100 YEARS OF QUANTUM MYSTERIES

SCIENTIFIC AMERICAN

FEBRUARY 2001 $4.95 WWW.SCIAM.COM

Are We Almost TAPPED OUT?

Safeguarding every drop of clean water

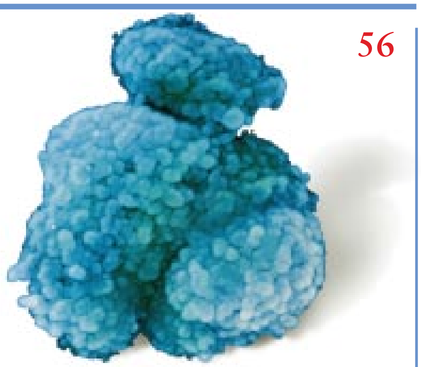
- Light-Emitting Diodes
- The Science of Persuasion
- Evolution of Sex Chromosomes

SCIENTIFIC AMERICAN

February 2001 Volume 284 www.sciam.com Number 2

Contents

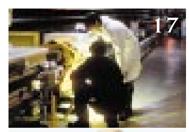
COVER STORY

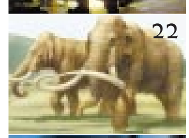


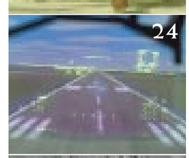
Contents

SCIENTIFIC AMERICAN

February 2001 | Volume 284 | www.sciam.com | Number 2







About the Cover
Illustration by Slim Films.

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

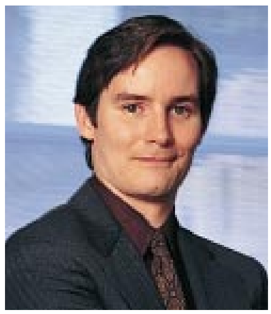
Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97, Canada $49, International $55. Postmaster: 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.

EDITOR_JOHN RENNIE

When Physics Goes Pop

When quantum theory first glimmered in 1900, the 45 United States of America had 144 miles of hard-surfaced roads, one telephone for every 13 homes and one bathtub for every seven. During that year, Walter Reed demonstrated that mosquitoes carried yellow fever, and this country suffered its first epidemic of bubonic plague; the average age at death in the U.S. was 47. Browning pistols and Brownie box cameras were introduced. China was torn by the Boxer Rebellion, and South Africa fought its Boer War. Count Ferdinand von Zeppelin built his first dirigible.

The world has changed and accepted much since then, obviously. But has it learned to embrace quantum theory? The words can still induce panic attacks among the physics-challenged. Few nonscientists would even claim to understand what quantum mechanics is. Nevertheless, it has slowly gained at least some kind of broad cultural currency.



ERICA LANSNER

Quantum theory's most successful foray has been through technology, of course. People don't need to know what quanta are to enjoy the benefits of their application. As Max Tegmark and John Archibald Wheeler celebrate in their article "100 Years of Quantum Mysteries" (see page 68), 30 percent of this country's gross national product derives from instruments that operate on quantum principles: the transistor, the laser, MRI scanners, superconducting magnets and much more.

"Uncertainty principle" is a phrase tailor-perfect for our anxious times.

Ideas plucked from quantum physics have also developed a life of their own in the popular imagination, most often as metaphors. They cling to some pith of their original meaning, although distortions can settle in, too. Consider the expression "It took a quantum leap forward." The speaker almost always means that something has advanced by a large, sudden increment. How did extremely tiny leaps transmogrify into huge ones?

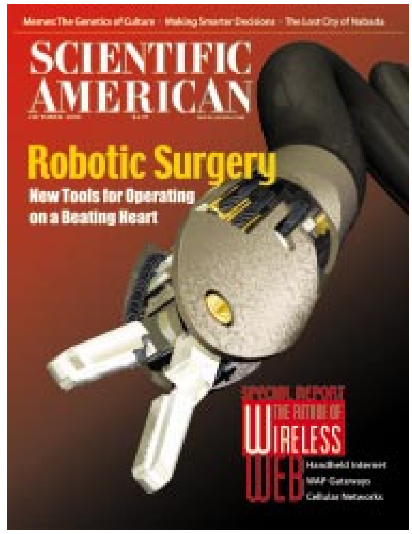
"Uncertainty principle" is a phrase tailor-perfect for our anxious times. What a relief: physics gives us an excuse for never being too sure about anything. Michael Frayn's play *Copenhagen* may be one of the more sublime results of that inspiration, in its exploration of murky human motives through an argument between Niels Bohr and Werner Heisenberg. The concept of parallel worlds, which science fiction embraced for decades, has attained even more widespread popularity for framing "what if" fantasies, as in the movies *Sliding Doors* and *Run, Lola, Run*. (But does that make *Rashomon* about relativity?)

The worst results of quantum physics infiltrating pop culture must be the shelves of cheesy physics-cum-philosophy tracts that bridge the science and New Age sections of bookstores. Wishfully citing quantum jargon, these authors find a basis for telepathy and other paranormal phenomena. Never mind; some misunderstanding is par for the course. In another 100 years, maybe even children will understand quantum theory. After all, it's not going away.

editors@sciam.com

Scientific American, Inc.
415 Madison Avenue
New York, NY 10017-1111
PHONE: (212) 754-0550
FAX: (212) 755-1976
WEB SITE: www.sciam.com

EDITORS@SCIAM.COM



The Joy of Memetics

Might imitative ability be a by-product of the evolution of the peculiar human trait for problem solving ["The Power of Memes," by Susan Blackmore]? Problem solving seems to be closer to the core of what was needed for early hominid hunters to survive. Hunters benefit from language and auditory skills as well as depth perception and the ability to abstract. It could be that the capacities for music, art and philosophy are just secondary frills of the brain complexity needed for higher problem solving.

DOUG BERGER
Department of Psychiatry
Albert Einstein College of Medicine

Imitation is largely useless without creativity. Among genes, "creative" events result from simple, mindless mechanisms of mutation and gene duplication, drift and recombination, followed by fixation through copying. Human creativity is far more subtle and resistant to reductionism.

Contrary to memetics, human evolutionary advantage and sexual attractiveness should go not to the best imitators but rather to individuals who can best create, understand or selectively employ the most useful memes in crucial situations. As noted by Lee Alan Dugatkin ["Animals Imitate, Too"], one of Blackmore's multiple definitions of "imitation" includes both a selection and a copying step; it would seem much better to keep these very different concepts explicitly separate, as they are when describing Darwinism.

PAUL E. DRIEDGER
Woburn, Mass.

Blackmore replies:

Driedger implies that imitation and creativity are opposite processes. But this commonsense view is turned inside out by memetic thinking, which treats human creativity as an evolutionary process that depends on human imitation for its copying mechanism. This is why imitation—apparently paradoxically—turns out to be the source of our amazing creativity. I agree that we would do well to study the copying step and the selection step separately, for both are complex and poorly understood.

Berger reiterates the usual biologically based argument. The joy of memetics is that it provides a completely different view—that the familiar evolutionary process working on a new replicator explains how we acquired all these other skills.

Reports of Humanity's Death ...

Listed as one of the "Paul Ehrlich: Fast Facts" ["Six Billion and Counting," by Julie Lewis, Profile, News and Analysis] is that he "turned down medical school." Thank goodness! Had he become a medical doctor, humanity (or his patients, at least) might have actually faced the premature demise that he has been predicting for decades. Ehrlich has been famously wrong throughout his entire career yet remains virtually unscathed. Exactly how many times must the evidence contradict the hypothesis before the idea is discredited? If the world survives half as long as Ehrlich's arrogance, death is a long way off.

EDWARD SIEBER
Alexandria, Va.

THE_MAIL

"WHAT A LONG ROAD humankind has traveled over the past 4,500 years," writes Leigh Ramsay of San Diego, commenting on the October 2000 issue, "and yet how little has changed. In 'Nabada: The Buried City,' Joachim Bretschneider notes that clay tablets provided 'a meticulous record' of the daily activities of Nabadian society. In 'The Internet in Your Hands,' Fiona Harvey observes that Nokia's conceptual phone could 'perform a plenitude of tasks' to support the daily activities of our world society. One wonders if 4,500 years from now Bretschneider's long-distant descendant will find a buried cache of plastic tablets in what may once have been a landfill and remark that 'the tablets are curious in one aspect: the language is English, but the script is Nokian.'"

Starting above, a selection of letters on other October articles.

PATRICIA J. WYNNE (*top*); COURTESY OF LERNOUT & HAUSPIE AND NOKIA (*bottom*)

Options for Coronary Surgery

The implication of Cornelius Borst's "Operating on a Beating Heart" is that the off-pump CABG is a much better alternative than the heart-lung machine. This article could frighten the hundreds of thousands of patients who will have very successful cardiopulmonary bypass operations this year. Although complications exist, the incidence has gone way down as the technology has improved dramatically in the past decade. The off-pump CABG is a good operation for certain individuals; however, it has not been demonstrated to be safer or less expensive in any scientific study to date.

LAWRENCE H. COHN
Chief, Division of Cardiac Surgery
Brigham and Women's Hospital
Harvard Medical School

Borst replies:

The majority of the most recently published nonrandomized studies in selected patients suggest that beating-heart surgery is associated with comparable technical revascularization success, fewer complications, shorter hospital stay, earlier return to normal activities and lower overall cost.

At this stage in the transition of conventional to beating-heart coronary surgery, the choice of treatment will depend on the balance between the medical history and condition of the individual patient and the available surgical and anesthesiologic expertise to perform this more demanding surgical technique.

Violence, Drugs, Guns (and Switzerland)

"The Roots of Homicide," by Rodger Doyle [By the Numbers, News and Analysis], ignored an obvious and important cause of violence. Prohibition of alcohol in the 1920s and the current "war on drugs," which began in the 1960s, both led to gang violence and drive-by shootings. This is the real root of the current homicide rates in America.

GERARD MURPHY
Honolulu, Hawaii

In the largest sample ever analyzed on the topic (36 Western nations, including the U.S.), there was no significant correlation between gun ownership rates and homicide rates. More generally, the best available evidence indicates that gun-ownership levels have no net effect on violence rates and that the association sometimes observed between the two is related to the effect of the latter on the former (for example, higher homicide rates motivate people to acquire guns for self-protection), rather than the reverse.

GARY KLECK
School of Criminology
and Criminal Justice
Florida State University

Doyle replies:

I confined my analysis to 11 countries on the basis that it is desirable to compare countries that are alike in terms of general social characteristics. Kleck finds no correlation using 36 countries because he is increasing the number of confounding variables. My key point is that the combination of easy access to guns and an extraordinary readiness to use them helps make the U.S. homicide rate so high.

More than a dozen readers wondered why I didn't mention Switzerland, which maintains an armed militia and a low homicide rate. According to criminologist Martin Killias of the University of Lausanne, the everyday availability of these weapons has led to the high suicide rates there, but firearm use for other purposes is limited because ammunition is provided in a sealed box that may be opened only in a wartime emergency.

Letters to the editors should be sent by e-mail to editors@sciam.com or by post to Scientific American, 415 Madison Ave., New York, NY 10017. Letters may be edited for length and clarity.

ERRATUM

In "The Third-Generation Gap" [Special Industry Report: The Wireless Web], Leander Kahney stated that Sprint's CDMA network will migrate to W-CDMA. In fact, Sprint is moving to 1XRTT, a cdma2000 technology.

SCIENTIFIC AMERICAN

Denise Anderman
PUBLISHER
danderman@sciam.com

Gail Delott
ASSOCIATE PUBLISHER
gdelott@sciam.com

NEW YORK ADVERTISING OFFICES
415 Madison Avenue
New York, NY 10017
212-451-8893 fax 212-754-1138

David Tirpack
Sales Development Manager
dtirpack@sciam.com

Wanda R. Knox
wknox@sciam.com

Hunter Millington
hmillington@sciam.com

Darren Palmieri
dpalmieri@sciam.com

Stan Schmidt
sschmidt@sciam.com

DETROIT
Edward A. Bartley
Midwest Manager
248-353-4411 fax 248-353-4360
ebartley@sciam.com

LOS ANGELES
Stephen Dudley
Los Angeles Manager
310-234-2699 fax 310-234-2670
sdudley@sciam.com

SAN FRANCISCO
Debra Silver
San Francisco Manager
415-403-9030 fax 415-403-9033
dsilver@sciam.com

CHICAGO
ROCHA & ZOELLER MEDIA SALES
312-782-8855 fax 312-782-8857
mrrocha@aol.com
kzoeller1@aol.com

DALLAS
THE GRIFFITH GROUP
972-931-9001 fax 972-931-9074
lowcpm@onramp.net

CANADA
FENN COMPANY, INC.
905-833-6200 fax 905-833-2116
dfenn@canadads.com

U.K.
Anthony Turner
Simon Taylor
THE POWERS TURNER GROUP
+44 207 592-8323 fax +44 207 592-8324
aturner@publicitas.com

FRANCE AND SWITZERLAND
Patricia Goupy
+33-1-4143-8300
pgoupy@compuserve.com

GERMANY
Rupert Tonn
John Orchard
PUBLICITAS GERMANY GMBH
+49 69 71 91 49 0 fax +49 69 71 91 49 30
rtonn@publicitas.com
jorchard@publicitas.com

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

OTHER EDITIONS OF SCIENTIFIC AMERICAN

SPEKTRUM DER WISSENSCHAFT
Verlagsgesellschaft mbH
Vangerowstrasse 20
69115 Heidelberg, GERMANY
tel: +49-6221-50460
redaktion@spektrum.com

POUR LA SCIENCE
Éditions Belin
8, rue Férou
75006 Paris, FRANCE
tel: +33-1-55-42-84-00

LE SCIENZE
Piazza della Repubblica, 8
20121 Milano, ITALY
tel: +39-2-29001753
redazione@lescienze.it

INVESTIGACION Y CIENCIA
Prensa Científica, S.A.
Muntaner, 339 pral. 1.ª
08021 Barcelona, SPAIN
tel: +34-93-4143344
precisa@abaforum.es

MAJALLAT AL-OLOOM
Kuwait Foundation for
the Advancement of Sciences
P.O. Box 20856
Safat 13069, KUWAIT
tel: +965-2428186

SWIAT NAUKI
Proszynski i Ska S.A.
ul. Garazowa 7
02-651 Warszawa, POLAND
tel: +48-022-607-76-40
swiatnauki@proszynski.com.pl

NIKKEI SCIENCE, INC.
1-9-5 Otemachi, Chiyoda-ku
Tokyo 100-8066, JAPAN
tel: +813-5255-2821

SVIT NAUKY
Lviv State Medical University
69 Pekarska Street
290010, Lviv, UKRAINE
tel: +380-322-755856
zavadka@meduniv.lviv.ua

SCIENTIFIC AMERICAN HELLAS SA
35–37 Sp. Mercouri Street
Gr 116 34 Athens, GREECE
tel: +301-72-94-354
sciam@otenet.gr

KE XUE
Institute of Scientific and
Technical Information of China
P.O. Box 2104
Chongqing, Sichuan
PEOPLE'S REPUBLIC OF CHINA
tel: +86-236-3863170

50, 100 & 150 Years Ago

The Common Cold, An Early Dial Telephone, 1901

FEBRUARY 1951

HEAT AND CHEMISTRY—"John A. Swartout of the Oak Ridge National Laboratory, in a comprehensive paper on the chemistry connected with nuclear reactors, revealed that this research had opened a whole new field of 'high-temperature chemistry.' Most chemical research in the past, he pointed out, has been conducted at room temperatures, and relatively little study has been given to chemical reactions above 100 degrees Centigrade. In the program looking toward the development of reactors for power, chemists must study how chemicals react at temperatures far above this level."

COMMON COLD—"An attack on the cold problem has been carried out since 1946 in the Common Cold Research Unit of the Medical Research Council at Salisbury, England. Of 2,000 volunteers, those who received the harmless control inoculations remained satisfactorily free from colds during their 10-day stay. Of those who received the active secretions taken from people with colds, some 50 per cent caught colds. An interesting point is that many of those who were inoculated with active materials seemed to be starting a cold on the second day or third day but next day had lost all their symptoms: the cold had aborted. It is easy to see why remedies purporting to cure the common cold so often gain a wholly unmerited reputation."

FEBRUARY 1901

TYPHOID AND WAR—"Typhoid fever in every war has claimed more victims not only than wounds caused by weapons of destruction, but even more than any other disease. The recent report of the commission appointed to inquire into the various causes of death among our soldiers during the Spanish-American war says that enteric fever was responsible for the great majority of fatalities. What is needed is an effectual method of purifying drinking-water. According to the *Medical Magazine,* filtration 'is too tedious for practical use with great bodies of troops. Boiling is also inconvenient and the cooling period entails waiting. Formalin leaves an objectionable taste.' The German government has a preference for bromine, but its method of employment would seem to be too elaborate for use with soldiers on the march."

SCIENTIFIC AMERICAN

TELEPHONE subscriber's new transmitter/receiver, 1901

TESLA'S TELEGRAPHY—"Long distance wireless telegraphy, if we may believe the current story, is about to take an enormous stride, for we are shortly to be in possession of a means of wireless telegraphic communication across the Atlantic, by which we can send messages at considerable greater speed than is possible by the present cable. The feat is to be accomplished by the Nicola Tesla 'oscillator.' We are, all of us, fairly well familiar with the Marconi system in which Hertzian waves are transmitted through the ether. Mr. Tesla, however, manipulates his recently discovered 'stationary electrical waves in the earth' by setting up 'vibratory currents which can be transmitted through the terrestrial globe, just as through a wire, to the greatest distances.'"

GOOD-BYE, OPERATOR—"Inventors have dreamed of devising some means to permit telephone subscribers to call one another without the aid of the central office, and in 20 years several apparatuses have been proposed and tried (and failed). The Direction Générale des Postes et des Télégraphes, of France, has installed a trial apparatus invented by an American and called the 'Auto-Commutator,' which gives direct communication, and assures entire secrecy of the conversation. Each subscriber has an instrument [*see illustration at left*] which comprises a battery, transmitter and receiver, a call bell, and a special mechanism which is indicated at the exterior by a dial provided with numbers. The dial in its motions actuates, via an electric current, the commutator placed in the central office."

