorage. "You will always need a human expert, comparing evidence under a microscope, to say 'Yes, this bullet came from that gun.'" —*Mark Fischetti*

DIFFERENT GUN BARRELS
have different rifling patterns. Common templates include 4/right (four lands and grooves, twisting to the right) (*far right*), 8/left, and so on. Each pattern cuts a distinct, inverse image in a bullet propelled through the barrel (*second photograph at right*). Each land and groove also leaves unique microscopic striations on the bullet (*third photograph at right*).



PULLING A GUN'S TRIGGER
slams a firing pin into a bullet's primer, which ignites gunpowder. The explosion drives the cartridge case backward into the breech face and out of the gun, leaving telltale scars, and propels the bullet forward. A spiral of raised lands and shallow grooves along the barrel spins the speeding bullet like a thrown football, improving its flight accuracy.

DANIELS & DANIELS; BALLISTICS PHOTOGRAPHS BY SCOTT DOYLE, FROM WWW.FIREARMSID.COM

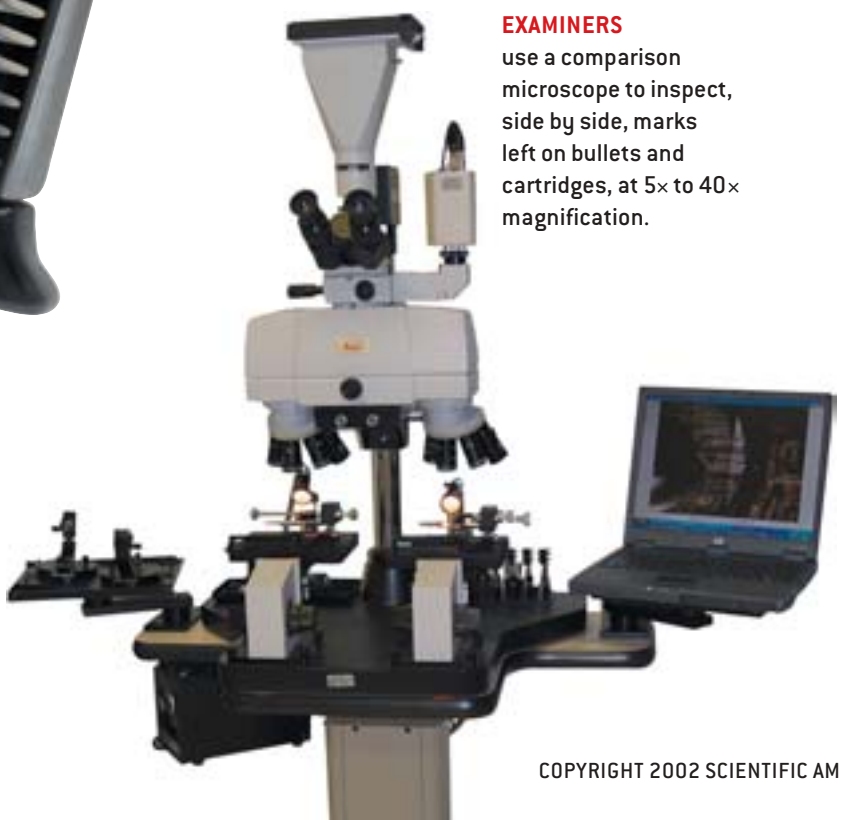
Have a topic for a future column?
Send it to workingknowledge@sciam.com

DID YOU KNOW . . .

➤ X CALIBER: A .22-caliber bullet has a diameter of roughly 0.22 inch. The designation on its cartridge typically includes the caliber, the manufacturer's name and adjectives describing variations; a .32 S&W Long indicates a .32-caliber Smith & Wesson bullet that is longer than the standard model. Casings found during the Beltway sniper shootings were .223; the "3" indicates not a measurement but an elongated .222 Remington cartridge that contains extra gunpowder to propel the bullet faster. More than 25 models of .223 rifles are sold in the U.S.

➤ MATCHMAKERS: The FBI is gradually expanding three nascent national databases to help forensic investigators. The Integrated Automated Fingerprint Identification System houses 40 million sets of fingerprints of prior suspects and convicted criminals. The Combined DNA Index System has 1.2 million DNA profiles. And the National Integrated Ballistic Information Network has half a million images of bullets and cartridges. In each case, local examiners upload data from crime scenes and search for matches from solved and unsolved cases. Civil liberties groups worry that the information could be misused.