FEBRUARY 1851

THE POISONER COOK—"The *Barnstable Patriot* writes that a letter received from Capt. Wm. Loring, of the bark [ship] *Governor Hinckley,* says that when ten days out of New York for London, an attempt was made by the cook to poison the officers and passengers, by introducing some poisonous substances into their coffee: the victims partook of the coffee but not in sufficient quantities to prove fatal to any one of them. Now, all this might have happened without the least attempt on the part of the poor cook. If coffee be kept hot in a copper vessel for five or six hours, it will dissolve part of the copper and become a poisonous drink. Coffee should not be kept in any other metallic vessel than tin or silver."

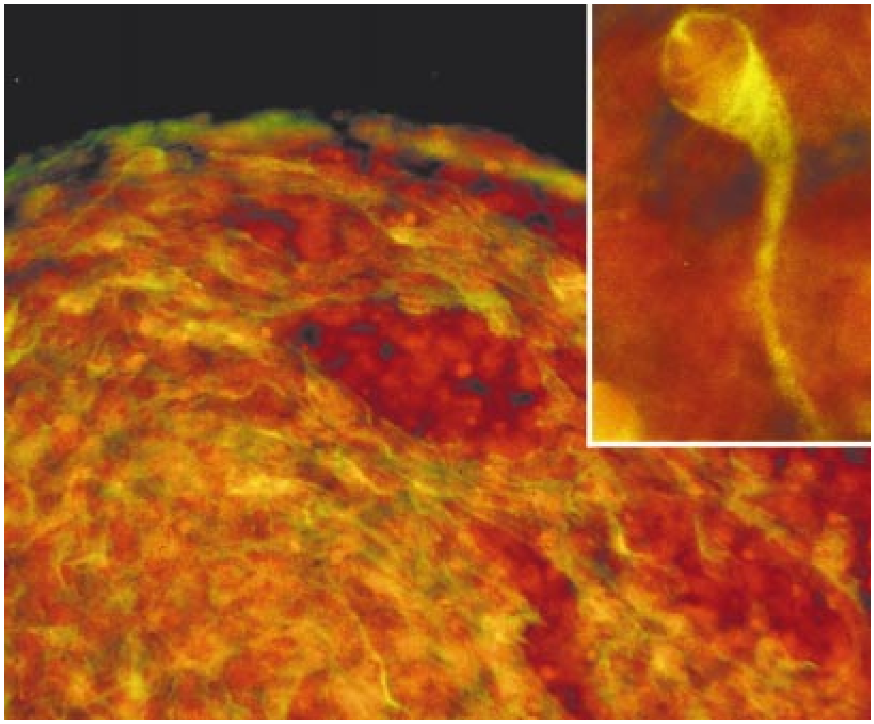
FROM *SCIENTIFIC AMERICAN*

NEUROSCIENCE_STEM CELLS

Biological Alchemy

The discovery that skin and bone marrow cells can transform into neurons raises hopes—and many questions

NEW ORLEANS—Two years ago Fred H. Gage set neurologists buzzing when he, his co-workers at the Salk Institute in La Jolla, Calif., and collaborators in Sweden disproved a long-standing "fact" that the human brain cannot grow new neurons once it reaches adulthood. That buzz has recently intensified into a hum of excitement as new observations of stem cells—immature cells that can divide repeatedly and give rise to many different kinds of tissues, including neurons—have found that the cells appear to be more accessible and more malleable than scientists had dared hope. Tantalized by the prospect of growing petri dishes full of neurons from a patient's own skin or marrow, several scientists spoke dreamily to reporters at a November 2000 conference in New Orleans about their hopes that transplanted stem cells could repair the nervous wreckage left by Parkinson's disease, Alzheimer's disease, multiple sclerosis, stroke or head trauma. Major newspapers ran with the story.



SPHERE OF STAR-SHAPED CELLS called astrocytes harbor new, growing neurons (*inset*). Scientists first identified these so-called neural stem cells in the brains of infant mice in December 2000. Most of the stem cells become dormant during childhood.

A close look at the details, however, suggests that the story has run ahead of the science. The most important recent experiments have uncovered three surprising properties of stem cells that together do raise the possibility of new therapies. But the results also raise a host of difficult questions.

Revelation number one is that stem cells from several places other than fetal tissue—a scarce and controversial source—can apparently be coaxed to produce neurons. Gage's group has isolated stem cells from the brains of recently deceased children and young adults. Cultured in a cocktail of nutrients, growth factors, antibiotics and serum from newborn calves, a tiny fraction of the cells lit up when the culture was stained with labels that stick to neurons. Dale Woodbury and Ira B. Black of the Robert Wood Johnson Medical School in Camden, N.J., cultured stem cells out of marrow from rats and adult humans. A different elixir, they found, forced as many as 80 percent of the cells to send out neuronlike arms and to express some of the same proteins that neurons do. And a team at McGill University led by Freda Miller presented similar results for stem cells that they have culled from the scalps of adult humans and the skin of rats.

The second surprise came when researchers injected neural stem cells into the spinal column or the fluid-filled ventricles of the brain. In almost every case, some of the cells migrated into injured tissue. One team saw this migration in monkeys whose nerves had been stripped of insulation to mimic the damage of multiple sclerosis. Another scientist found it in mice whacked on the head to mimic head trauma. Still others reported the phenomenon in rats injected with amyloid protein (a culprit in Alzheimer's), infected with a virus that kills motor neurons (as ALS, or Lou Gehrig's disease, does), or given a stroke in a surgical operation. In three of the rodent experiments, the animals that received stem cells regained more function than did control animals. Taken together, said Jeffrey Rothstein of Johns Hopkins University, the latest research indicates that stem cell transplants might enter human clinical trials within one to two years.

That optimistic forecast is hard to square with the scientific data, which are still clouded with uncertainties. One question is whether the cells that stain as neurons really are neurons. "Two or three markers don't make a neuron," said Theodore D. Palmer of Stanford University, who worked with Gage on the cadaver cells. "We still need to show that these cells snap onto other neurons and send electrical signals back and forth."

"There are many questions and caveats," Rothstein conceded later. "How long do these cells survive in the body? Do they become neurons? Do they make connections to the appropriate targets?"

The answers aren't yet known even for the cells transplanted into rats, mice or monkeys. And the results may well differ for humans.

The idea of growing human gray matter under glass still faces thorny issues as well. Black told reporters that "these cells grew so fast we had trouble keeping up with them," but Woodbury privately said that that was true only of the rat cells. Stem cells from human marrow stopped dividing after just four generations. Scientists at Geron, a biotech firm in Menlo Park, Calif., reported in New Orleans that they have a stem cell line taken from human embryos that is still dividing after 250 generations. But when they injected human stem cells into the brains of rats, the cells failed to transform into neurons. What is worse, surrounding brain tissue began to die.

Before stem cells can go into humans, researchers will have to make a convincing case that the benefits outweigh the risks. So far the improvements seen in animal studies, though measurable, have been small: previously paralyzed mice can flex their legs or splay their toes, for example, but they cannot stand. "To move into human trials based on this would, I think, be unethical," commented Martin E. Schwab, a neurologist at the University of Zurich.

Stem cells will probably be of little use to medicine until scientists solve a fundamental mystery about them: What combination of external signals and internal programming determines their fate in the human body? To solve this mystery, neurologists need to know which cells in the brain are the stem cells that give birth to neurons. In December, Pasko Rakic of Yale University and his collaborators claimed to have a firm answer. The stem cells are—at least in mice—not nondescript, youthful-looking cells, they concluded, but rather mature, star-shaped cells called astrocytes. During the brief window of infancy, these cells differentiate into neurons in all parts of the brain. Then the window closes at some point in childhood, and the stem cells fall dormant except in tiny regions of the ventricles and hippocampus, where neurogenesis continues.

Their paper concludes with a truly tantalizing idea: preliminary studies, they write, suggest that changing the chemical environment of even dormant astrocytes may reawaken their latent stem cell properties. Perhaps—many years or decades from now when the puzzle is solved—doctors will be able to repair brain damage from raw material that lies not in our bones or our skin but throughout the brain itself. *—W. Wayt Gibbs*

PHYSICS_ELEMENTARY PARTICLES

Higgs Won't Fly

CERN declines a massive opportunity to find the Higgs particle

In a move that surprised and dismayed many physicists, one of the world's leading laboratories has chosen not to continue an experiment that showed every sign of being on the verge of discovering an elusive particle that would have placed the capstone on a century of particle physics. The experiment was the last gasp of the venerable Large Electron-Positron collider (LEP), located near Geneva, Switzerland, and part of the European laboratory for particle physics (CERN). The particle was the long-sought Higgs, which is profoundly unlike any other particle discovered in human history and is the final jigsaw piece needed to complete the Standard Model of particle physics. The decision came down to the judgment of one man, Luciano Maiani, CERN's director general, who chose to shut down LEP on schedule to avoid delaying construction of CERN's next big experiment, the Large Hadron Collider (LHC), which is slated to be turned on in 2005.

Postulated independently by British physicist Peter Higgs and others in 1964, the Higgs plays a unique role in particle physics. In one guise, the Higgs is a field permeating the universe and giving the other particles their mass. If the field were turned off, the particles making up your body would presumably fly apart at the speed of light like so many photons. We have no way of directly detecting the all-pervasive Higgs field, but its other guise—individual Higgs particles, like tiny concentrated knots in the field—should be producible in violent collisions at accelerators. By studying the particle, physicists can verify the theory and pin down the Higgs's many unknown properties.

In 2000 researchers optimized the 11-year-old LEP to conduct one last search for the Higgs, pushing it to achieve collision energies of 206.5 billion electron

NOT THE LIGHT OF DISCOVERY: **Technicians in 1999 worked on one of the 3,368 electromagnets in LEP's 27-kilometer-long tunnel. Last November crews began dismantling LEP, despite hints that another major discovery may have been imminent.**

CERN

volts (GeV)—about 14 GeV beyond its original design parameters. Most likely the Higgs would be too massive to fall within LEP's extended reach, but in the summer, physicists saw signs of Higgs particles. Out of millions of collisions, nine produced Higgs candidates. A one-month extension to LEP yielded additional results, sufficient to conclude that the odds that the results were noise were one in 250—a tantalizing result but much too uncertain to proclaim "discovery." The data indicated that the Higgs has a mass of about 115 GeV (the remaining collision energy goes into creating a so-called *Z* particle at 91 GeV). By comparison, a proton is 1 GeV. A 115-GeV Higgs would agree nicely with predictions of supersymmetry models—the idea that particles in the Standard Model have "supersymmetric" partners.

Hoping to gain enough data to reduce the odds of error below the one in a million needed for a discovery, experimenters pleaded for a year's reprieve to LEP's scheduled dismantling, but after vigorous debate they were turned down. It was time to make way for the $4-billion LHC, which is to occupy the same 27-kilometer-circumference tunnel as LEP. Running LEP in 2001 would have cost CERN $65 million, including $40 million in civil-engineering contract penalties for delaying the LHC.

Chris Tully, the Higgs coordinator for one of the four LEP detectors and the person responsible for combining the data from all four, complains that what is most frustrating is the perceived failure of CERN's scientific decision-making process. Two different review boards discussed the Higgs evidence and the extension request, and both failed to recommend whether to proceed or not. Each board had roughly equal numbers of LEP and LHC scientists. Tully feels that part of the problem was the boards' not keeping to their proper terms of reference. For example, the LEP Scientific Committee, instead of limiting itself to the scientific issues, also considered the potential effect on LHC finances.

Maiani's decision could have been overturned at a special November 17 meeting of the CERN Council, representatives of CERN's 20 member countries—but again the result was a deadlock, and so Maiani's decision stood. "CERN is following a scientific program based on indecision," Tully says. Yet he doesn't fault Maiani, who, he considers, "made the wisest choice" from the perspective of a director general, who must give highest priority to the future of the laboratory, meaning the LHC.

LHC advocates insist that the decision was based on the science. Ana Henriques Correia, who leads construction on part of the LHC's ATLAS detector, says, "The scientific evidence [for Higgs] was not strong enough to postpone LHC." She points out that a sizable chance remained of no discovery by LEP even after a 2001 run.

Supporters argue that LEP was uniquely positioned to discover or rule out a 115-GeV Higgs promptly: after 11 years LEP's experimenters had a very good understanding of the performance of the accelerator and its four detectors. By comparison, the LHC's extremely complicated detectors are unknown quantities. Although the LHC is scheduled to collide its first protons in July 2005, collection of scientific data will not begin until 2007—after the lengthy process of commissioning, understanding and calibrating the accelerator and its detectors. Furthermore, CERN is discussing moving the start-up date back to the end of 2005.

The opportunity to discover the Higgs now passes to the Tevatron proton collider at the Batavia, Ill., Fermi National Accelerator Laboratory. The Tevatron discovered the top quark in 1995 and starts up again in March after a major upgrade. But it will take until about 2006 to gather sufficient data to claim discovery of the Higgs, if it is near 115 GeV (the device could see Higgs evidence up to 180 GeV). Paul Grannis, a member of the D-Zero experiment at the Tevatron, cautions that he doesn't know enough about the various factors in play to second-guess the CERN decision, but nonetheless he has "a hard time imagining why they did not" choose to continue. "We would be globally in so much better shape if we knew whether the Higgs were there or not, in trying to map out the future [accelerator] program."

These matters interest experimenters planning what to build *after* the LHC. The U.S., Japan and Germany are working on plans for the next electron-positron colliders, which will explore higher energies than LEP had. These devices would map out the detailed properties of the Higgs and other new particles, such as supersymmetric particles, expected to be discovered at the LHC. A Higgs under 130 GeV favors supersymmetry, and physicists understand very well what kind of program is needed to find and study supersymmetry. Above 130 GeV, "it is most likely not supersymmetry," Grannis says, "and then we're on a fishing expedition to figure out what the hell is going on."

—Graham P. Collins

OPTOELECTRONICS_LASERS

Cheap Light

Microlasers go deeper into the infrared to boost optical networking

It was so '80s, the dream of building an optical computer faster and more flexible than its electronic counterpart. That vision foundered because of the intrinsic challenges of processing light: simple things, like storing zeros and ones in the form of photons, proved inordinately difficult. These labors were not all wasted, however. The search for devices sufficiently small to meet the specifications for optical processors led to the development of lasers only a few millionths of a meter in width.

Although these small, cheap lasers, which can be integrated with a microchip, still won't make optical computing a reality, they are now opening new vistas in the still hot, Internet-driven market for optical communications. In the past few years, microlasers have reached the commercial marketplace, serving as transmitters for the dozens and dozens of fiber connections among the switching circuit cards in the huge routers (sometimes channeling trillions of bits each second) that send data packets along different paths in the network.

Sales for primarily short-reach, microlaser-based transmitter-receivers—including those in local-area networks—will increase from $262 million in 1999 to $14 billion in 2009, according to market researcher ElectroniCast. "They've blown away other types of lasers in terms of the quality of the light they produce and the cost of manufacturing," notes a report at Light Reading, a Web site that covers optical technologies.

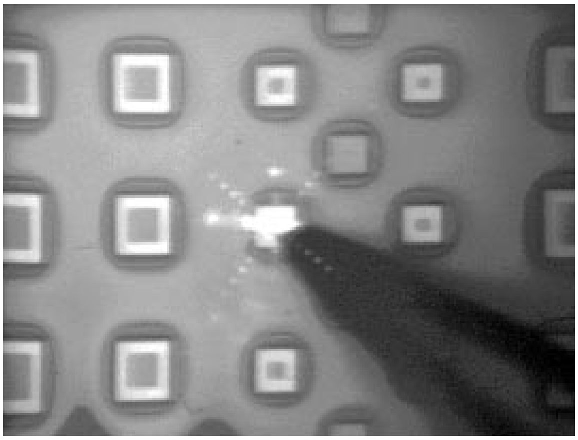
The technical name for these lasers is vertical-cavity surface-emitting lasers, or VCSELs ("vixels"). The technology differs from that of other types of semiconductor lasers, which are long, rectangular structures that beam light out the sides and that show up in everything from CD players to transoceanic fiber-optic networks. The tiny vertical lasers, in contrast, eject photons from the surface of the semiconductor wafer on which they are manufactured, a design that offers distinct manufacturing advantages over edge emitters.

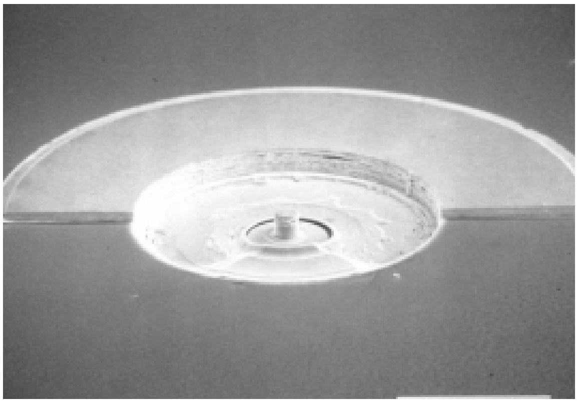
For instance, VCSELs can be tested by the thousands on the silicon wafer on which they are made and so do not have to be diced up and probed individually, as edge emitters must be. The sub-10-micron-diameter high-efficiency devices can be manufactured as an array that is integrated with surrounding electronics. Their cylindrical beam permits the narrow emission to move easily into the fiber and lets the units be packaged more easily into transmitter-receiver modules.

Unfortunately, materials and manufacturing issues have confined VCSEL light to the near infrared (around 0.85 micron), which can be used to transmit to distances up to one kilometer, though often much less. But new systems under development may make microlasers ubiquitous throughout local phone networks, and perhaps one day they may even show up in the home. "If everyone is to have a laser in his or her computer, then you can't have one costing $1,000. So cheap lasers will become increasingly important," says Dennis Deppe, a VCSEL researcher at the University of Texas at Austin.

To reach this objective, VCSELs must emit light farther into the infrared, at the 1.31- and 1.55-micron wavelengths needed for medium- and long-range telecommunications. The big technical glitch relates to the spacing of atoms in the crystal lattice of the semiconductor materials used to make up the primary components of a laser: the active region, the lasing material that resides inside a cavity, and the mirrors that bookend it.

A VCSEL works when electricity or light is "pumped" into the semiconductor cavity, thereby exciting the electrons to a higher energy level. When they jump back to a lower energy state, the electrons combine with "holes" (positively charged areas) to emit photons, which then bounce back and forth off the mirrors and cause other electrons to leap the energy gap and emit new photons. Eventually the photons penetrate through one of the mirrors as a coherent beam.