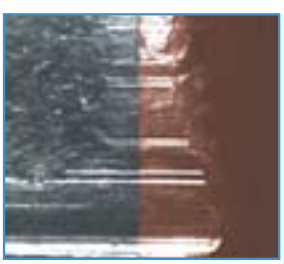
➤ TANKED: Firearms inspectors test-fire guns from one end of a rectangular water tank about three feet wide, three feet high and 10 feet long. The bullet from a revolver travels only about five feet before it drops to the bottom. "The friction is tremendous," says firearms specialist Scott Doyle, "and the projectile rapidly loses its energy."



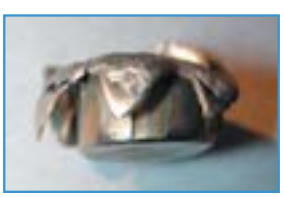
EXAMINERS use a comparison microscope to inspect, side by side, marks left on bullets and cartridges, at 5× to 40× magnification.

BREECH MARKS on the base of a cartridge found at a crime scene (*right*) match marks on a cartridge test-fired from a gun recovered at the scene (*left*).

TWO BULLETS each have six right-turning land and groove impressions, but their pitch differs, indicating that the bullets were fired from different guns.



STRIATIONS ON A BULLET found at a crime scene (*left*), caused by tiny imperfections unique to a gun's lands and grooves, match striations on a bullet test-fired from a recovered gun (*right*).



BADLY TORN FRAGMENT of a bullet still shows enough land and groove impressions and striations to help examiners solve a crime.

Peering into the Earth

A RIM WITH A VIEW IN HAWAII VOLCANOES NATIONAL PARK BY MARGUERITE HOLLOWAY

We lift off from the Hilo airport on Hawaii's Big Island, and our pilot, Robert Blair, swings the helicopter toward a horizon gray with mist, steam and volcanic gas. For the first several minutes, we pass over 2,500 acres of green macadamia plantations while Hilo's famous rain, between 130 and 200 inches a year, splatters the front window. The morning's torrent is clearly subsiding, though—fortunately for us, because most of the drops are coming in where the helicopter's doors used to be. Four passengers and a pilot, we are about to fly over an active volcano—one spewing sulfurous fumes, scarlet lava and searing steam—with just a seatbelt tethering us to a metal contraption whose doors have been removed so we can commune with nature. Against the volcano and its fields of lava, all things seem puny, but at the moment, this helicopter most of all.

The wind is blowing to the west, taking the giant plume of gas and smoke with it, so we hover to the east side of the Pu`u `O`o vent, which is on the eastern rift zone of the Kilauea volcano and has been erupting for 20 years. The ashy brown sides of the vent rise conelike; inside them and through four windows—or skylights, as they are also called—that have opened in the crater floor, we see the incandescent glow of liquid rock. Lava fields stretch around the vent, a vast plain of older brown and newer shiny black cooled rock.

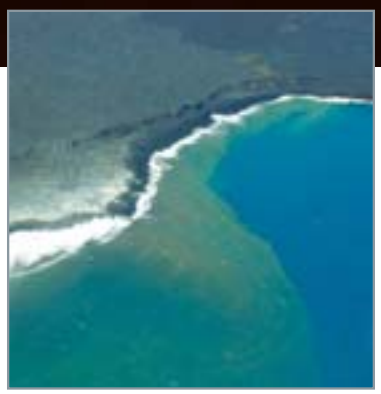
Blair flies tight circles around Pu`u `O`o, taking us lower and lower. Soon we can make out equipment from the U.S. Geological Survey's Hawaiian Volcano Observatory—which is located in Hawaii Volcanoes National Park on the rim of the now quiet Kilauea caldera—and we feel the heat and smell the sulfur. (The vent releases more than 1,500 tons of sulfur dioxide a day, enough to fill 100 blimps, according to the USGS.) After several passes, we veer to the south, following a lava tube, a tunnel of hardened lava through which the roughly 2,100 degree Fahrenheit molten rock travels, down to the Pacific Ocean. There red lava flows into turquoise water, and thick clouds of bright white steam laced with shards of silica and hydrochloric acid rise high above the black sand of the beach. These are the colors of early earth: red, black and blue.