MICROLASER LIGHT emits from a 1.31-micron VCSEL at Sandia National Laboratories (*top*); the University of California at Santa Barbara made a 1.55-micron version.

KENT GEIB *Sandia National Laboratories (top)*; UNIVERSITY OF CALIFORNIA AT SANTA BARBARA *(bottom)*

In long-wavelength systems, it is very difficult to line up the atoms that constitute the active region with those in the mirrors. Misalignments strain the materials and can cause defects that make a good laser go dark. Some companies fabricate the mirrors on different wafers from those in the active region and then fuse the components. But critics say this approach adds cost and complexity—although it may be closest to the commercial marketplace. Recent excitement relates to research that circumvents bonding by growing epitaxially, or layer by layer, a single unitary crystal. "There's a race today to make things lattice-matched," says Larry Coldren, a VCSEL researcher at the University of California at Santa Barbara, adding, "The goal is to grow one wafer, process it and stick it in a package; you don't want to grow three wafers and stick one to the other."

Using a material originally pioneered by Hitachi, Sandia National Laboratories and Cielo Communications put forward one solution to the lattice mismatch problem last year when they reported a method for growing a 1.31-micron VCSEL as a single crystal. Researchers made the active region from a compound material—indium gallium arsenide nitride—that matched closely enough the atomic spacing of the mirrors (alternating layers of gallium arsenide and aluminum gallium arsenide) to make a working 1.31-micron laser.

Another way investigators are improving VCSELs is through better confinement of electrons and holes in the active region, which both facilitates the emission of photons in the laser and allows tailoring of the desired wavelength. The active region of today's VCSELs contains quantum wells, layers of semiconductors that confine electrons and holes in a flat, two-dimensional space. A few researchers have taken this confinement approach much further, replacing quantum wells with quantum dots. Such dots function as nanotechnological boxes, holding electrons and holes essentially to a single point—at the precise energy level and location at which they can lase.

A quantum dot thus becomes the ultimate designer material for tailoring the wavelength desired. Just as important, the dot size, perhaps 100 atoms across, means that the bonds in the atoms in the dot and the surrounding cavity can distort to conform to the atoms in the lattice of the mirrors without causing dislocations. VCSELs could become the first commercial application for quantum dots. Other VCSEL efforts target the 1.55-micron wavelength used in most long-haul fiber telecommunications, including tunable lasers that can change wavelengths to reroute a transmission or deliver bandwidth on demand.

Still, engineering long-wavelength VCSELs—making quantum dots that have the necessary performance characteristics, for example—remains a challenge. But the payoff, cheap optical networks that reach all the way into your bedroom, suggests that these devices may experience a better fate than the optical computer. *—Gary Stix*

PALEONTOLOGY_EXTINCTION

Mammoth Kill

Did humans hunt giant mammals to extinction? Or give them lethal disease?

MEXICO CITY—Although it's hard to imagine in this age of urban sprawl and automobiles, North America once belonged to mammoths, camels, ground sloths as large as cows, bear-size beavers and other formidable beasts. Some 11,000 years ago, however, these large-bodied mammals and others—about 70 species in all—disappeared. Their demise coincided roughly with the arrival of humans in the New World and dramatic climatic change—factors that have inspired several theories about the die-off. Yet despite decades of scientific investigation, the exact cause remains a mystery. Now new findings offer support to one of these controversial hypotheses: that human hunting drove this megafaunal menagerie to extinction.

The overkill model emerged in the 1960s, when it was put forth by Paul S. Martin of the University of Arizona. Since then, critics have charged that no evidence exists to support the idea that the first Americans hunted to the extent necessary to cause these extinctions. But at the annual meeting of the Society of Vertebrate Paleontology in Mexico City last October, paleoecologist John Alroy of the University of California at Santa Barbara argued that, in fact, hunting-driven extinction is not only plausible, it was unavoidable. He has determined, using a computer simulation, that even a very modest amount of hunting would have wiped these animals out.

Assuming an initial human population of 100 people that grew no more than 2 percent annually, Alroy determined that if each band of, say, 50 people killed 15 to 20 large mammals a year, humans could have eliminated the animal populations within 1,000 years. Large mammals in particular would have been vulnerable to the pressure because they have longer gestation periods than smaller mammals and their young require extended care.

Not everyone agrees with Alroy's assessment. For one, the results depend in part on population-size estimates for the extinct animals—figures that are not necessarily reliable. But a more specific criticism comes from mammalogist Ross D. E. MacPhee of the American Museum of Natural History in New York City, who points out that the relevant archaeological record contains barely a dozen examples of stone points embedded in mammoth bones (and none, it should be noted, are known from other megafaunal remains)—hardly what one might expect if hunting drove these animals to extinction. Furthermore, some of these species had huge ranges—the giant Jefferson's ground sloth, for example, lived as far north as the Yukon and as far south as Mexico—which would have made slaughtering them in numbers sufficient to cause their extinction rather implausible, he says.

JEFFERSON'S GROUND SLOTH, woolly mammoths and other North American megafauna vanished mysteriously around 11,000 years ago.

MacPhee agrees that humans most likely brought about these extinctions (as well as others around the world that coincided with human arrival), but not directly. Rather he suggests that people may have introduced hyperlethal disease, perhaps through their dogs or hitchhiking vermin, which then spread wildly among the immunologically naive species of the New World. As in the overkill model, populations of large mammals would have a harder time recovering. Repeated outbreaks of a hyperdisease could thus quickly drive them to the point of no return. So far MacPhee does not have empirical evidence for the hyperdisease hypothesis, and it won't be easy to come by: hyperlethal disease would kill far too quickly to leave its signature on the bones themselves. But he hopes that analyses of tissue and DNA from the last mammoths to perish will eventually reveal murderous microbes.

The third explanation for what brought on this North American extinction does not involve human beings. Instead its proponents blame the loss on the weather. The Pleistocene epoch witnessed considerable climatic instability, explains paleontologist Russell W. Graham of the Denver Museum of Nature and Science. As a result, certain habitats disappeared, and species that had once formed communities split apart. For some animals, this change brought opportunity. For much of the megafauna, however, the increasingly homogeneous environment left them with shrinking geographical ranges—a death sentence for large animals, which need large ranges. Although these creatures managed to maintain viable populations through most of the Pleistocene, the final major fluctuation—the so-called Younger Dryas event—pushed them over the edge, Graham says.

For his part, Alroy is convinced that human hunters demolished the titans of the Ice Age. The overkill model explains everything the disease and climate scenarios explain, he asserts, and makes accurate predictions about which species would eventually go extinct. "Personally, I'm a vegetarian," he remarks, "and I find all of this kind of gross—but believable."

—Kate Wong

See www.sciam.com/interview/2001/010201macphee/index.html for Ross MacPhee's explanation of his hyperdisease hypothesis.

KAREN CARR

AVIATION_SAFETY

Collision Decision

New radar systems may prevent deadly accidents on congested runways

It was pitch-black and raining hard—Typhoon Xangsane was moving into northern Taiwan. The pilot of Singapore Airlines Flight 006 turned onto runway 05R at Chiang Kai-shek International Airport in Taipei and advanced the throttles on the Boeing 747-400. The airplane began rolling in the blinding rain. Then the pilot suddenly cried out, "Something there!" and pulled back hard on the yoke to hop over the object. But the plane plowed into a barricade at a speed estimated at up to 163 miles per hour. It disintegrated and erupted into flames. Of the 179 people on board, 83 perished.

The October 31 tragedy happened because the pilot had been cleared to take off from runway 05L, not runway 05R, which had construction equipment on it. In the previous 10 years, 63 people died in such "runway incursions." As airports grow busier, that number is expected to rise substantially. The past five years have already seen a 60 percent increase in incursions. In 1999 airlines reported 321 incidents, and in 2000 they had logged 403 incidents by early December. A study released last November found that, in the U.S. alone, the next two decades could see 700 to 800 deaths and 200 injuries from runway collisions if nothing is done to improve safety.

In light of the hazards, the Federal Aviation Administration developed ASDE-3 (Airport Surface Detection Equipment, version 3), which it has installed at the 34 busiest U.S. airports. ASDE-3 is essentially a ground-based radar that detects a vehicle, calculates its intended path, and broadcasts the information onto the air-traffic controller's radar screen. The controller then must radio instructions to the vehicle. "Our big effort is in heightened awareness for controllers and pilots," says William Shumann, an FAA spokesperson. The FAA has also begun installing an enhancement to ASDE-3 called AMASS, for Airport Movement Area Safety System, which provides the controller with an aural and visual alert. The early version of AMASS gave frequent false alarms whenever pilots approached a runway "hold" line. "But," Shumann states, "we restructured the program in the summer of 1999" to eliminate those problems.

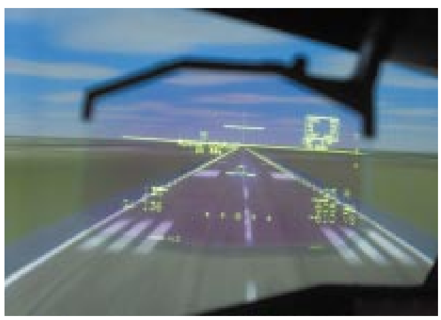
Critics, though, argue that relying on human controllers for action is too passive a strategy for events that happen in a matter of seconds. Moreover, "radars have problems with ground clutter and blockage," explains Warren Morningstar, vice president of communications for the Aircraft Owners and Pilots Association, a pilots advocacy group based in Frederick, Md. "It's very difficult to design a system that can see an entire surface area."

The National Aeronautics and Space Administration may have the answer: the Runway Incursion Prevention System, or RIPS, a radar system more active than the FAA's improved ASDE-3. "It's one part of an entire system we are developing to give pilots a clear electronic picture of what's outside their window," explains Kathy Barnstorff, a spokesperson for the NASA Langley Research Center in Hampton, Va. RIPS attacks incursions with a three-pronged approach. First, pilots can use a color "head-down" moving map on the control panel, which graphically illustrates the runway or taxiway and warns of conflicts in either yellow (for runway traffic) or red (for runway conflict). Second, they can use a transparent head-up display, similar to that on a fighter jet, that flashes a text warning. Finally, they can hear a two-stage auditory warning.

Last October, NASA tested RIPS, along with the FAA system, in a specially equipped Boeing 757 at Dallas–Fort Worth International Airport. As the airliner approached to land, a van on the ground would cross a hold line and enter the runway. Out of 47 test runs, NASA's system alerted pilots 42 times, compared with 36 times for the FAA's system. "I see some tremendous benefits to the [RIPS] system," says Richard Grue, a technical pilot with American Airlines who tested the system. "I'd be willing to bet that with the aural incursion system, the Singapore Airlines accident wouldn't have occurred."

But it will take a year or so before NASA's moving map could be on the market, Barnstorff says. And the FAA won't finish installing AMASS, whatever its shortcomings, until September 2002. In the meantime, other safety measures can be enacted right now. Morningstar advocates widening stop bars on runways and improving and updating signage. "There are low-tech solutions out there that ought to be implemented immediately," he says, "well in advance of looking for a high-tech silver bullet." —*Phil Scott*

PHIL SCOTT, based in New York City, specializes in aviation technology.



HAZARD-ALERT system by NASA warns pilots via a moving-map display (*left, behind yoke and outlined by orange stripe*) and a "head-up" display (*above*).

NASA

Debit or Credit?

Whether CO_2-consuming trees can offset global warming is far from certain

Should countries be able to count forests as credits against the amount of carbon dioxide and other heat-trapping greenhouse gases they are allowed to emit? That is perhaps the biggest question that stymied recent negotiations on how to implement the Kyoto Protocol, an international plan to curb global warming by cutting emissions of greenhouse gases.

When the agreement was hammered out at a United Nations conference in 1997, the participating countries agreed to count forests planted since 1990 as carbon sinks—and, in doing so, as credits that would offset required cuts in emissions. Oceans and forests absorb more than half of the CO_2 put out by burning fossil fuels, so it seemed to make sense to count both sinks and sources in this international accounting game.

But when the U.S. claimed a whopping 310 million metric tons of carbon in its forests as emissions credits, other countries balked at the idea. This dispute raises a second question: Does science know enough to verify the role of carbon sinks in a quantifiable way?

"Hard to say," answers Allen M. Solomon, a senior global ecologist with the U.S. Environmental Protection Agency. "These sinks are highly variable." The week before Kyoto negotiators met in the Hague, the Netherlands, last November, a handful of scientific reports were reiterating the uncertainty over how much carbon forests actually soak up.

Many climate-change scenarios assume that extra carbon dioxide in the atmosphere will make trees grow faster—"a bit like dumping fertilizer on them," says Jorge L. Sarmiento, an atmospheric chemist at Princeton University. But no one has ever uncovered unequivocal evidence of this phenomenon, called CO_2 fertilization. In fact, a November report in *Science* indicated the contrary. Princeton ecologist John P. Caspersen and his colleagues failed to detect signals of CO_2 fertilization when they analyzed growth rates of more than 20,000 forest plots in five eastern U.S. states. Bottom line: if there is no CO_2 fertilization, the forest carbon sink will eventually disappear. Solomon, who advised White House representatives during the November negotiations, says Caspersen's conclusions were "right on the button."

Even if CO_2 fertilization does occur, according to a recent report in *Nature,* the benefit doesn't last forever. Peter M. Cox of the Hadley Center for Climate Prediction and Research in Berkshire, England, and his colleagues conducted a global climate simulation and found that by the middle of the century, the land becomes a source of CO_2 rather than a sink. Apparently, the estimated atmospheric warming of about 5.5 degrees Celsius over land would invigorate soil microbes that give off CO_2 as they digest fallen tree leaves and other organic material. The CO_2 output of the soil microbes would eventually outpace trees' ability to absorb carbon, and by 2100 an extra six billion metric tons of carbon will remain in the atmosphere every year. Compensating for this net gain in CO_2 by making additional cuts in greenhouse-gas emissions could cost $1.2 trillion a year, according to Sarmiento's estimates.

Carbon sinks can vary dramatically on short timescales as well. Even in the five years of the Kyoto Protocol's first commitment period, land-based ecosystems could give off more carbon dioxide than they absorb, Solomon says. A year of rampant wildfires, for instance, would release loads of extra CO_2 as the forests and grasslands turned to ash.

In the end, scientific uncertainty took a backseat to politics and economics in the Kyoto negotiations. And many scientists question the idea of using carbon sinks as emissions credits in the first place. "Kyoto was about *reducing* emissions," Solomon says. "I don't think sinks make a good deal of sense, because they don't solve the problem."

The U.S. softened its controversial stance in the final hours of the conference by reducing its credit claims from 310 million metric tons of carbon to 20 million tons. But European negotiators found even the scaled-back U.S. plan unacceptable.

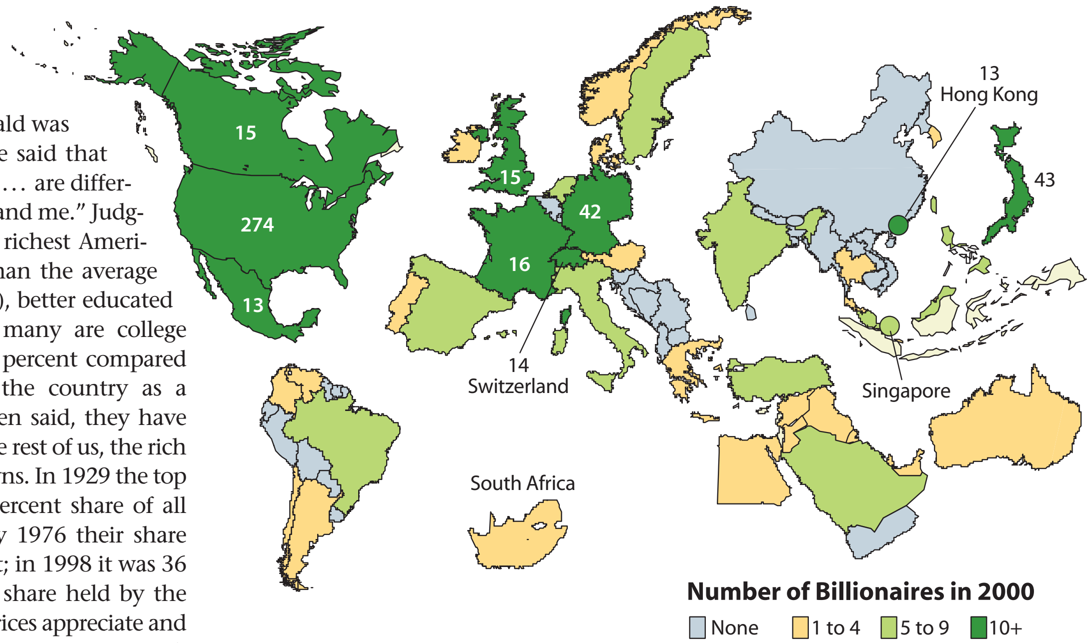
Some 180 nations are scheduled to rehash the details again this spring, but time is running short: the first emission-reduction targets—7 percent below 1990 levels in the U.S.— must be met by 2012. We have to start somewhere, Sarmiento says. "Not as a scientist but as a citizen, I'd like to see them strike a bargain," he adds, "because I'm really concerned about what's going to happen if we do nothing."

—Sarah Simpson

OUT OF SINK: **The U.S. and Europe disagree on how to count forests, which act as a sink for carbon dioxide released when factories and cars burn fossil fuels.**

NICHOLAS PAVLOFF *Photonica (left)*; GERRY JOHANSSON *Photonica (right)*

The Rich and Other Americans



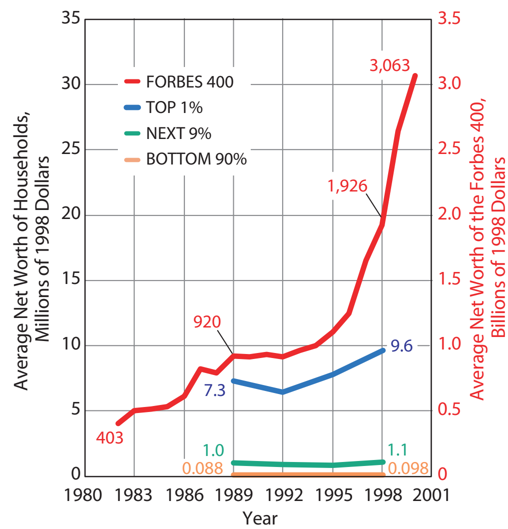
F. Scott Fitzgerald was right when he said that the "very rich . . . are different from you and me." Judging by the Forbes 400 richest Americans, they are older than the average American (by 12 years), better educated (more than twice as many are college graduates), whiter (95 percent compared with 71 percent for the country as a whole) and, as has been said, they have better teeth. But like the rest of us, the rich have their ups and downs. In 1929 the top 1 percent held a 44 percent share of all personal assets, but by 1976 their share had sunk to 20 percent; in 1998 it was 36 percent. Typically the share held by the rich rises when stock prices appreciate and the price of housing, the preeminent middle-class asset, rises less swiftly—precisely what happened in the 1990s.

The single biggest reason for the spectacular increase in average assets of the Forbes 400 is the growth of electronic technology, based not only in the computer, software and Internet sectors but also in retailing, finance and mass media. In 1998 the Forbes 400 accounted for an estimated 2.6 percent of total personal net worth held by all Americans, compared with 33 percent held by the remaining one million households in the top 1 percent. The 9.2 million households in the next 9 percent held 34 percent, and the bottom 92.3 million households held 31 percent.

Perhaps a more pertinent indicator is financial wealth, which is calculated as net worth less net equity in owner-occupied dwellings and so is a measure of the more liquid assets available. An analysis by economist Edward N. Wolff of New York University showed that the bottom 40 percent of middle-aged householders in 1998 had virtually no financial wealth and thus were exceptionally vulnerable to economic shocks or personal disability. The financial wealth of the middle 20 percent would typically carry them for two to four months. The figures for the next 20 percent and the top 20 percent are, respectively, eight to 18 months and two to seven years.

The measure of wealth used in the chart is net worth—that is, assets such as real estate, securities, businesses, checking accounts and so on, less any debts. Factoring in Social Security and other pensions, however, lowers the shares held by the rich: by one estimate, the top 1 percent in 1992 held 34 percent of personal net worth but only 20 percent of the total when pensions are included.

In 1998, 27 percent of black and 36 percent of Hispanic households had zero or negative net worth, compared with 15 percent of non-Hispanic whites. Inheritance plays a crucial role in wealth disparities: 24 percent of white households in 1998 reported ever receiving an inheritance (average value $115,000 in 1998 dollars) compared with 11 percent of black households (average value $32,000). Blacks' efforts to accumulate wealth have historically also been stymied by inferior access to credit and housing markets.



SOURCE: Based on Forbes*'s 2000 tabulations, the* 400 Richest Americans and the World's Richest People. *Data used with permission of Forbes.com. Chart also includes data from the Federal Reserve Board's triennial studies,* Survey of Consumer Finance. *The Federal Reserve estimates for the top 1 percent group specifically exclude those in the Forbes 400. Figures on chart indicate average wealth in millions of dollars.*

By *Forbes*'s estimate, in 2000 there were 590 billionaires worldwide, including nine kings, queens and dictators, plus 13 family fortunes in which multiple heirs participated. The U.S., with about half the total, has been most successful in producing billionaires. For complicated historical and cultural reasons, such as the distinctive American emphasis on individuality, the U.S. taxes the rich far less than most other industrial countries do.

—Rodger Doyle (rdoyle2@aol.com)

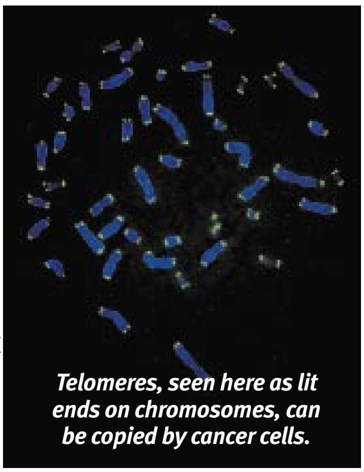
RODGER DOYLE

News Briefs

ONCOLOGY

Copycaps

In a normal cell, caps on the ends of chromosomes, called telomeres, get shorter each time a cell divides, thereby limiting its life span. Cancer cells rely on the enzyme telomerase to repair the telomeres, enabling them to keep dividing indefinitely. Some cancer cells, though, can repair their telomeres without the telomerase. Melissa A. Dunham and her colleagues at the Children's Medical Research Institute in Sydney have now found that these cells do it by copying existing telomeres. The team marked telomeres in human cells with pieces of bacterial DNA. These tags later showed up in the telomeres of other chromosomes in the cell. The next step is to identify the enzymes that drive the copying, because they will be the targets for new anticancer drugs. The work appears in the December 2000 *Nature Genetics*. —*Diane Martindale*



JERRY SHAY University of Texas Southwestern Medical Center

Telomeres, seen here as lit ends on chromosomes, can be copied by cancer cells.

CHEMISTRY

Carbon Original

The structure of tetrahedral carbon compounds is drummed into every student who has survived organic chemistry: a central carbon atom bonds to four other atoms to form a tripod (three below and one above the carbon). But a glaring anomaly has just been calculated to be possible. Instead of being three-dimensional, molecules could exist in which carbon lies in the center of a plane, maintaining bonds with six other atoms. Reporting in the December 8, 2000, *Science*, Paul von Ragué Schleyer and Kai Exner of the University of Georgia found in their computational research that the electronic arrangement of such flat, six-bonded carbon molecules would be related to that of well-known benzene. Synthesizing such stable, flat carbon molecules could yield compounds with novel properties. —*Steve Mirsky*

DATA POINTS

Death Defying

Average time between sentencing and execution in U.S.:	10.6 years
Percent of death-penalty sentences found to have a serious error on subsequent appeals:	68
Percent for noncapital cases:	15
Percent of those convicted who are later determined to be innocent:	5
Number mistakenly executed since 1900:	at least 23
Approximate cost of a murder trial in Los Angeles County:	$625,000
Cost when the death penalty is sought:	$1.9 million
Cost to New York State to put five men on death row (since 1995):	$23 million
Percent of law-enforcement officials who do not believe capital punishment reduces the homicide rate:	67

Average homicide rate per 100,000 for
- Death-penalty states: 9.3
- Entire U.S.: 9

MATT COLLINS

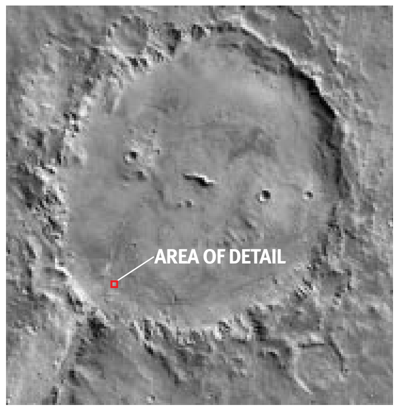
SOURCES: A Broken System: Error Rates in Capital Cases, 1973–1995, *by James S. Liebman et al., 2000 (http://justice.policy.net/jpreport/); Death Penalty Focus (www.deathpenalty.org); National Association of Criminal Defense Lawyers; American Civil Liberties Union*

PLANETARY SCIENCE

Mars Water

NASA/JPL/MALIN SPACE SCIENCE SYSTEMS

The latest Mars Observer images lend even more credence to the supposition that Mars was once a good place to swim. Researchers reporting in the December 8, 2000, *Science* revealed photographs showing sedimentary rock layers that have filled impact craters near the Martian equator. Most likely, water is responsible for the rocky buildup, much the way it is for sedimentary rocks on Earth. The layers are estimated to have formed some four billion years ago—about the time life was beginning on Earth. Future expeditions to the Red Planet may target these areas in a search for past life. *—Philip Yam*



Holden Crater contains rounded slopes* (detail above) *near where a valley enters the crater, suggesting that water once drained there.

ANCIENT ASTRONOMY

Stellar Work

Much aligned

YASUMITSU OKURA *Photonica*

If Egyptians were building the pyramids today, they could use Polaris, the North Star, to orient their constructs to the celestial pole. But 4,500 years ago there was no star to align to north, thanks to the precession, or wobbling, of Earth's axis, which shifts the pole around on a 26,000-year cycle. It's been a mystery how the Egyptians managed to orient the pyramids so accurately—the eastern and western sides of the Khufu (or Cheops) pyramid deviate only three arc minutes, or about 1⁄10 the apparent diameter of the moon as seen from Earth, from celestial north. In the November 16, 2000, *Nature,* Kate Spence of the University of Cambridge describes how an Egyptian with a plumb line in 2467 B.C.E. could have done the job: that year the celestial pole fell within a straight line drawn from Mizar, in the Big Dipper, down to Kochab, in the Little Dipper. Pyramids built before Khufu seem skewed somewhat to the west, whereas those constructed afterward steer slightly to the east—by amounts largely in accordance with Earth's precession. The correlation suggests that the age of the pyramids can be dated to within five years—a vast improvement over the previous 100-year-wide error margins. *—P.Y.*

News Briefs

PALEONTOLOGY

Bad Breathosaur

Some experts have speculated that, like today's Komodo dragons, *Tyrannosaurus rex* and other large meat-eaters of the past inadvertently cultivated bacteria on bits of flesh trapped in their teeth. After a quick bite, such creatures would infect, and perhaps ultimately kill, their prey with the bacteria created by their poor dental hygiene. Now two Mexican researchers have found a half-inch-long tooth from a wolf-size, meat-eating dinosaur, or theropod, that they surmise was adapted specifically to harbor toxic bacteria. They point to a dimpled groove running along the curved tooth. "It represents a true venom groove, an extremely specialized structure that houses infectious bacteria," says Rubén A. Rodríguez de la Rosa, a paleontologist at the Museum of the Desert in Saltillo, Mexico. Rodríguez presented his findings at the Society of Vertebrate Paleontology meeting in Mexico City last fall along with co-author Francisco Aranda-Manteca of the Autonomous University of Baja California in Ensenada, Mexico.

COURTESY OF RUBÉN A. RODRÍGUEZ DE LA ROSA *Museum of the Desert*

Fossil tooth of an unnamed theropod has a groove that may have harbored deadly bacteria.

Snakes and several other reptiles have internal canals for conducting venom, but this external groove is unique, Rodríguez says. The dinosaur itself could have been protected from the nasty bacteria by a thin layer of skin from the gums.

The tooth was actually discovered by an undergraduate in 1989. It sat unnoticed in Ensenada until Rodríguez sifted through the university's collection earlier this year. The theropod is as yet unnamed but may belong to the *Dromaeosaurus* genus.

Philip Currie, director of the Royal Tyrrell Museum in Drumheller, Alberta, has examined the tooth and believes the groove may serve the same purpose as a groove on a bayonet—to make it easier to pull out of the flesh. "If you have a poison groove, it ends up being much deeper," Currie says. "This is a deep groove but not that deep." As for the infection theory, Currie remarks, "I didn't buy it for *T. rex,* and it's harder to buy it with this one."

Aranda-Manteca and his students went back this month to the coastal Baja site. "The only way to know" for sure if theropods infected their prey, Aranda-Manteca says, "is to find more teeth and bones."

—Eric Niiler

ASTROPHYSICIST_RICHARD A. MULLER

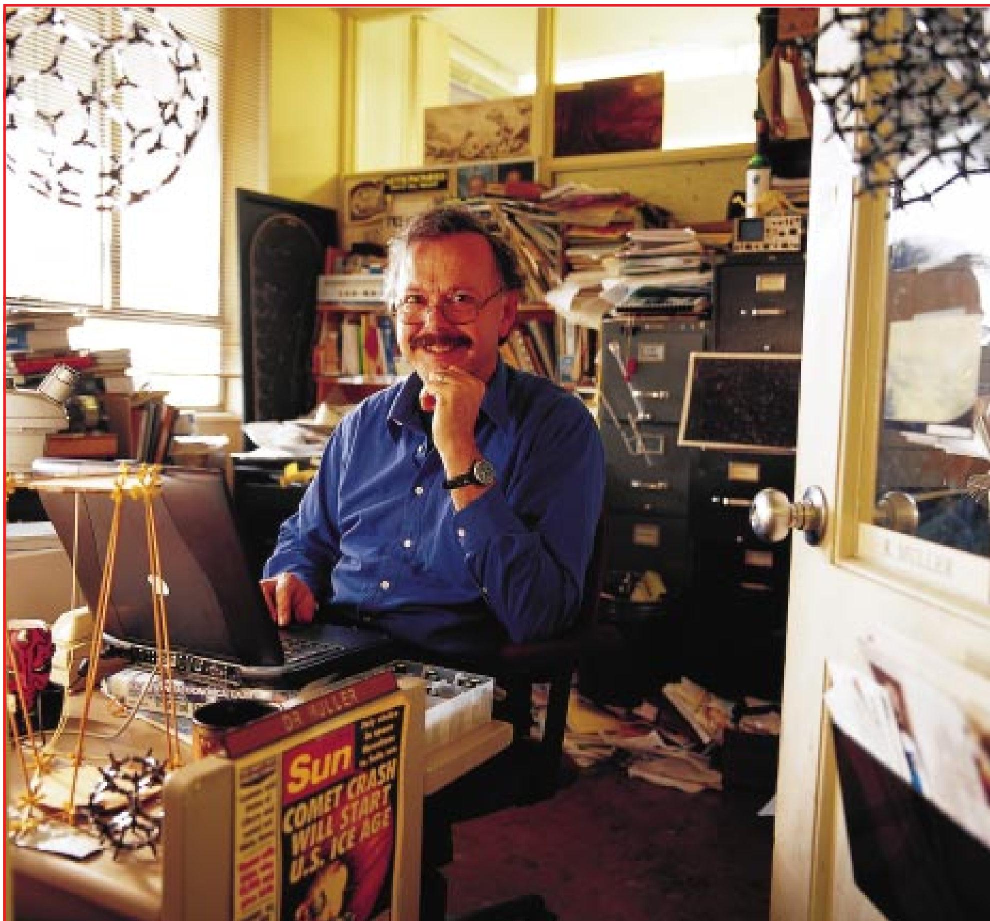
One Disaster after Another

The father of the idea that a sibling of the sun periodically wreaks havoc on Earth finds inspiration in catastrophes

BERKELEY, CALIF.—I first meet Richard A. Muller during a record-breaking heat wave. The astrophysicist is on his way to get a refreshment. Bottles of his favorite cold dairy drink—mocha milk—are stacked in a nearby vending machine. Through the clear front, the scientist notices something out of place: a juice can trapped obliquely against the glass. "I'll get either two drinks or none," he predicts playfully, inserting his change and selecting the beverage he thinks is most likely to knock the can free. Muller is unconcerned (or perhaps oblivious) that this selection is vanilla, not the flavor he came for. His purchase grazes the target but fails to knock the bottle down.

Gambles like this one typify the life of Richard Muller—although usually the stakes are higher. The restless researcher loves to prowl for new scientific territory to conquer. "You need to have one interesting idea every day," he says. His graduate research concerned particle physics, but his accomplishments range from inventing an improved technique for carbon dating to designing an experiment for measuring the cosmic background radiation left over from the big bang about 15 billion years ago.

These and other accomplishments won Muller a MacArthur Fellowship in 1982, a year after these so-called genius awards began. It was a turning point. After that, Muller felt liberated to do "crazy things," as he puts it. "Just like James Bond has a license to kill, I had a license to depart from the normal path of a scientist."



EDWARD CALDWELL

RICHARD A. MULLER: SCIENTIFIC FREE SAFETY

- ***Born in New York City, 1944; wife, Rosemary, heads architectural firm; two daughters, Betsy and Melinda***
- ***Published novel on the life of Jesus, providing commonsense explanations of miracles (water walking done by surreptitious use of a submerged dock, for example)***
- ***Invention in progress: A way to spray water 10 kilometers or more to extinguish fires***
- ***Professional philosophy: "My best achievements have come when I strike out and do something crazy"***

On the surface at least, he fits the stereotype of a scientist. He will head to the lab in the middle of the night when an idea strikes him. His cluttered office, which overlooks the Berkeley campus of the University of California, where he has been since he received his Ph.D. in particle physics here 32 years ago, could be a set from an absentminded-professor comedy. There's hardly enough floor space for a visitor amid filing cabinets and desks and cartons overflowing with journals and papers. His in-box groans under a two-foot-high stack. "My research has been one disaster after another," Muller puckishly offers. This well-rehearsed line is quite literally true. He did work on the big bang. He studied the violent supernova explosion preceding the creation of the sun. And then there's his Nemesis.

"Nemesis" refers to a seemingly bizarre hypothesis concerning the evolution of life on Earth. Muller hatched it one day in 1983 when his mentor, Nobel laureate Louis Alvarez, enlisted the young physicist to debunk a research paper showing that Earth has sustained significant plant and animal extinctions at regular intervals—every 26 million years. Alvarez and his son, Walter, had recently advanced the theory that dinosaurs were the casualty of a Mount Everest–size comet that hit the planet 65 million years ago. At the time, the hypothesis was scoffed at; now it is generally accepted. Playing devil's advocate for Alvarez, Muller conjured up a scenario. Suppose, he suggested, the sun has a sibling around which it do-si-dos every 26 million years. And suppose that once each revolution the star swings through the Oort cloud, a calving ground for comets between four trillion and 10 trillion miles from us. Perhaps some of those icy balls, of which there are billions, would be knocked off-kilter and sent hurtling into Earth.

At first the idea seemed preposterous, even to Muller himself. But neither Muller nor Alvarez could think of any reason why the theory couldn't be true. With a touch of whimsy, Muller dubbed the star Nemesis, after the Greek goddess who fends off human folly. "We worry that if the companion is not found," he stated in the scientific article introducing the theory, "this paper will be our nemesis."

It seems counterintuitive that the solar system could be looping around an unknown star, but in fact most stars have partners: some 85 percent have some kind of companion. The only way to identify which, if any, of the catalogued stars is the sun's sibling requires measuring the distances to them. Muller says the elliptical orbit of Nemesis would get no farther than about 18 trillion miles from Earth, about three light-years away and three quarters the distance to the closest known star, Alpha Centauri. It could be a red dwarf star, which might be bright enough to be seen with a small telescope, or, less likely, a brown dwarf, which might not be visible at all.

When he dreamed up the theory nearly two decades ago, Muller thought he would locate Nemesis in just a few years. Given its putative distance and brightness, it should be easy to find such a star through parallax measurements—seeing how it shifts against the more distant stellar background as Earth moves along in its orbit. But the search, short on funds for telescope time, languished and stalled. Muller says most astronomers think his theory was disproved, when in fact it is simply in limbo.