The hour-long helicopter tour is over, and we return to Hilo, flying back over a largely bleak landscape, a battlefield of burned trees and buried houses—although a few structures stand stranded in small patches of rain forest, spared by the flow. Pu`u `O`o destroyed the Royal Gardens community here between 1983 and 1986 and covered part of the park's Chain of Craters Road. From the air it is easy to appreciate the power and reach of the volcano, the primal force shaping this island, creating new land (about 600 acres' worth so far), incinerating everything in its path. From the ground this power feels even more vast and unruly. Down here you can hear the crackle of cooling lava as you walk over it.

Volcanoes National Park is perhaps the only place on earth where visitors arrive in continuous carloads to peer at ashy craters, calderas, cones, plumes of gas, and skeletons of trees and to clamber over sharp rock, desperate to see lava. Despite many posted warnings, people in open-toed sandals and shorts eagerly trot along newly hardened, still hot lava to peek through a sudden opening and catch a glimpse of the red flow. The desire to be as close as possible to this force of nature has cost five people their lives in the past decade because they ignored warnings either about lava hazards or about medical conditions that can be aggravated by sulfuric fumes, says Jim Martin, the park's superintendent. At the same time, however, Hawaiian volcanoes are, as volcanoes go, gentle ones—so if you are going

GAS FROM the Pu`u `O`o vent often affects Big Island residents. The volcanic smog—or vog, as it is called—contains sulfur dioxide, which turns rainwater acidic and causes respiratory health problems.

PHOTOGRAPHS BY U.S. GEOLOGICAL SURVEY, HAWAIIAN VOLCANO OBSERVATORY



LAVA CONES from Pu`u `O`o are among the volcanic iterations found in the park (*bottom image*)—along with jets of lava (*left*), devastated forests (*right*), and clouds of steam and hydrochloric acid (*below*), where molten rock hits the Pacific Ocean.

lava hunting, this is the place to do it. The magma here is more fluid than most and contains less gas, so it is less explosive and gives rise to what are called shield volcanoes because of their sloping profiles.

The volcanoes that make up the Hawaiian Islands are not the result of the subduction of one tectonic plate under another—as are Mount Pinatubo, Mount Fuji, Mount Saint Helens and other volcanoes in the Ring of Fire. Instead magma spurts up from an unmoving hot spot located about 60 miles below the seafloor, in the earth's mantle. The magma cools, building huge mountains that rise out of the sea. As the Pacific plate moves northwest, over millions of years, these volcanic islands pull away from the hot spot and become dormant. New mountains then form above the hot spot. The youngest of these are Kilauea and, 20 miles or so off the coast of the Big Island, a seamount called Lo`ihi, which might break through the waves in 10,000 years.

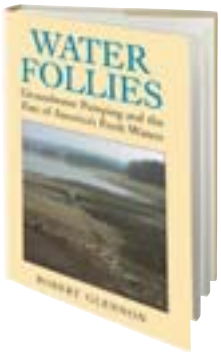
Volcanoes National Park contains volcanoes in various stages, from Pu`u `O`o to Mauna Loa, the largest volcano in the world, and its many cooled lava fields. Mauna Loa last erupted in 1984 but was stirring again this past year, and headlines warned that it might become active again. (Some volcanologists estimate that the hot spot is about 200 miles in diameter, so vents can open or reopen in several places across it.) Visitors can walk through the hollow Thurston Lava Tube, and every Wednesday a park volunteer leads a limited tour through the less accessible Pu`u lava tube (call 808-985-6000 well in advance to reserve a spot). The park is also a good place to see the rain forests and grasslands that eventually colonize the mineral-rich lava fields.

The Volcanoes National Park visitor center is open from 7:45 A.M. until 5 P.M., and there are regular video shows and ranger talks that explain the geology of the region and provide information about native plants and animals (www.nps.gov/havo). The Jagger Museum, located right next to the USGS observatory, is open from 8:30 A.M. to 5 P.M. The exhibits describe the various techniques volcanologists use to study their subjects, and there are seven operational seismographs on display (http://hvo.wr.usgs.gov).