It is no coincidence that so much of his career has been spent studying such tumultuous events. For centuries, scientists have predicated theories about Earth's evolution on the principles of uniformitarianism and gradualism, which posit that by and large the planet evolved slowly, relying on the same forces we see at work today, such as erosion and continental drift. Muller, however, believes infrequent, violent events are just as important—a doctrine some call catastrophism. Muller says neglect of catastrophic explanations gives him a strategic opportunity: "That's where the discoveries are."

Most recently, Muller has begun delving into the ice ages. Geologists still have a hard time explaining why they come and go. Muller insists the answer is of much more than academic interest. Springing from his office chair, he heads to a blackboard in an adjoining room—he couldn't locate any chalk in his office—and sketches a graph of global temperature since the industrial revolution. Overall, global temperature has gone up about 1.5 degrees Fahrenheit in the past 120 years—and 15 to 20 degrees since glaciers receded 12,000 years ago. "Anything that can have an impact of 15 degrees is probably having an impact on the present climate," he reasons.

INVADING ICE, from this January 1929 issue of *Amazing Stories*, sparked a young Muller's interest in ice ages.

Ice ages come and go at approximately 100,000-year intervals. The conventional explanation, refined and popularized by Serbian mathematician Milutin Milankovitch in the decades before World War II, involves subtle irregularities in Earth's motion. The theory mainly posits that the eccentricity, or out-of-roundness, of Earth's orbit varies the amount of sunlight bathing our planet.

Painstaking reconstructions of Earth's past movements show that the planet's orbit around the sun goes from almost perfectly round to slightly oval and back in 100,000 years, matching the interval between ice ages. But there are problems. For instance, the modest change in orbital eccentricity does not make nearly enough difference in sunlight reaching Earth to produce ice ages. Another problem is that some ice ages appear to have begun before the orbital changes that supposedly caused them.

Although adherents think that more research will explain such conflicts, Muller believes the textbook Milankovitch theory is hopelessly flawed. His own answer rests on a different aspect of Earth's orbit: Imagine the solar system is a vinyl record. Earth travels precisely on the record, called the ecliptic, only some of the time. At other times, the orbit is inclined a few degrees to the disk. Over a 100,000-year cycle, Earth's orbit begins in the ecliptic, rises out of it, then returns to where it started. This slow rocking, Muller proposes, is responsible for Earth's ice ages. He says the regions above and below the ecliptic are laden with cosmic dust, which cools the planet.

Muller's inclination theory got a shot in the arm in 1995, when Kenneth Farley, a geochemist at the California Institute of Technology, published a paper on cosmic dust found in sea sediments. He began the research expecting to give Muller's theory a knockout punch but discovered that cosmic dust levels do indeed wax and wane in sync with the ice ages.

But most researchers seem to echo the sentiment of Wallace Broecker, a geochemist at Columbia University, who thinks Muller is fooling himself. In 1996 Broecker brought a group of top-flight climate researchers together to hear Muller's theory. He says they found the presentation "riveting," but "they didn't buy it."

"There's no mechanism attached to the idea," states Nicholas J. Shackleton, a marine geologist at the University of Cambridge and a leading proponent of the Milankovitch theory. He questions how small changes in interplanetary dust could result in effects as dramatic as the coming and going of ice ages. Muller responds that dust from space influences cloud cover on Earth and could have profound climatic implications. He says his theory, if viewed objectively, does just as well at explaining the facts as Milankovitch's.

Referring to football, Muller calls himself a free safety of science, a generalist who scores intellectual touchdowns because he is unrestrained by questionable preconceived ideas. "Every once in a while there's a fumble" that no one notices, Muller says, "and I can grab that ball and run into the end zone." *—Daniel Grossman*

DANIEL GROSSMAN is a freelance writer based in Watertown, Mass.

COURTESY OF WIZARDS OF THE COAST 2000

Shrinking to Enormity

DNA microarrays are reshaping basic biology—but scientists fear they may soon drown in the data

A small start-up firm in Santa Clara, Calif., had a big idea five years ago. By adapting the methods of microprocessor manufacturing, it created microchips that contain thousands of distinct DNA probes on glass in place of transistors on silicon. The company figured that researchers would immediately find such "gene chips" useful, and doctors would eventually find them indispensable. With a chip, a tissue sample and a scanner, a technician can get a snapshot of the secret lives of the cells in that tissue, a detailed picture showing which genes are most active and which have been silenced. The idea that this might lead to customized preventive medical treatments was a compelling one for investors, who bid the stock of **Affymetrix** up 2,700 percent from July 1996 to March 2000.

Success like that attracts competition, and numerous companies now make several different kinds of DNA microarrays. All the chips work on the same principle: the glass is coated with a grid of tiny spots, 20 to 100 microns diameter; each spot contains millions of copies of a short sequence of DNA; and a computer keeps track of which DNA sequences are where. To make their snapshot, scientists extract from their sample cells messenger RNA (mRNA). Using enzymes, they make millions of copies of the mRNA molecules, tag them with fluorescent dye and break them up into short fragments. The tagged fragments are washed over the chip and, overnight, perform a remarkable feat of pattern matching, randomly bumping into the DNA probes fixed to the chip until they stick to one that contains a perfect genetic match. Although there are occasional mismatches, the millions of probes in each spot ensure that it lights up only if complementary mRNA is present. The brighter the spot fluoresces when scanned by a laser, the more mRNA of that kind was in the cell.

Affymetrix now makes more than 100,000 chips a year using light, masks and photosensitive chemicals to build DNA probes on chips one nucleotide at a time. **Agilent**, **Hitachi** and **Protogene Laboratories**, among others, use modified ink-jet printers, whose heads squirt *A, T, G* and *C* nucleotides instead of cyan, magenta, yellow and black inks. **Canon** is reportedly working with bubble jets to deposit DNA sequences, whereas **Corning**, **Motorola** and **Incyte Genomics** employ precision robots that place microdroplets of presynthesized sequences onto prepared slides. Although firms are spreading into almost every viable niche, none has yet submitted a medical diagnostic to the U.S. Food and Drug Administration for approval. Beyond the relatively straightforward obstacles—gene chip systems are still too expensive, for example, and few doctors know how to interpret their results—lies a much deeper question.

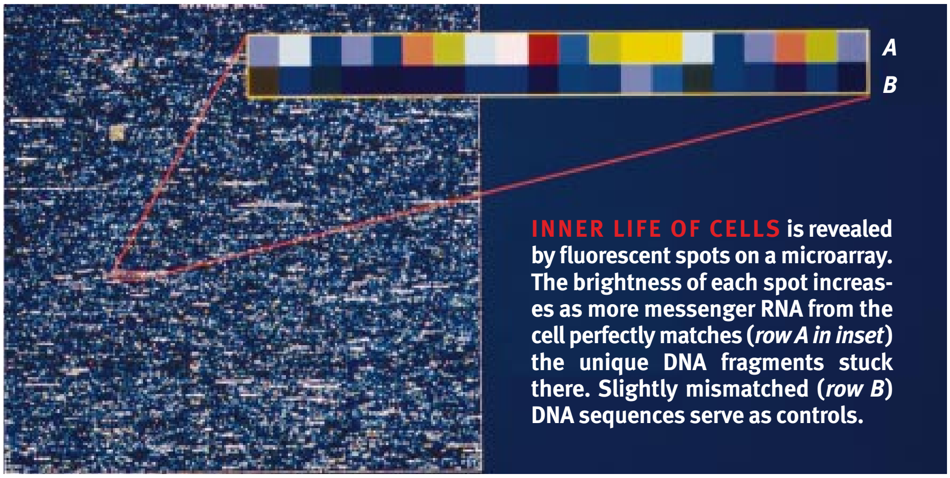
"Humans populations are outbred," remarks Lee Hartwell, director of the **Fred Hutchinson Cancer Research Center** in Seattle. Even well-understood genetic diseases involve myriad possible mutations; more than 1,000 have been linked to cystic fibrosis, for instance. An accurate diagnostic chip may have to include them all.

Although it may be many years before DNA microarrays find routine use by physicians, they have already begun to change experimental biology in profound ways. "They allow us to be vastly more productive—by a factor of 1,000 or so," says Richard A. Young of the **Whitehead Institute for Biomedical Research** of the Massachusetts Institute of Technology. In December 2000 his group reported that they had used microarrays for yeast to rediscover, in a matter of weeks, seven genes known to control a particular protein—research that originally took about 30 scientist-years to complete by conventional means. And the microarray experiments identified three additional genes that had been missed.

"The productivity boost is great," Young continues. "But what microarrays are really useful for is asking radically new questions about an entire system. At the moment, we understand how only half a dozen genes in any organism are regulated. If we knew the complete regulatory circuitry—how all genes are turned on or off and coordinate their activity with one another to deal with the environment—such a map would vastly increase our capacity to develop drugs for serious medical problems."

DNA CHIPS in handheld housings can sense the on/off state of up to 400,000 genes in a tissue sample.

COURTESY OF AFFYMETRIX, INC.



INNER LIFE OF CELLS is revealed by fluorescent spots on a microarray. The brightness of each spot increases as more messenger RNA from the cell perfectly matches (*row A in inset*) the unique DNA fragments stuck there. Slightly mismatched (*row B*) DNA sequences serve as controls.

COURTESY OF AFFYMETRIX, INC.

A team led by Timothy R. Hughes and Matthew J. Marton of **Rosetta Inpharmatics** in Kirkland, Wash., recently demonstrated one way to sketch such a map. Using some 700 chips, the scientists measured what happened to every gene in yeast cells when they were perturbed in 300 different ways: they deleted 279 genes and treated the cells with 13 different drugs. The study was able to work out the function of eight mysterious yeast genes, pinpoint the target of a common drug and even uncover a strong clue to a new human gene involved in cholesterol production. The project mined 10 million data points, in which more nuggets of knowledge undoubtedly remain.

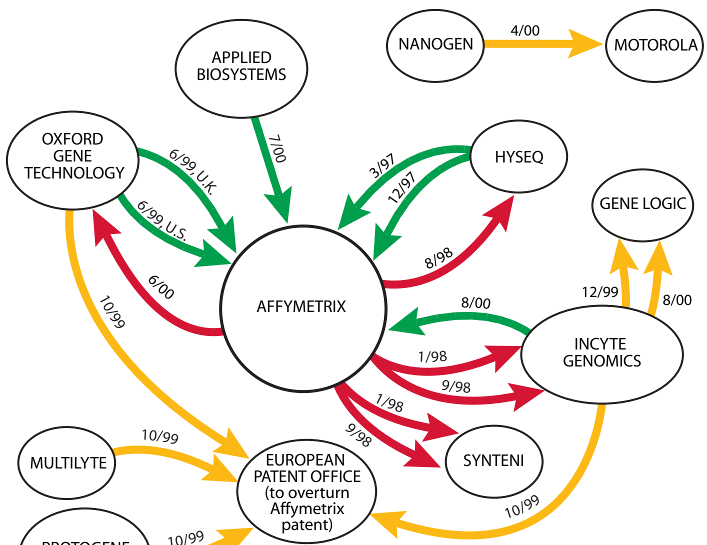
With each successive generation of microarray technology, the size of the probe spots shrinks, the number of genes per chip rises, and biologists' schemes for using the devices swell in grandeur. "We can now put over 60 million probes on a single glass wafer," Fodor says excitedly. He figures the entire human genome will fit on 200 to 300 wafers. And in fact, in September, Affymetrix spun off **Perlegen**, a subsidiary that plans to use microarrays to sequence, from scratch, the genomes contained in both chromosomes of 50 people to detect the subtle variations both within and among them. "In these patterns we will find the signature of human evolution. The potential for scientific discovery," Fodor boasts, "is fantastic."

So is the potential for confusion and error, Young and others caution. Hughes and Marton showed that genetic profiles are most powerful when compared with hundreds or thousands of others in a reference database. Such databases will be huge, because each profile contains about 50 megabytes of data. "How do we translate the data from an Affymetrix array to compare it with data from an array built by Corning?" Young asks. "It hasn't been done yet. And how do we encode the effects of one gene on another? It's all probabilistic, even though biologists tend to talk in terms of A *causing* B. We need a new mathematical language," he says. That may lead in turn, he suggests, to new theories that explain how the rich patterns of life arise from the complex chemistry of DNA.

—W. Wayt Gibbs

Patently Inefficient

A new industry is thrashed by waves of litigation



Who's Suing Whom
Plaintiff → Defendant
Litigation as of November 30, 2000

JOHNNY JOHNSON

Flipping through the quarterly report that Affymetrix issued last November, investors may have noticed a section entitled "The company may lose customers unless it improves its ability to manufacture its products and ensure their proper performance." Indeed, the firm took almost five years to address frequent complaints from researchers that it delivered chips that sometimes gave spurious results and often arrived months after they had been ordered. Fortunately for Affymetrix, until recently it had no real competitors to lose customers to, thanks largely to a formidable portfolio of issued and pending patents that now number more than 400, according to Stephen P. A. Fodor, its chief executive. "We have license agreements with 20 other companies," Fodor says. But he acknowledges that the licenses restrict those other firms to making arrays that have only about a tenth as many genetic probes as Affymetrix's gene chips do.

Other microarray producers responded in two ways: with lawsuits and with patents of their own on different microarray designs. Incyte Genomics, for example, uses robots to deposit up to 10,000 presynthesized genetic probes onto a glass slide. Motorola has prototypes of chips that hold the probes inside a thin slab of gel. But companies' aggressive patenting has led to a bewildering web of lawsuits (*above*)—and it may only get worse. "If we want to make a medical diagnostic with 40 genes on it, and 20 companies hold patents on those genes, we may have a big problem," says Nicholas J. Naclerio, head of Motorola's BioChip division. "It isn't at all clear how this is going to work out."

—W.W.G.

No E(asy) Cure

Electronic voting won't fix butterfly ballots, dimpled chads or W.'s presidency

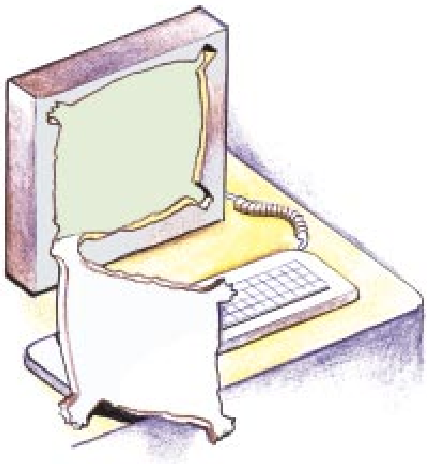
The lengthy machinations of last November's election proved one thing: that the instruments we have for measuring the people's will are not precise. Ballot recounting largely amounts to delving into statistical noise. What probably astonished most people was the sheer range of voting systems in the U.S. Certainly it was a source of astonishment in Britain, where people were genuinely puzzled about why Americans needed anything more complicated than the simple piece of paper and stub of pencil that British voters use to choose their members of Parliament.

It took only a few days before war was declared on chad and people started talking about electronic and online voting systems as the answer to everything. By that time the experts were already heading them off at the pass. Electronic and online voting systems are not going to provide perfect systems, basically because the specifications for elections are very complex. Voting machines (of whatever type) must be absolutely reliable: they must not invisibly lose or create votes. They must be easy for the broadest of audiences to use. They must be verifiably resistant to electoral fraud. They must protect voters' anonymity and privacy.

Talk to the people selling digital signatures, cryptographic products and online voting systems, and you'll be told that all these problems can be solved. The Italian Parliament, for example, votes via "smart cards." That's fine for a relatively small, educated group of professionals whose votes are a matter of public record anyway; for the 140-million-plus registered U.S. voters, whose choices are anonymous, you'd be looking at vast expense. Currently about 9 percent of them vote electronically.

In a posting to the RISKS Forum (http://catless.ncl.ac.uk/Risks/) on Election Day, Douglas W. Jones, chair of the Iowa State Board of Examiners for Voting Machines and Electronic Voting Systems, pointed out that although Federal Election Commission guidelines require that custom-built software be reviewed by an independent third party, "industry standard components" are acceptable without such a review. Increasingly, he wrote, custom voting software is being replaced with off-the-shelf, proprietary software. In other words: Windows. And who knows what's going on in there?

As Thomas went on to say, a dedicated individual out to fix an election—not now, perhaps, but in the future—might find himself a job within a relevant software company. He could seek to be assigned to the right group of programmers to allow him to modify code that when the right date came along could swing, say, 10 percent of the votes away from a specified party and distribute them in random amounts to other parties. In such a case, you wouldn't see anything as obvious as Palm Beach County's now famous anomalous blip for Pat Buchanan.



DAVID SUTER

Doctored software isn't the only risk. There are also power failures, bugs, hacker attacks and uncertainty whether the software inside the voting machine is the same software that was approved by the state. In Internet voting, there's the political issue of shifting the burden of supplying and maintaining the voting infrastructure from election officials to individual voters. Not to mention the fact that not everyone has access to the Internet. Even the argument of lowered costs is specious, says Rebecca Mercuri in the November 2000 *Communications of the ACM,* when you compare the costs of mailing out passwords and authenticating voters with the costs of today's well-understood absentee ballots. (Mercuri, a faculty member at Bryn Mawr College, successfully defended her doctoral dissertation on the perils of electronic voting last fall; when published, it is expected by some to be one of the most comprehensive contributions to the subject.)

"All the experts agree on some things," says Lorrie Cranor, a researcher at AT&T Labs Research who has written extensively on voting systems. "For example, that Internet voting is a huge can of worms, that there is no perfect system—all technology solutions are going to have problems—and that punch-card ballots are the worst thing we could have. The place where the experts don't all agree is if you get rid of punch cards, what do you replace them with?"

One suggestion has been direct-recording electronic (DRE) devices. Such machines, which register votes directly into a computer, have no audit trail, cannot be made rigorously bipartisan and may be expensive. Mercuri, for these reasons, is adamantly against DRE systems. Carnegie Mellon's Michael Shamos, on the other hand, has been saying for nearly a decade that it is naive to believe that mechanical and paper-based systems are more trustworthy than electronic ones.

Even so, in a local election in South Brunswick, N.J., an electronic machine was shown to have failed to record votes. In such a case, there's no ballot box to find in the back of a car and no way to restore the lost votes. Or rather, as the vendor told the newspaper there: machines don't lose votes; votes aren't cast. That kind of subtlety may be lost on voters.