Several companies offer helicopter tours, including flights with the doors on, with the doors off, and with detours to nearby waterfalls. Blair works for Tropical Helicopters: 808-961-6810 (www.tropicalhelicopters.com). Other companies found at the Hilo airport are Paradise Helicopters (808-969-7392), Sunshine Helicopters (808-882-1223) and Blue Hawaiian Helicopters (808-961-5600). Fifty-minute flights run from about $100 to $175 a person. SA

Out of Sight, Out of Mind

AN ONCOMING CRISIS OVER MISUSE OF A HIDDEN RESOURCE—AMERICA'S AQUIFERS **BY DOUGLAS JEHL**



WATER FOLLIES: GROUNDWATER PUMPING AND THE FATE OF AMERICA'S FRESH WATERS
by Robert Glennon
Island Press, Washington, D.C., 2002 ($25)

In the high plains of Texas the farmers who grow cotton, alfalfa and other crops are entitled by law to as much underground water as they can reasonably use. No matter that this water comes from the Ogallala Aquifer, that vast underground reservoir whose levels have dropped precipitously since 1940. No matter that the overpumping threatens eventually to put thousands of farmers across seven states out of business. The illusion, codified in the law not just in Texas but in much of the U.S., is that groundwater is somehow boundless, or in a category apart from lakes, rivers and streams, and ought not be regulated, even for the common good.

Now comes Robert Glennon to puncture this illusion, in a book as rich in detail as it is devastating in its argument. Its focus on groundwater brings overdue attention to a category that accounts for nearly a quarter of American freshwater use. Its title, *Water Follies,* sets the tone for tales that can be tragicomic; this is a book about water being squandered, so it is also, as the author puts it, a book about "human foibles, including greed, stubbornness, and especially, the unlimited human capacity to ignore reality."

Take, for example, his story of the fast-food french fry. It used to be that potatoes were grown on unirrigated land, he writes, but Americans' love of processed foods changed that. Uneven moisture leads to small, knobby, misshapen potatoes, so most American growers, even in places such as Minnesota, routinely irrigate their lands, to produce products acceptable to the industry and customers like McDonald's.

But in Minnesota the groundwater that farmers pump for potatoes turned out to be the same water that helps to sustain the Straight River, a major trout fishery. Even modest pumping for potatoes, a federal study eventually concluded, had the potential to reduce the river's flow by one third during irrigation season, with adverse impact on the brown trout. For now, the trout are not in danger, but that could change if Minnesota were to approve applications from farmers still eager to see potato planting and irrigation widen.

"One long-term answer, of course," Glennon notes, with characteristic wryness, "is for us, as American consumers, to accept french fries that have slightly different colors, or minor discolorations, or even ones that are not long enough to stick out from a supersize carton."

Farmers are not the only ones who get a hard time for their shortsightedness. Bottled-water purveyors, particularly Perrier, are tarred for their pursuit, in places such as Wisconsin, of cool, underground (and highly profitable) springwater in quantities so vast as to prove devastating to the ecology of nearby rivers. The gold-mining industry is called to account for "dewatering" operations in, for example, Nevada, where it makes way for its deep operations by pumping away groundwater at a stunning rate. And planners in Tampa, Fla., and San Antonio, Tex., come under fire for their cavalier reliance on perishable underground sources such as Texas's Edwards Aquifer to fuel development they are finding difficult to sustain.

The cumulative picture painted by the

PHELAN M. EBENHACK *AP Photo*

SINKHOLE, 60 feet deep and 50 feet wide, created by groundwater pumping in Orlando, Fla., June 2002.

author is a grim one. Already four states—Florida, Nebraska, Kansas and Mississippi—use more groundwater than surface water, and more and more are looking underground to support growing populations. Becoming equally apparent are the consequences in dry rivers, land subsidence, and aquifers drawn down far faster than they can ever be recharged. "The country cannot sustain even the current levels of groundwater use," Glennon writes, "never mind the projected increases in groundwater consumption over the next two decades."

Why is it that groundwater has become subject to such abuse? One reason, of course, is that buried below the surface, it is hidden from the kind of relentless monitoring that in recent decades has helped clean up rivers such as the Erie and the Hudson.