Overall, it seems unlikely that electronic voting would fix the kind of problem that happened in Florida, where the margin for error in the voting systems was greater than the margin of victory. Of course, e-voting would have spared all those dedicated poll workers from hours of ballot checking, prevented the seemingly endless court battles over recounts and kept chad jokes at bay. But even if you could prove that electronic systems were the most reliable—doubtful, considering the Y2K bug—democracy is in part about perception and the reinforcement of trust. There is a comforting, ritual quality to that painstaking ballot counting and its close, bipartisan observance. In a narrow election decided wholly by electronic voting, there would be no comparable way to convince people that every vote really did count. *—Wendy M. Grossman*

WENDY M. GROSSMAN wrote From Anarchy to Power: The Net Comes of Age, *due out this month from NYU Press.*

Safeguarding

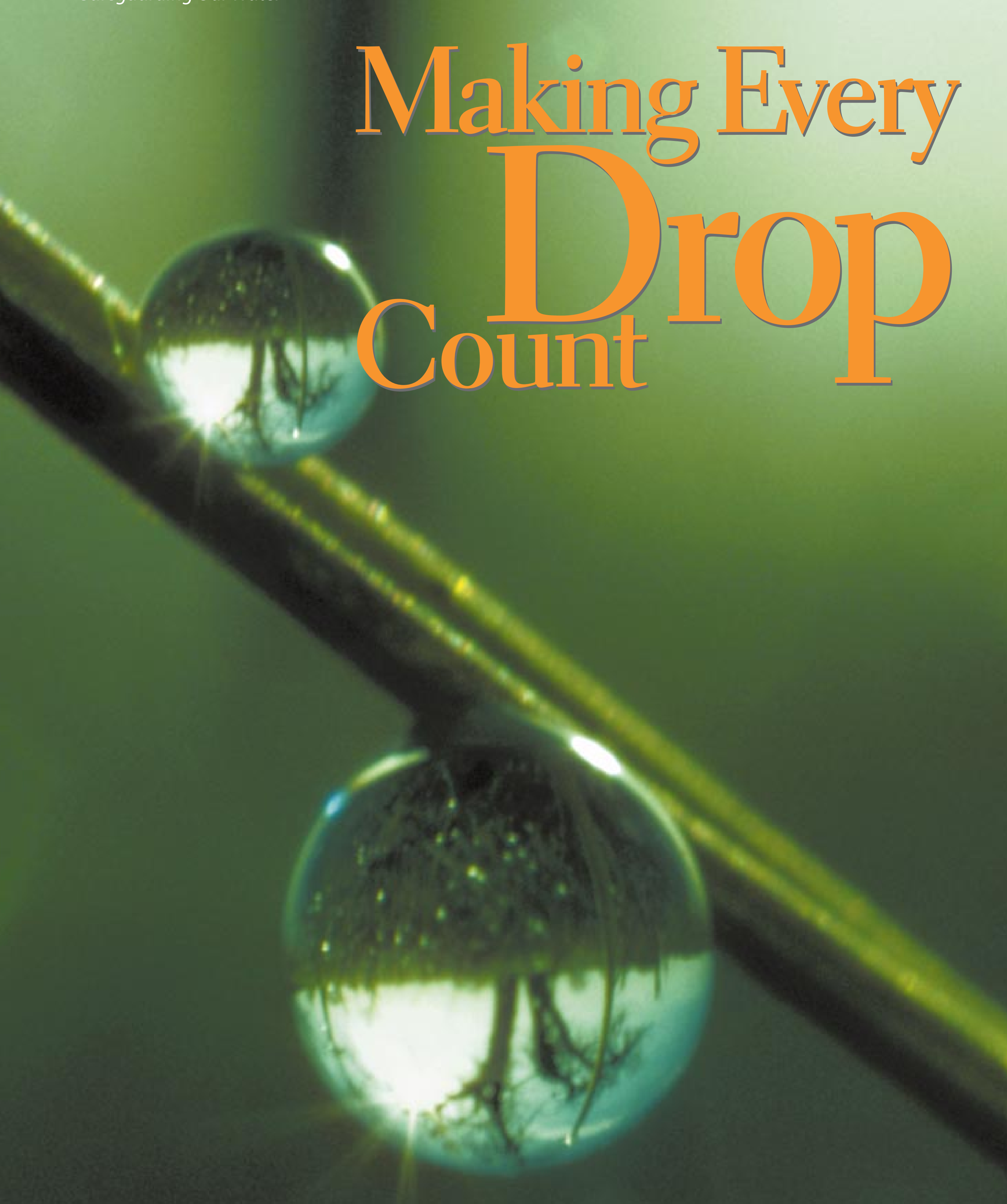
Our Water

Drip, trickle, splash. Water is one of the most common substances in the universe, and our ocean-wrapped planet is blessed with a generous share of it. Unfortunately, 97 percent of that share is salty, and much of the rest is locked up in ice. Obtaining an adequate supply of freshwater has consequently been the focus of human ingenuity and passions throughout history. Water has been the prize (and sometimes the weapon) in conflicts around the world. Even in the century ahead, impressive gains in technological capabilities to find, transport and conserve freshwater may not be able to accommodate increasing demand, particularly in the developing world. Local mismatches between need and supply could push groups to violence, retard economic progress and devastate populations.

In the following pages, Peter H. Gleick of the Pacific Institute for Studies in Development, Environment and Security describes the magnitude of the world's pressing water problems in terms of skyrocketing usage and ominous limits to the known supplies. Sandra Postel of the Global Water Policy Project then narrows the discussion to irrigation, the single largest use for freshwater, and to the prospects for improving this vital agricultural technology. Lest anyone think that other options for staving off water shortages are lacking, we also consider a quartet of other approaches, including desalination, "bag and drag" transport, recycling and increased plumbing efficiency. A water crisis may be in the cards for some, but not if we act quickly to develop all the solutions at our disposal. *—The Editors*

PAUL EDMONDSON *Stone*

Making Every Drop Count



RICHARD HAMILTON SMITH *Corbis*

We drink it, we generate electricity with it, we soak our crops with it. And we're stretching our supplies to the breaking point. Will we have enough clean water to satisfy all the world's needs?

by Peter H. Gleick

The history of human civilization is entwined with the history of the ways we have learned to manipulate water resources. The earliest agricultural communities emerged where crops could be cultivated with dependable rainfall and perennial rivers. Simple irrigation canals permitted greater crop production and longer growing seasons in dry areas. Five thousand years ago settlements in the Indus Valley were built with pipes for water supply and ditches for wastewater. Athens and Pompeii, like most Greco-Roman towns of their time, maintained elaborate systems for water supply and drainage.

As towns gradually expanded, water was brought from increasingly remote sources, leading to sophisticated engineering efforts, such as dams and aqueducts. At the height of the Roman Empire, nine major systems, with an innovative layout of pipes and well-built sewers, supplied the occupants of Rome with as much water per person as is provided in many parts of the industrial world today.

During the industrial revolution and population explosion of the 19th and 20th centuries, the demand for water rose dramatically. Unprecedented construction of tens of thousands of monumental engineering projects designed to control floods, protect clean water supplies, and provide water for irrigation and hydropower brought great benefits to hundreds of millions of people. Thanks to improved sewer systems, water-related diseases such as cholera and typhoid, once endemic throughout the world, have largely been conquered in the more industrial nations. Vast cities, incapable of surviving on their local resources, have bloomed in the desert with water brought from hundreds and even thousands of miles away. Food production has kept pace with soaring populations mainly because of the expansion of artificial irrigation systems that make possible the growth of 40 percent of the world's food. Nearly one fifth of all the electricity generated worldwide is produced by turbines spun by the power of falling water.

Yet there is a dark side to this picture: despite our progress, half of the world's population still suffers with water services inferior to those available to the ancient Greeks and Romans. As the latest United Nations report on access to water reiterated in November of last year, more than one billion people lack access to clean drinking water; some two and a half billion do not have adequate sanitation services. Preventable water-related diseases kill an estimated 10,000 to 20,000 children every day, and the latest evidence suggests that we are falling behind in efforts to solve these problems. Massive cholera outbreaks appeared in the mid-1990s in Latin America, Africa and Asia. Millions of people in Bangladesh and India drink water contaminated with arsenic. And the surging populations throughout the developing world are intensifying the pressures on limited water supplies.

The effects of our water policies extend beyond jeopardizing human health. Tens of millions of people have been forced to move from their homes—often with little warning or compensation—to make way for the reservoirs behind dams. More than 20 percent of all freshwater fish species are now threatened or endangered because dams and water withdrawals have destroyed the free-flowing river ecosystems where they thrive. Certain irrigation practices degrade soil quality and reduce agricultural productivity, heralding a premature end to the green revolution. Groundwater aquifers are being pumped down faster than they are naturally replenished in parts of India, China, the U.S. and elsewhere. And disputes over shared water resources have led to violence and continue to raise local, national and even international tensions [*see box on page 44*].

At the outset of the new millennium, however, the way resource planners think about water is beginning to change. The focus is slowly shifting back to the provision of basic human and environ-

The Author

PETER H. GLEICK is director of the Pacific Institute for Studies in Development, Environment and Security, a nonprofit policy research think tank based in Oakland, Calif. Gleick co-founded the institute in 1987. He is considered one of the world's leading experts on freshwater problems, including sustainable use of water, water as it relates to climate change, and conflicts over shared water resources.

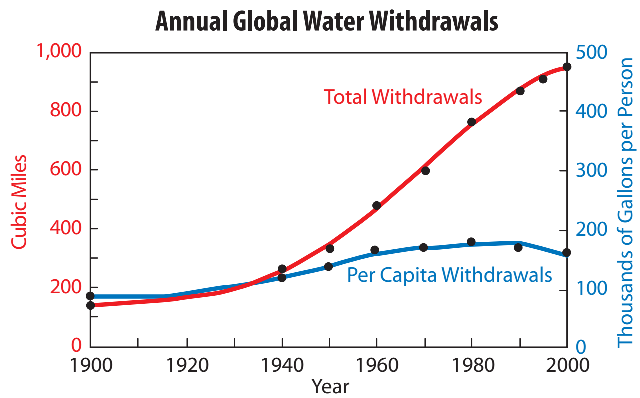
mental needs as the top priority—ensuring "some for all, instead of more for some," as put by Kader Asmal, former minister for water affairs and forestry in South Africa. To accomplish these goals and meet the demands of booming populations, some water experts now call for using existing infrastructure in smarter ways rather than building new facilities, which is increasingly considered the option of last, not first, resort. The challenges we face are to use the water we have more efficiently, to rethink our priorities for water use and to identify alternative supplies of this precious resource.

This shift in philosophy has not been universally accepted, and it comes with strong opposition from some established water organizations. Nevertheless, it may be the only way to address successfully the pressing problems of providing everyone with clean water to drink, adequate water to grow food and a life free from preventable water-related illness. History shows that although access to clean drinking water and sanitation services cannot guarantee the survival of a civilization, civilizations most certainly cannot prosper without them.

Where the Water Will Be in 2025

The total amount of water withdrawn globally from rivers, underground aquifers and other sources has increased ninefold since 1900 (*chart*). Water use per person has only doubled in that time, however, and it has even declined slightly in recent years. Despite this positive trend, some experts worry that improvements in water-use efficiency will fail to keep pace with projected population growth. Estimated annual water availability per person in 2025 (*map*) reveals that at least 40 percent of the world's 7.2 billion people may face serious problems with agriculture, industry or human health if they must rely solely on natural endowments of freshwater. Severe water shortages could also strike particular regions of water-rich countries, such as the U.S. and China.

People's access to water also depends on factors not reflected here, such as political and economic conditions, changing climate patterns and available technology. —*P.H.G.*



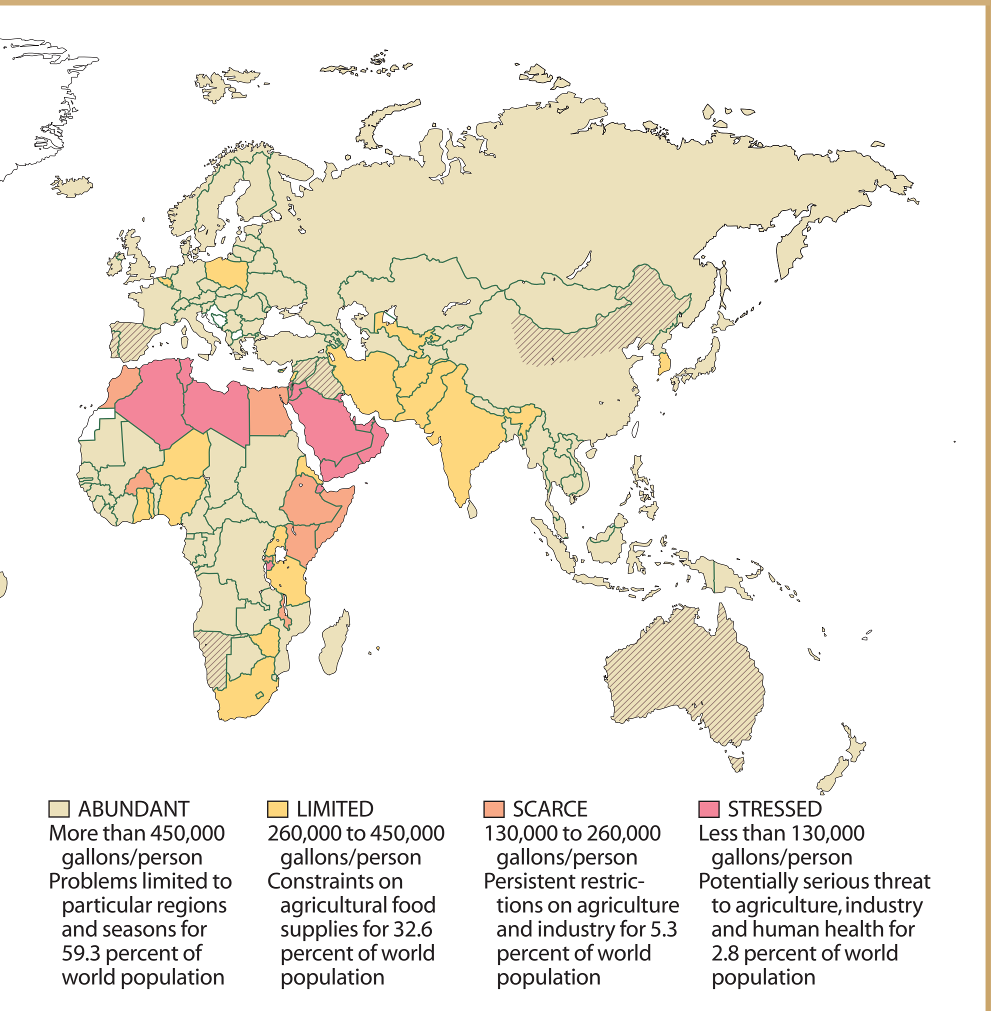
Damage from Dams

Over the past 100 years, humankind has designed networks of canals, dams and reservoirs so extensive that the resulting redistribution of freshwater from one place to another and from one season to the next accounts for a small but measurable change in the wobble of the earth as it spins. The statistics are staggering. Before 1900 only 40 reservoirs had been built with storage volumes greater than 25 billion gallons; today almost 3,000 reservoirs larger than this inundate 120 million acres of land and hold more than 1,500 cubic miles of water—as much as Lake Michigan and Lake Ontario combined. The more than 70,000 dams in the U.S. are capable of capturing and storing half of the annual river flow of the entire country.

In many nations, big dams and reservoirs were originally considered vital for national security, economic prosperity and agricultural survival. Until the late 1970s and early 1980s, few people took into account the environmental consequences of these massive projects. Today, however, the results are clear: dams have destroyed the ecosystems in and around countless rivers, lakes and streams. On the Columbia and Snake rivers in the northwestern U.S., 95 percent of the juvenile salmon trying to reach the ocean do not survive passage through the numerous dams and reservoirs that block their way. More than 900 dams on New England and European rivers block Atlantic salmon from their spawning grounds, and their populations have fallen to less than 1 percent of historical levels. Perhaps most infamously, the Aral Sea in central Asia is disappearing because water from the Amu Darya and Syr Darya rivers that once sustained it has been diverted to irrigate cotton. Twenty-four species of fish formerly found only in that sea are currently thought to be extinct.

As environmental awareness has heightened globally, the desire to protect—and even restore—some of these natural resources has grown. The earliest environmental advocacy groups in the U.S. mobilized against dams proposed in places such as Yosemite National Park in California and the Grand Canyon in Arizona. In the 1970s plans in the former Soviet Union to divert the flow of Siberian rivers away from the Arctic stimulated an unprecedented public outcry, helping to halt the projects. In many developing countries, grassroots opposition to the environmental and social costs of big water projects is becoming more and more effective. Villagers and community activists in India have encouraged a public debate over major dams. In China, where open disagreement with government policies is strong-

The water lost from Mexico City's leaky supply system is enough



JOHNNY JOHNSON; SOURCE: PETER H. GLEICK

ly discouraged, protest against the monumental Three Gorges Project has been unusually vocal and persistent.

Until very recently, international financial organizations such as the World Bank, export-import banks and multilateral aid agencies subsidized or paid in full for dams or other water-related civil engineering projects—which often have price tags in the tens of billions of dollars. These organizations are slowly beginning to reduce or eliminate such subsidies, putting more of the financial burden on already strained national economies. Having seen so much ineffective development in the past—and having borne the associated costs (both monetary and otherwise) of that development—many governments are unwilling to pay for new structures to solve water shortages and other problems.

to meet the needs of a city the size of Rome

A handful of countries are even taking steps to remove some of the most egregious and damaging dams. For example, in 1998 and 1999 the Maisons-Rouges and Saint-Etienne-du-Vigan dams in the Loire River basin in France were demolished to help restore fisheries in the region. In 1999 the Edwards Dam, which was built in 1837 on the Kennebec River in Maine, was dismantled to open up an 18-mile stretch of the river for fish spawning; within months Atlantic salmon, American shad, river herring, striped bass, shortnose sturgeon, Atlantic sturgeon, rainbow smelt and American eel had returned to the upper parts of the river. Altogether around 500 old, dangerous or environmentally harmful dams have been removed from U.S. rivers in the past few years.