But Glennon, a professor of law at the University of Arizona, finds buried in the law some further reasons for the neglect. Even now, he says, most American laws affecting groundwater do not recognize any connection between underground and surface waters, despite abundant evidence of such links. They remain rooted in 19th-century ideas that underground flows were something so mysterious that they could not be understood, an assumption that has been translated into lax or nonexistent regulation.

In most parts of the U.S., the author points out, surface water is subject to doctrines of riparian law or prior appropriation, with water rights carefully parceled out to various claimants. Groundwater, in contrast, is most often subject to the rule of capture, which, as Glennon observes, essentially means that "the biggest pump wins," notwithstanding the impact on surface water or the aquifer itself.

To Glennon, the plight of the country's groundwater has come increasingly to represent what biologist Garrett Hardin called "the tragedy of the commons," a direct result of allowing citizens unlimited use of a common area. Among his recommendations for the future is an immediate halt to unregulated groundwater pumping. To some ears, especially those of high-plains Texas farmers, that is certain to sound like an unconscionable assault on property rights. But *Water Follies* makes the case that groundwater is something that we all should regard as very public indeed. SA

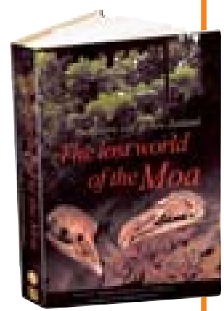
Douglas Jehl, a reporter for the New York Times, *writes frequently on water issues for that publication.*

REVIEWS

THE EDITORS **RECOMMEND**

THE LOST WORLD OF THE MOA: PREHISTORIC LIFE OF NEW ZEALAND

by Trevor H. Worthy and Richard N. Holdaway. Principal photography by Rod Morris. Indiana University Press, Bloomington, Ind., 2002 ($75)



This book is about much moa than moas. It places these extinct giant birds in the world of their curious contemporary fauna—the kiwis and tuataras, the walking worms and giant land mollusks that lived with the moas in "the land that time forgot." Now an archipelago, New Zealand was once part of the large southern continent of Gondwana. "Before the end of the age of dinosaurs, a fragment of Gondwana broke free, carrying with it a ready-made fauna," the authors write. Isolated, this fauna was protected from overland invasions by other animals and followed its own unique evolutionary path, along which birds—and even bats—became flightless. Some 80 million years later primitive forms still dominated. Sadly, most of these species were ill prepared to face the new mammalian predators brought by humans. Worthy, a research associate at the Museum of New Zealand Te Papa Tongarewa, and Holdaway, an independent researcher who works for the New Zealand government and the University of Canterbury, have given us the definitive book on this world and its demise—more than 600 pages of scholarly, copiously illustrated, lucidly presented information.

THE BEST AMERICAN SCIENCE AND NATURE WRITING 2002

Edited by Natalie Angier. Houghton Mifflin, Boston, 2002 ($27.50)

THE BEST AMERICAN SCIENCE WRITING 2002

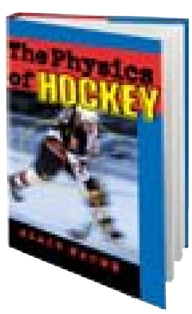
Edited by Matt Ridley. Ecco, New York, 2002 ($27.50)

The two editors, both science writers, set out with the same objective: culling good science writing from U.S. magazines and newspapers published in 2002. Intriguingly, their collections have only one article in common—Sarah Blaffer Hrdy's "Mothers and Others," from *Natural History*. Good science writing is evidently plentiful. The 47 articles reproduced in the two books cover a broad range of subjects and make for edifying, even entertaining, reading.

THE PHYSICS OF HOCKEY

by Alain Haché. Johns Hopkins University Press, Baltimore, 2002 ($24.95)

Haché brings to this informative study the perspective of a physicist (he is assistant professor of physics at the University of Moncton in New Brunswick, Canada) and amateur hockey player (goalie). He stints on neither the physics, which he presents clearly, nor the hockey, making the reader feel like going to a game. Hockey, he says, perhaps involves more physics than any other sport. "Because it is played on ice, we need to take into account elements of thermodynamics and molecular physics. Skating makes use of a great deal of mechanics, as does shooting. Puck trajectories are influenced by air drag and ice friction, which involve fluid dynamics. And because hockey is a contact sport, the physics of collisions is also part of the game." After chapters on the ice and aspects of play, Haché considers the game as a whole and offers a betting tip: "Bet on the team that is in the middle of a losing streak (or against the team that seems to be on a roll)."