Fortunately—and unexpectedly—the demand for water is not rising as rapidly as some predicted. As a result, the pressure to build new water infrastructures has diminished over the past two decades. Although population, industrial output and economic productivity have continued to soar in developed nations, the rate at which people withdraw water from aquifers, rivers and lakes has slowed. And in a few parts of the world, demand has actually fallen.

Demand Is Down—But for How Long?

What explains this remarkable turn of events? Two factors: people have figured out how to use water more efficiently, and communities are rethinking their priorities for water use. Throughout the first three quarters of the 20th century, the quantity of freshwater consumed per person doubled on average; in the U.S., water withdrawals increased 10-fold while the population quadrupled. But since 1980 the amount of water consumed per person has actually decreased, thanks to a range of new technologies that help to conserve water in homes and industry. In 1965, for instance, Japan used approximately 13 million gallons of water to produce $1 million of commercial output; by 1989 this had dropped to 3.5 million gallons (even accounting for inflation)—almost a quadrupling of water productivity. In the U.S., water withdrawals have fallen by more than 20 percent from their peak in 1980.

As the world's population continues to grow, dams, aqueducts and other kinds of infrastructure will still have to be built, particularly in developing countries where basic human needs have not been met. But such projects must be built to higher standards and with more accountability to local people and their environment than in the past. And even in regions where new projects seem warranted, we must find ways to meet demands with fewer resources, minimum ecological disruption and less money.

The fastest and cheapest solution is to expand the productive and efficient use of water. In many countries, 30 percent or more of the domestic water supply never reaches its intended destinations, disappearing from leaky pipes, faulty equipment or poorly maintained distribution systems. The quantity of water that Mexico City's supply system loses is enough to meet the needs of a city the size of Rome, according to recent estimates. Even in more modern systems, losses of 10 to 20 percent are common.

When water does reach consumers, it is often used wastefully. In homes, most water is literally flushed away. Before 1990 most toilets in the U.S. drew about six gallons of water for each flush. In 1992 the U.S. Congress passed a national standard mandating that all new residential toilets be low-flow models that require only 1.6 gallons per flush—a 70 percent improvement with a single change in technology. It will take time to replace all older toilets with the newer, better ones. A number of cities, however, have found the water conservation made possible by the new technology to be so significant—and the cost of saving that water to be so low—that they have established programs to speed up the transition to low-flow toilets [see "Leaking Away," by Diane Martindale, on page 54].

Even in the developing world, technologies such as more efficient toilets have a role to play. Because of the difficulty of finding new water sources for Mexico City, city officials launched a water conservation program that involved replacing 350,000 old toilets. The replacements have already saved enough water to supply an additional 250,000 residents. And numerous other options for both industrial and nonindustrial nations are available as well, including better leak detection, less wasteful washing machines, drip irrigation and water-conserving plants in outdoor landscaping.

The amount of water needed for industrial applications depends on two factors: the mix of goods and services demanded by society and the processes chosen to generate them. For instance, producing a ton of steel before World War II required 60 to 100 tons of water. Current technology can make a ton of steel with less than six tons of water. Replacing old technology with new techniques reduces water needs by a factor of 10. Producing a ton of aluminum, however, requires only one and a half tons of water. Replacing the use of steel with aluminum, as has been happening for years in the automobile industry, can further lower water use. And telecommuting from home can save the hundreds of gallons of water required to produce, deliver and sell a gallon of gasoline, even accounting for the water required to manufacture our computers.

The largest single consumer of water is agriculture—and this use is largely inefficient. Water is lost as it is distributed to farmers and applied to crops. Consequently, as much as half of all water diverted for agriculture never yields any food. Thus, even modest improvements in agricultural efficiency could free up huge quantities of water [see "Growing More Food with Less Water," by Sandra Postel, on page 46]. Growing tomatoes with traditional irrigation systems may require 40 percent more water than growing tomatoes with drip systems. Even our diets have an effect on our overall water needs. Growing a pound of corn can take between 100 and 250 gallons of water, depending on soil and climate conditions and irrigation methods. But growing the grain to produce a pound of beef can require between 2,000 and 8,500 gallons. We can conserve water not only by altering how we choose to grow our food but also by changing what we choose to eat.

Shifting where people use water can also lead to tremendous gains in effi-

Why should we raise all water to drinkable standards and then use it to flush toilets?

Continuing Conflict over Freshwater

Myths, legends and written histories reveal repeated controversy over freshwater resources since ancient times. Scrolls from Mesopotamia, for instance, indicate that the states of Umma and Lagash in the Middle East clashed over the control of irrigation canals some 4,500 years ago.

Throughout history, water has been used as a military and political goal, as a weapon of war and even as a military target. But disagreements most often arise from the fact that water resources are not neatly partitioned by the arbitrary political borders set by governments. Today nearly half of the land area of the world lies within international river basins, and the watersheds of 261 major rivers are shared by two or more countries. Overlapping claims to water resources have often provoked disputes, and in recent years local and regional conflicts have escalated over inequitable allocation and use of water resources.

A small sampling of water conflicts that occurred in the 20th century demonstrates that treaties and other international diplomacy can sometimes encourage opposing countries to cooperate—but not always before blood is shed. The risk of future strife cannot be ignored: disputes over water will become more common over the next several decades as competition for this scarce resource intensifies. —*P.H.G.*

AP PHOTO

WATER SUPPLY LINE in Novi Sad, Yugoslavia, was destroyed along with this Danube River bridge during a NATO airstrike in April 1999.

U.S. 1924

Local farmers dynamite the Los Angeles aqueduct several times in an attempt to prevent diversions of water from the Owens Valley to Los Angeles.

India and Pakistan 1947 to 1960

Partitioning of British India awkwardly divides the waters of the Indus River valley between India and Pakistan. Competition over irrigation supplies incites numerous conflicts between the two nations; in one case, India stems the flow of water into Pakistani irrigation canals. After 12 years of World Bank–led negotiations, a 1960 treaty helps to resolve the discord.

Egypt and Sudan 1958

Egypt sends troops into contested territory between the two nations during sensitive negotiations concerning regional politics and water from the Nile. Signing of a Nile waters treaty in 1959 eases tensions.

Israel, Jordan and Syria 1960s and 1970s
Clashes over allocation, control and diversion of the Yarmouk and Jordan rivers continue to the present day.

South Africa 1990
A pro-apartheid council cuts off water to 50,000 black residents of Wesselton Township after protests against wretched sanitation and living conditions.

Iraq 1991
During the Persian Gulf War, Iraq destroys desalination plants in Kuwait. A United Nations coalition considers using the Ataturk Dam in Turkey to shut off the water flow of the Euphrates River to Iraq.

India 1991 to present
An estimated 50 people die in violence that continues to erupt between the Indian states of Karnataka and Tamil Nadu over the allocation of irrigation water from the Cauvery River, which flows from one state into the other.

Yugoslavia 1999
NATO shuts down water supplies in Belgrade and bombs bridges on the Danube River, disrupting navigation.

A comprehensive chronology of water-related conflicts can be found at www.worldwater.org/conflictIntro.htm

ciency. Supporting 100,000 high-tech California jobs requires some 250 million gallons of water a year; the same amount of water used in the agricultural sector sustains fewer than 10 jobs—a stunning difference. Similar figures apply in many other countries. Ultimately these disparities will lead to more and more pressure to transfer water from agricultural uses to other economic sectors. Unless the agricultural community embraces water conservation efforts, conflicts between farmers and urban water users will worsen.

The idea that a planet with a surface covered mostly by water could be facing a water shortage seems incredible. Yet 97 percent of the world's water is too salty for human consumption or crops, and much of the rest is out of reach in deep groundwater or in glaciers and ice caps. Not surprisingly, researchers have investigated techniques for dipping into the immense supply of water in the oceans. The technology to desalinate brackish water or saltwater is well developed, but it remains expensive and is currently an option only in wealthy but dry areas near the coast. Some regions, such as the Arabian Gulf, are highly dependent on desalination, but the process remains a minor contributor to overall water supplies, providing less than 0.2 percent of global withdrawals [see "Sweating the Small Stuff," by Diane Martindale, on page 52].

With the process of converting saltwater to freshwater so expensive, some companies have turned to another possibility: moving clean water in ships or even giant plastic bags from regions with an abundance of the resource to those places around the globe suffering from a lack of water [see "Bagged and Dragged," by Peter H. Gleick, on page 53]. But this approach, too, may have serious economic and political constraints.

Rather than seeking new distant sources of water, smart planners are beginning to explore using alternative *kinds* of water to meet certain needs. Why should communities raise all water to drinkable standards and then use that expensive resource for flushing toilets or watering lawns? Most water ends up flowing down the drain after a single use, and developed countries spend billions of dollars to collect and treat this wastewater before dumping it into a river or the ocean. Meanwhile, in poorer countries, this water is often simply returned untreated to a river or lake where it may pose a threat to human health or the environment. Recently attention has begun to focus on reclaiming and reusing this water.

Wastewater can be treated to different levels suitable for use in a variety of applications, such as recharging groundwater aquifers, supplying industrial processes, irrigating certain crops or even augmenting potable supplies. In Windhoek, Namibia, for instance, residents have used treated wastewater since 1968 to supplement the city's potable water supply; in drought years, such water has constituted up to 30 percent of Windhoek's drinking water supply [see "Waste Not, Want Not," by Diane Martindale, on page 55]. Seventy percent of Israeli municipal wastewater is treated and reused, mainly for agricultural irrigation of nonfood crops. Efforts to capture, treat and reuse more wastewater are also under way in neighboring Jordan. By the mid-1990s residents of California relied on more than 160 billion gallons of reclaimed water annually for irrigating landscapes, golf courses and crops, recharging groundwater aquifers, supplying industrial processes and even flushing toilets.

New approaches to meeting water needs will not be easy to implement: economic and institutional structures still encourage the wasting of water and the destruction of ecosystems. Among the barriers to better water planning and use are inappropriately low water prices, inadequate information on new efficiency technologies, inequitable water allocations, and government subsidies for growing water-intensive crops in arid regions or building dams.

Part of the difficulty, however, also lies in the prevalence of old ideas among water planners. Addressing the world's basic water problems requires fundamental changes in how we think about water, and such changes are coming about slowly. Rather than trying endlessly to find enough water to meet hazy projections of future desires, it is time to find a way to meet our present and future needs with the water that is already available, while preserving the ecological cycles that are so integral to human well-being. SA

Further Information

THE WORLD'S WATER 1998–1999. Peter H. Gleick. Island Press, 1998.
INTERNATIONAL RIVER BASINS OF THE WORLD. Aaron T. Wolf et al. in *Water Resources Development*, Vol. 15, No. 4, pages 387–427; December 1999.
THE WORLD'S WATER 2000–2001. Peter H. Gleick. Island Press, 2000.
Information on the world's water resources can be found at www.worldwater.org
United Nations Environment Program Global Environment Monitoring System's Freshwater Quality Program can be found at www.cciw.ca/gems/
VISION 21: A SHARED VISION FOR HYGIENE, SANITATION AND WATER SUPPLY. Water Supply and Sanitation Collaborative Council. Available at www.wsscc.org/vision21/docs/index.html

Growing more Food with less Water

by Sandra Postel

DAVID BARNES Stone

Six thousand years ago farmers in Mesopotamia dug a ditch to divert water from the Euphrates River. With that successful effort to satisfy their thirsty crops, they went on to form the world's first irrigation-based civilization. This story of the ancient Sumerians is well known. What is not so well known is that Sumeria was one of the earliest civilizations to crumble in part because of the consequences of irrigation.

Sumerian farmers harvested plentiful wheat and barley crops for some 2,000 years thanks to the extra water brought in from the river, but the soil eventually succumbed to salinization—the toxic buildup of salts and other impurities left behind when water evaporates. Many historians argue that the poisoned soil, which could not support sufficient food production, figured prominently in the society's decline.

Far more people depend on irrigation in the modern world than did in ancient Sumeria. About 40 percent of the world's food now grows in irrigated soils, which make up 18 percent of global cropland [*see illustration on page 50*]. Farmers who irrigate can typically reap two or three harvests every year *and* get higher crop yields. As a result, the spread of irrigation has been a key factor behind the near tripling of global grain production since 1950. Done correctly, irrigation will continue to play a leading role in feeding the world, but as history shows, dependence on irrigated agriculture also entails significant risks.

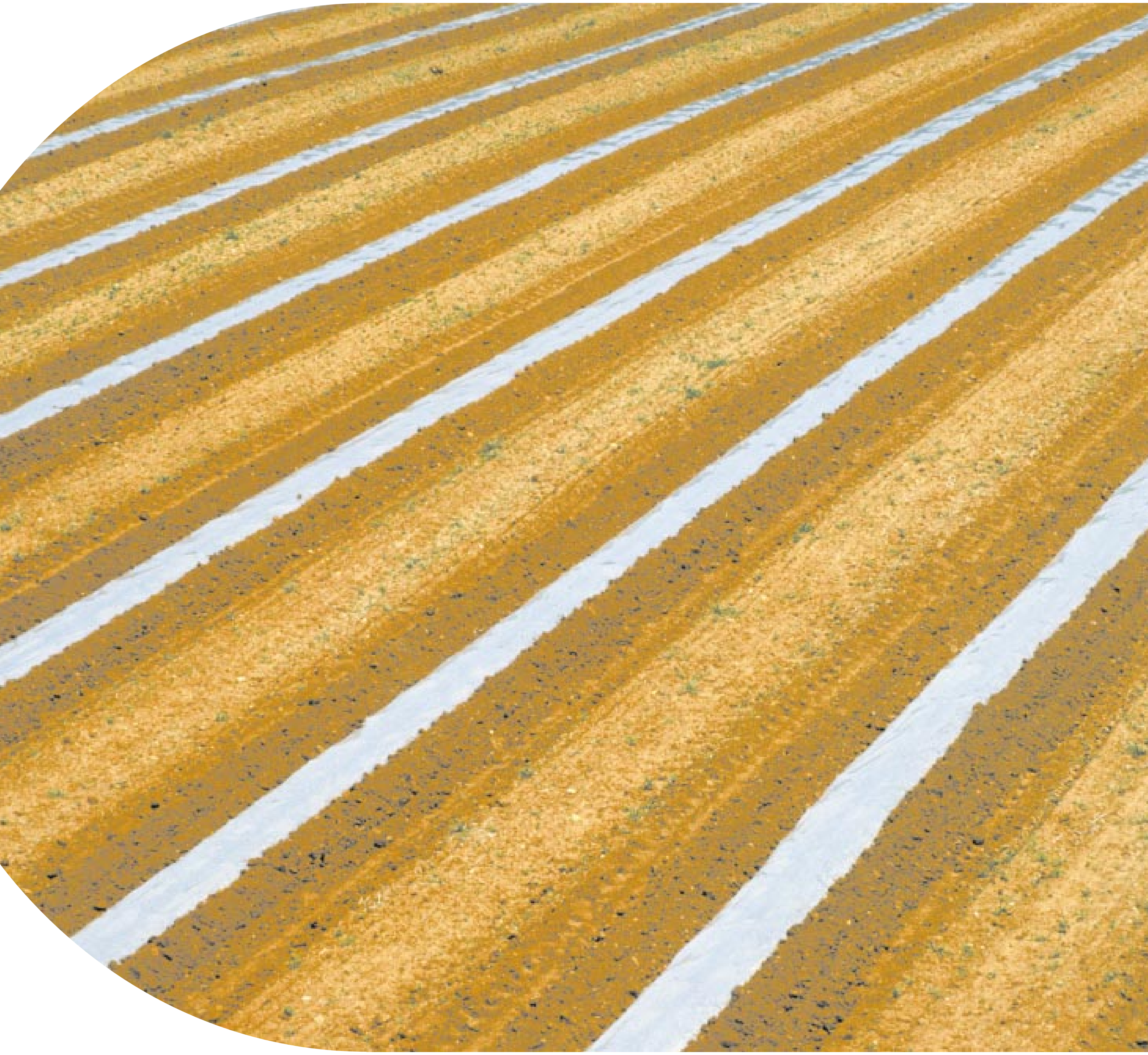
Today irrigation accounts for two thirds of water use worldwide and as much as 90 percent in many developing countries. Meeting the crop demands projected for 2025, when the planet's population is expected to reach eight billion, could require an additional 192 cubic miles of water—a volume nearly equivalent to the annual flow of the Nile 10 times over. No one yet knows how to supply that much additional water in a way that protects supplies for future use.

Severe water scarcity presents the single biggest threat to future food production. Even now many freshwater sources—underground aquifers and rivers—are stressed beyond their limits. As much as 8 percent of food crops grows on farms that use groundwater faster than the aquifers are replenished, and many large rivers are so heavily diverted that they don't reach the sea for much of the year. As the number of urban dwellers climbs to five billion by 2025, farmers will have to compete even more aggressively with cities and industry for shrinking resources.

The Author

SANDRA POSTEL directs the Global Water Policy Project in Amherst, Mass., and is a visiting senior lecturer in environmental studies at Mount Holyoke College. She is also a senior fellow of the Worldwatch Institute, where she served as vice president for research from 1988 to 1994.

If the world hopes to feed its burgeoning population, irrigation must become less wasteful and more widespread



FLOODING CROP FURROWS is a traditional but often wasteful irrigation method. Much of the water soaks into the ground or evaporates without assisting the plants.

Despite these challenges, agricultural specialists are counting on irrigated land to produce most of the additional food that will be needed worldwide. Better management of soil and water, along with creative cropping patterns, can boost production from cropland that is watered only by rainfall, but the heaviest burden will fall on irrigated land. To fulfill its potential, irrigated agriculture requires a thorough redesign organized around two primary goals: cut water demands of mainstream agriculture and bring low-cost irrigation to poor farmers.

Fortunately, a great deal of room exists for improving the productivity of water used in agriculture. A first line of attack is to increase irrigation efficiency. At present, most farmers irrigate their crops by flooding their fields or channeling the water down parallel furrows, relying on gravity to move the water across the land. The plants absorb only a small fraction of the water; the rest drains into rivers or aquifers, or evaporates. In many locations this practice not only wastes and pollutes water but also degrades the land through erosion, waterlogging and salinization. More efficient and environmentally sound technologies exist that could reduce water demand on farms by up to 50 percent.