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

Protein Chime

BY DENNIS E. SHASHA

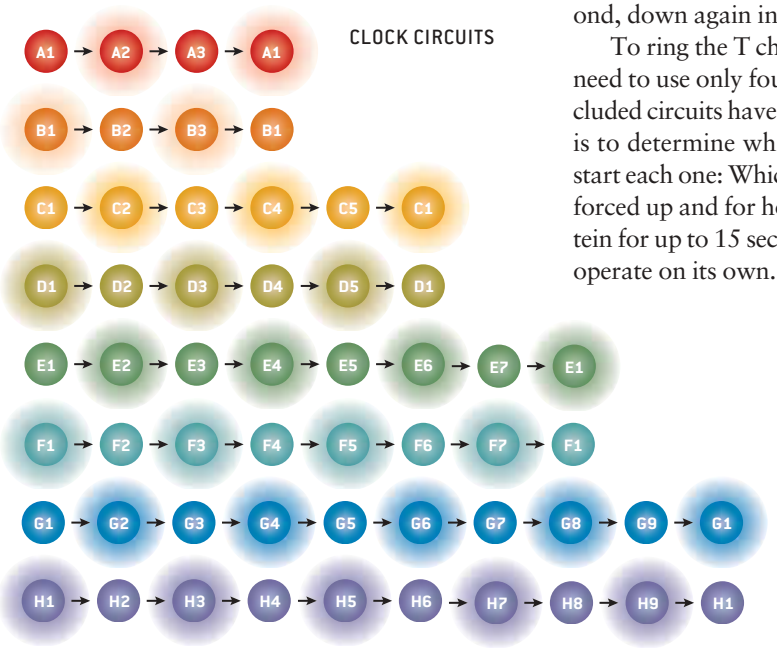
In the 1950s Jacques Monod and François Jacob of the Pasteur Institute in Paris showed that certain regulator proteins in *Escherichia coli* bacteria can repress the production of other proteins. Let's use the notation X → Y to mean that protein X represses protein Y. If X goes up (that is, the protein appears), then Y will go down (the protein disappears) after a short time (say, one second). If X goes down and no other repressor for Y is up, then Y will go up one second after X goes down.

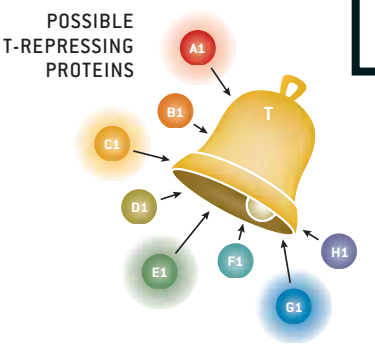
Now consider three proteins: A, B and C, such that A → B → C → A. If A goes up, B goes down one second later. After another second, C goes up, and after one more second A goes back down. Then the pattern continues: B up, C down, A up, B down, and so on. We call this a circuit of proteins: A, B and C periodically appear and disappear, acting like a biochemical clock. Such a clock was actually constructed a few years ago by Michael Elowitz, then a graduate student at Princeton University, and his adviser Stanislas Leibler.

The goal of this puzzle is to create a "protein chime" that will ring every 70 seconds. There are eight circuits, labeled A through H, each containing three, five, seven or nine proteins. No protein in any circuit has an effect on proteins in different circuits. But one protein in each circuit represses T, the special chiming protein that rings when it goes up. If any of these eight T-repressing proteins (labeled A1 through H1) goes up, T will go down one second later. And T will not go back up until one second after all eight proteins are down.

To start any of the circuits in this clock, you must force the production of one of the proteins in the circuit. For example, to start the C circuit, you can keep C4 forced up for five seconds. One second after the forcing begins, C5 goes down. In the succeeding seconds, C1 goes up, C2 goes down and C3 goes up. But C4 does not go down at the end of five seconds, because it is still being forced up. C3 does not push C4 down until one second after the forcing stops (that is, six seconds after the start). Then the cycle resumes without interruption: in the seventh second, C5 goes up, and in the eighth second C1 goes down. C1 goes up again in the 13th second, down again in the 18th, and so on.

To ring the T chime every 70 seconds, you will need to use only four of the eight circuits. (The excluded circuits have no effect on T.) Your challenge is to determine which circuits to use and how to start each one: Which protein in the circuit must be forced up and for how long? You may force a protein for up to 15 seconds; after that, the clock must operate on its own.





SARA CHEN

Answer to Last Month's Puzzle

If the advisers are labeled A_0 through A_8, the director should give tidbits to 25 foursomes determined by this protocol:

$A_7, A_x, A_{x+1}, A_{x+3}$
$A_8, A_x, A_{x+2}, A_{x+3}$
$A_x, A_{x+1}, A_{x+2}, A_{x+4}$
A_7, A_8, A_0, A_1
A_7, A_8, A_2, A_3
A_7, A_8, A_4, A_5
A_7, A_8, A_6, A_0

where *x* ranges from 0 to 6 and the + operation is modulo 7, which means 6 + 2 = 1. Once a tidbit produces a leak, the director determines which three-member subsets of the foursome cannot yet be exonerated. If more than one threesome remains suspect, he gives tidbits to all but one threesome. This strategy will identify the leakers directly or by elimination.

Web Solution

For a full explanation of last month's puzzle and a peek at the answer to this month's problem, visit www.sciam.com

Good Fellows

HOLMES IS WHERE THE HEART IS, BUT HE DESERVES SOME COMPANY **BY STEVE MIRSKY**

Truth is stranger than fiction, it is often said. Well, this is the truth—in October 2002 England's Royal Society of Chemistry granted an honorary fellowship to a fictional character who is no stranger: Sherlock Holmes. A spokesperson for the society was quoted as saying that the recognition was for Holmes's "love of chemistry, and the way that he wielded such knowledge for the public good, employing it dispassionately and analytically."

If any fabricated person deserves membership in a real science organization, Sherlock's surely a (gum)shoo-in. But don't other pretend people and assorted chimerical characters also merit fellowships in various real science societies? In the Holmesian spirit, here are a few nominations.

American Ornithologists' Union: The Road Runner, for research on the predator-prey relationship, specifically within the context of avian flightlessness.

American Institute of Physics: Wile E. Coyote, for his literally groundbreaking studies, performed in collaboration with the Road Runner, on falling objects and rapid deceleration.

International Society for Prosthetics and Orthotics: Captain Ahab, for overcoming adversity and maintaining a position of authority despite being differently abled.

American Society of Transplant Surgeons: Dr. Victor Frankenstein, for his innovative research on the art and science of multiple organ, limb, bone—well, pretty much everything—transplantation.

Academy of Criminal Justice Sciences: Jean Valjean, for his long-term study of recidivism and techniques involved in successful rehabilitation of convicted felons.

American Society for Adolescent Psychiatry: Holden Caulfield, for his studies of the American postwar adolescent experience.

American Board of Sport Psychology: Mighty Casey, for striking out and thus illustrating the pitfalls of unfounded high self-esteem, while at the same time bringing no joy to Mudville, thereby illustrating the dangers of investment in external sources of gratification.

Society for Endocrinology: Paul Bunyan, for actualizing his personal experience, despite an obvious and potentially debilitating endocrine disorder. (Note: The Sierra Club may protest this citation.)

International Society of Entomology: Gregor Samsa, for his serendipitous study of the social life of the common cockroach.

The International Association for Research in Economic Psychology and the American Society of Agronomy, jointly: Willy Loman, for his enduring efforts to educate others about the importance of being well liked for success in sales and marketing and for his forceful hypothesis that a man is not a piece of fruit.

American Association for Marriage and Family Therapy: Anna Karenina, for identifying individualized unhappiness within the family unit.

American Association of Amateur Astronomers: Popeye, for the countless new stars and other celestial objects he discovered—thanks to his spinach-enhanced punches—orbiting Bluto's head.

American Meteorological Society and National Geographic Society, jointly: Dorothy Gale, for employing tornadoes in the discovery of new lands and peoples.

American Heart Association: The Scarecrow, for his tenacious efforts to obtain a heart despite his apparent capacity to get along fine without one.

American Society of Metallurgists: Rumplestiltskin, for his imaginative insights in the field of transmutation, specifically involving the conversion of common base material into precious metal.

American Podiatric Medical Association: Achilles, for his elucidation of the vital importance of foot health care.

Linguistic Society of America: Presidents Josiah Bartlet (*The West Wing*), Tom Beck (*Deep Impact*), Andrew Shepherd (*The American President*) and James Marshall (*Air Force One*), for their ability to construct complete, grammatically coherent sentences.

And, finally:
Royal Agricultural Society of England: Dr. John Watson, for being the quintessential second banana (and thereby offering evidence disputing the theoretical agronomy of Willy Loman). SA

PHOTOGRAPH BY FRANK VERONSKY

ASK THE EXPERTS

How do Internet search engines work?

—A. Dharia, Houston

Javed Mostafa, Victor H. Yngve Associate Professor of Information Research Science and director of the Laboratory of Applied Informatics, Indiana University, offers this answer:

Publicly available Web services—such as Google, InfoSeek, Northernlight and AltaVista—employ various techniques to speed up and refine their searches. The three most common methods are known as preprocessing the data, "smart" representation and prioritizing the results.

One way to save search time is to match the Web user's query against an index file of preprocessed data stored in one location, instead of sorting through millions of Web sites. To update the preprocessed data, software called a crawler is sent periodically by the database to collect Web pages. A different program parses the retrieved pages to extract search words. These words are stored, along with the links to the corresponding pages, in the index file. New user queries are then matched against this index file.

Smart representation refers to selecting an index structure that minimizes search time. Data are far more efficiently organized in a "tree" than in a sequential list. In an index tree, the search starts at the "top," or root node. For search terms that start with letters that are earlier in the alphabet than the node word, the search proceeds down a "left" branch; for later letters, "right." At each subsequent node there are further branches to try, until the search term is either found or established as not being on the tree.

The URLs, or links, produced as a result of such searches are usually numerous. But because of ambiguities of language (consider "window blind" versus "blind ambition"), the resulting links would generally not be equally relevant. To glean the most pertinent records, the search algorithm applies ranking strategies. A common method, known as term-frequency-inverse document-frequency, determines relative weights for words to signify their importance in individual documents; the weights are based on the distribution of the words and the frequency with which they occur. Words that occur very often (such as "or," "to" and "with") and that appear in many documents have substantially less weight than do words that appear in relatively few documents and are semantically more relevant.

Link analysis is another weighting strategy. This technique considers the nature of each page—namely, if it is an "authority" (a number of other pages point to it) or a "hub" (it points to a number of other pages). The highly successful Google search engine uses this method to polish searches.

What is quicksand?

—S. Yamasaki, Brussels, Belgium

Darrel G. F. Long, a sedimentologist in the department of earth sciences, Laurentian University in Sudbury, Ontario, explains:

Quicksand is a mixture of sand and water or of sand and air; it looks solid but becomes unstable when it is disturbed by any additional stress. Grains frequently are elongated rather than spherical, so loose packing can produce a configuration in which the spaces between the granules, or voids, filled with air or water make up 30 to 70 percent of the total volume. This arrangement is similar to a house of cards, in which the space between the cards is significantly greater than the space occupied by the cards. In quicksand, the sand collapses, or becomes "quick," when force from loading, vibration or the upward migration of water overcomes the friction holding the particles in place. In normal sand, in contrast, tight packing forms a rigid mass, with voids making up only about 25 to 30 percent of the volume.

Most quicksand occurs in settings where there are natural springs, such as at the base of alluvial fans (cone-shaped bodies of sand and gravel formed by rivers flowing from mountains), along riverbanks or on beaches at low tide. Quicksand does appear in deserts, on the loosely packed, downwind sides of dunes, but this is rare. And the amount of sinking is limited to a few centimeters, because once the air in the voids is expelled, the grains nestle too close together to allow further compaction. SA

For a complete text of these and other answers from scientists in diverse fields, visit www.sciam.com/askexpert

MATT COLLINS