Drip systems rank high among irrigation technologies with significant untapped potential. Unlike flooding techniques, drip systems enable farmers to deliver water directly to the plants' roots drop by drop, nearly eliminating waste. The water travels at low pressure through a network of perforated plastic tubing installed on or below the surface of the soil, and it emerges through small holes at a slow but steady pace. Because the plants enjoy an ideal moisture environment, drip irrigation usually offers the added bonus of higher crop yields. Studies in India, Israel, Jordan, Spain and the U.S. have shown time and again that drip irrigation reduces water use by 30 to 70 percent and increases crop yield by 20 to 90 percent compared with flooding methods.

Sprinklers can perform almost as well as drip methods when they are designed properly. Traditional high-pressure irrigation sprinklers spray water high into the air to cover as large a land area as possible. The problem is that the more time the water spends in the air, the more of it evaporates and blows off course before reaching the plants. In contrast, new low-energy sprinklers deliver water in small doses through nozzles positioned just above the ground. Numerous farmers in Texas who have installed such sprinklers have found that their plants absorb 90 to 95 percent of the water that leaves the sprinkler nozzle.

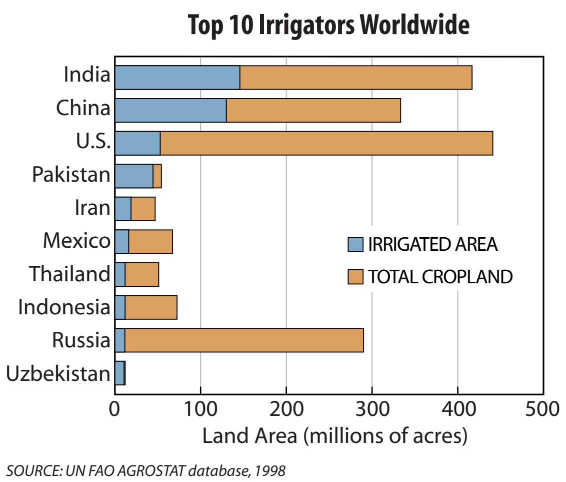
Despite these impressive payoffs, sprinklers service only 10 to 15 percent of the world's irrigated fields, and drip systems account for just over 1 percent. The higher costs of these technologies (relative to simple flooding methods) have been a barrier to their spread, but so has the prevalence of national water policies that discourage rather than foster efficient water use. Many governments have set very low prices for publicly supplied irrigation, leaving farmers with little motivation to invest in ways to conserve water or to improve efficiency. Most authorities have also failed to regulate groundwater pumping, even in regions where aquifers are overtapped. Farmers might be inclined to conserve their own water supplies if they could profit from selling the surplus, but a number of countries prohibit or discourage this practice.

Efforts aside from irrigation technologies can also help reduce agricultural demand for water. Much potential lies in scheduling the timing of irrigation to more precisely match plants' water needs. Measurements of climate factors such as temperature and precipitation can be fed into a computer that calculates how much water a typical plant is consuming. Farmers can use this figure to determine, quite accurately, when and how much to irrigate their particular crops throughout the growing season. A 1995 survey conducted by the University of California at Berkeley found that, on average, farmers in California who used this tool reduced water use by 13 percent and achieved an 8 percent increase in yield—a big gain in water productivity.

An obvious way to get more benefit out of water is to use it more than once. Some communities use recycled wastewater [see "Waste Not, Want Not," by Diane Martindale, on page 55]. Treated wastewater accounts for 30 percent of Israel's agricultural water supply, for instance, and this share is expected to climb to 80 percent by 2025. Developing new crop varieties offers potential as well. In the quest for higher yields, scientists have already exploited many of the most fruitful agronomic options for growing more food with the same amount of water. The hybrid wheat and rice varieties that spawned the green revolution, for example, were bred to allocate more of the plants' energy—and thus their water uptake—into edible grain. The widespread adoption of high-yielding and early-maturing rice varieties has led to a roughly threefold increase in the amount of rice harvested per unit of water consumed—a tremendous achievement. No strategy in sight—neither conventional breeding techniques nor genetic engineering—could repeat those gains on such a grand scale, but modest improvements are likely.

Yet another way to do more with less water is to reconfigure our diets. The typical North American diet, with its large share of animal products, requires twice as much water to produce as the less meat-intensive diets common in many Asian and some European countries. Eating lower on the food chain could allow the same volume of water to feed

Technologies exist that could reduce water demand on farms by up to 50 percent



JUST FOUR COUNTRIES account for half the world's 670 million acres of irrigated cropland.

JOHNNY JOHNSON

two Americans instead of one, with no loss in overall nutrition.

Reducing the water demands of mainstream agriculture is critical, but irrigation will never reach its potential to alleviate rural hunger and poverty without additional efforts. Among the world's approximately 800 million undernourished people are millions of poor farm families who could benefit dramatically from access to irrigation water or to technologies that enable them to use local water more productively.

Most of these people live in Asia and Africa, where long dry seasons make crop production difficult or impossible without irrigation. For them, conventional irrigation technologies are too expensive for their small plots, which typically encompass fewer than five acres. Even the least expensive motorized pumps that are made for tapping groundwater cost about $350, far out of reach for farmers earning barely that much in a year. Where affordable irrigation technologies have been made available, however, they have proved remarkably successful.

I traveled to Bangladesh in 1998 to see one of these successes firsthand. Torrential rains drench Bangladesh during the monsoon months, but the country receives very little precipitation the rest of the year. Many fields lie fallow during the dry season, even though groundwater lies less than 20 feet below the surface. Over the past 17 years a foot-operated device called a treadle pump has transformed much of this land into productive, year-round farms.

To an affluent Westerner, this pump resembles a StairMaster exercise machine and is operated in much the same way. The user pedals up and down on two long bamboo poles, or treadles, which in turn activate two steel cylinders. Suction pulls groundwater into the cylinders and then dispenses it into a channel in the field. Families I spoke with said they often treadled four to six hours a day to irrigate their rice paddies and vegetable plots. But the hard work paid off: not only were they no longer hungry during the dry season, but they had surplus vegetables to take to market.

Costing less than $35, the treadle pump has increased the average net income for these farmers—which is often as little as a dollar a day—by $100 a year. To date, Bangladeshi farmers have purchased some 1.2 million treadle pumps, raising the productivity of more than 600,000 acres of farmland. Manufactured and marketed locally, the pumps are injecting at least an additional $350 million a year into the Bangladeshi economy.

In other impoverished and water-scarce regions, poor farmers are reaping the benefits of newly designed low-cost drip and sprinkler systems. Beginning with a $5 bucket kit for home gardens, a spectrum of drip systems keyed to different income levels and farm sizes is now enabling farmers with limited access to water to irrigate their land efficiently. In 1998 I spoke with farmers in the lower Himalayas of northern India, where crops are grown on terraces and irrigated with a scarce communal water supply. They expected to double their planted area with the increased efficiency brought about by affordable drip systems.

Bringing these low-cost irrigation technologies into more widespread use requires the creation of local, private-sector supply chains—including manufacturers, retailers and installers—as well as special innovations in marketing. The treadle pump has succeeded in Bangladesh in part because local businesses manufactured and sold the product and marketing specialists reached out to poor farmers with creative techniques, including an open-air movie and village demonstrations. The challenge is great, but so is the potential payoff. Paul Polak, a pioneer in the field of low-cost irrigation and president of International Development Enterprises in Lakewood, Colo., believes a realistic goal for the next 15 years is to reduce the hunger and poverty of 150 million of the world's poorest rural people through the spread of affordable small-farm irrigation techniques. Such an accomplishment would boost net income among the rural poor by an estimated $3 billion a year.

Over the next quarter of a century the number of people living in water-stressed countries will climb from 500 million to three billion. New technologies can help farmers around the world supply food for the growing population while simultaneously protecting rivers, lakes and aquifers. But broader societal changes—including slower population growth and reduced consumption—will also be necessary. Beginning with Sumeria, history warns against complacency when it comes to our agricultural foundation. With so many threats to the sustainability and productivity of our modern irrigation base now evident, it is a lesson worth heeding. SA

LOW-COST TREADLE PUMPS have helped more than a million Bangladeshi farmers irrigate for the first time.

COURTESY OF INTERNATIONAL DEVELOPMENT ENTERPRISES

Further Information

SALT AND SILT IN ANCIENT MESOPOTAMIAN AGRICULTURE. Thorkild Jacobsen and Robert M. Adams in *Science,* Vol. 128, pages 1251–1258; November 21, 1958.

PILLAR OF SAND: CAN THE IRRIGATION MIRACLE LAST? Sandra Postel. W. W. Norton, 1999.

GROUNDWATER IN RURAL DEVELOPMENT. Stephen Foster et al. Technical Paper No. 463. World Bank, Washington, D.C., 2000.

Irrigation and land-use databases are maintained by the United Nations Food and Agriculture Organization at http://apps.fao.org

APPROACH 1: SEEK NEW SOURCES

Sweating the Small Stuff

Extracting freshwater from the salty oceans is an ancient technique that is gaining momentum in a high-tech way

A water-covered planet facing a water crisis seems paradoxical. And yet that is exactly the reality on planet Earth, where 97 percent of the water is too salty to quench human thirst or to irrigate crops. Tackling water-shortage issues with desalination—drawing fresh, drinkable water out of salty seawater—is common in the desert nations of the Middle East, the Caribbean and the Mediterranean. But as the cost of desalination drops and the price and demand for water climb, countries in temperate regions are turning more and more to the sea.

Large-scale desalination facilities are even turning up in the U.S., one of the world's most water-rich countries. As part of an ambitious plan to reduce pumping from depleted underground aquifers, water officials in the Tampa Bay, Fla., area are contracting the construction of a desalination plant capable of producing 25 million gallons of desalted water a day. They are relying on desalination to supplement the region's future water demands. Houston is also looking at desalinating water from the Gulf of Mexico to keep from going dry.

People have been pulling freshwater out of the oceans for centuries using technologies that involve evaporation, which leaves the salts and other unwanted constituents behind. Salty source water is heated to speed evaporation, and the evaporated water is then trapped and distilled.

This process works well but requires large quantities of heat energy, and costs have been prohibitive for nearly all but the wealthiest nations, such as Kuwait and Saudi Arabia. (One exception is the island of Curaçao in the Netherlands Antilles, which has provided continuous municipal supplies using desalination since 1928.) To make the process more affordable, modern distillation plans recycle heat from the evaporation step.

A potentially cheaper technology called membrane desalination may expand the role of desalination worldwide, which today accounts for less than 0.2 percent of the water withdrawn from natural sources. Membrane desalination relies on reverse osmosis—a process in which a thin, semipermeable membrane is placed between a volume of saltwater and a volume of freshwater. The water on the salty side is highly pressurized to drive water molecules, but not salt and other impurities, to the pure side. In essence, this process pushes freshwater out of saltwater.

Most desalination research over the past few years has focused on reverse osmosis, because the filters and other components are much smaller than the evaporation chambers used in distillation plants. Reverse-osmosis plants are also more compact and energy-efficient.

Although reverse-osmosis plants can offer energy savings, the earliest membranes, made from either polyamide fibers or cellulose acetate sheets, were fragile and had short life spans, often no longer than three years. These materials are highly susceptible to contaminants in the source water—particularly chlorine, which hardens the membranes, and microbes, which clog them. Pretreatment regimes, such as filtering out sediments and bacteria, must be extremely rigorous. A new generation of so-called thin composite membranes, made from polyamide films, promises to eliminate these problems. Though still susceptible to contamination, these new membranes are sturdier, provide better filtration and may last up to 10 years.

Technical performance is important, but it alone does not drive the adoption of desalination as a source of clean

JOHN K. HUMBLE *Stone*

water. With or without technical improvements, the market for desalination equipment will very likely show healthy growth in the next 10 years as cities and other consumers realize the potential and favorable economics of existing equipment, according to James D. Birkett, who runs West Neck Strategies, a private desalination consulting company based in Nobleboro, Me.

Hundreds of suppliers are already selling many thousands of pieces of equipment annually. These desalination units range in capacity from a few gallons a day (small emergency units for life rafts) to several million gallons a day (municipal systems). "So confident are the suppliers that they enter into long-term contracts with their customers," Birkett says, "thus assuming themselves the risks of performance and economics." The desalination plant on Tampa Bay, scheduled to be operational by the end of 2002, will be funded and operated in such a manner.

Today the best estimate is that about 1 percent of the world's drinking water is supplied by 12,500 desalination plants. No doubt, this is only the beginning. In the future, the water in your glass may have originated in the seas. *—Diane Martindale*

DIANE MARTINDALE is a science writer based in New York City who says she will trade her bottle of Evian for a taste of the sea anytime.

Transporting water in enormous bags may not be such a crazy idea

Bagged and Dragged

Pipelines make it possible to move freshwater cheaply over vast distances of land. If only the same were possible over the oceans. Dragging waterproof plastic or fabric containers behind tugboats may be the answer.

Beginning in 1997, the English company Aquarius Water Trading and Transportation Ltd has towed water from mainland Greece to nearby resort islands in enormous polyurethane bags, helping the tourist destinations deal with increased demand for drinking water during the peak season. Another company, Nordic Water Supply in Oslo, Norway, has made similar deliveries from Turkey to northern Cyprus using their own fabric containers.

WATER-BAG INVENTOR Terry G. Spragg stands atop one of his giant plastic pouches as it is towed through Puget Sound during a demonstration in 1996.

AP PHOTO

The seemingly far-fetched concept of water bags was born in the early 1980s out of the desire to move large amounts of water more cheaply than modified oil tankers can do. For many years, tankers and barges have been making deliveries to regions willing to pay premium prices for small amounts of freshwater, such as the Bahamas, Cyprus and other islands with inadequate sources. Tankers have also supplied water during short-term droughts and disasters such as the 1995 Kobe earthquake in Japan.

Aquarius has manufactured eight 790-ton bags and two 2,200-ton versions; the latter hold about half a million gallons of water each. Aquarius has also developed models that are 10 times larger than the ones in use today, and last year Nordic began manufacturing bags that can hold nearly eight million gallons.

Water bags could offer a less expensive alternative to tankers—bags in the Aquarius fleet cost anywhere from $125,000 to $275,000—but some technical problems remain. In particular, making such large bags that are capable of withstanding the strains of an ocean voyage is difficult. For freshwater deliveries to the Greek isles and to Cyprus, bags need be dragged no farther than 60 miles. The piping systems needed to connect the bags to water supplies on land can be built from existing technology, but bags have ripped during transport on several occasions.

A third water-bag inventor, Terry G. Spragg of Manhattan Beach, Calif., is solving the problems of both volume and towing in a different way. With the support of privately hired scientists and consultants, Spragg has patented specialized zippers, with teeth more than an inch long, that can link water bags like boxcars. He has demonstrated the technology but has yet to sell it for commercial use.

Thus far this technology has been used only for freshwater deliveries to emergency situations and to extremely water-scarce coastal regions with a reliable demand for expensive water. But for some communities with no other option, water bags may offer a new and clever solution. *—Peter H. Gleick*

PETER H. GLEICK is the author of "Making Every Drop Count," on page 40, in this special report.

APPROACH 3: REDUCE DEMAND

Leaking Away

More than one billion gallons of water flow through New York City every day, and hardly a drop is wasted

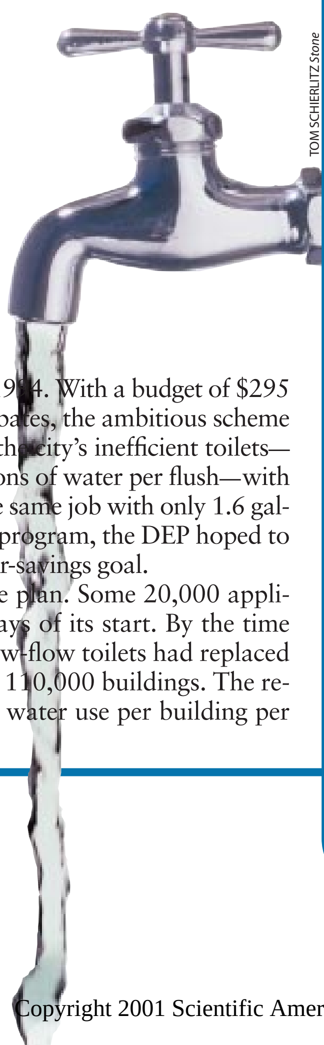
New York City is a metropolis of flamboyant excess, except when it comes to water. No one would suspect it, but the Big Apple has clamped down on water wasters, and after 10 years of patching leaky pipes and replacing millions of water-guzzling toilets, the city is now saving billions of gallons of water every year.

Back in the early 1990s New York City faced an imminent water shortage, and it was getting worse with every flush, shower and tooth brushing. With an influx of new residents and an increase in the number of drought years, the city needed to find an extra 90 million gallons of water a day—about 7 percent of the city's total water use. Instead of spending nearly $1 billion for a new pumping station along the Hudson River, city officials opted for a cheaper alternative: reduce the demand on the current water supply, which was piped in from the Catskill Mountains.

Officials knew that persuading New Yorkers to go green and conserve water would require some enticement—free toilets. The city's Department of Environmental Protection (DEP) stepped in with a three-year toilet rebate program, which began in 1994. With a budget of $295 million for up to 1.5 million rebates, the ambitious scheme set out to replace one third of the city's inefficient toilets—those using more than five gallons of water per flush—with water-saving models that do the same job with only 1.6 gallons per flush. With the rebate program, the DEP hoped to meet the largest part of its water-savings goal.

New Yorkers embraced the plan. Some 20,000 applications arrived within three days of its start. By the time the program ended in 1997, low-flow toilets had replaced 1.33 million inefficient ones in 110,000 buildings. The result: a 29 percent reduction in water use per building per year. The DEP estimates that low-flow toilets save 70 million to 90 million gallons a day citywide—enough to fill about 6,700 Olympic-size swimming pools.

But more efficient flushes weren't enough. The toilet rebate program happened concurrently with the city's water audit program, which continues today. For much of the city's history, the amount building owners paid for water



TOM SCHIERLITZ Stone

APPROACH 4: RECYCLE

Waste Not, Want Not

In the world's arid regions, even sewage water cannot be thrown away

Namibia is the driest African country south of the Sahara Desert. Blistering heat evaporates water faster than rains can rejuvenate the parched landscape, and there are no year-round rivers. Residents of the capital city, Windhoek, must do more than just conserve water to secure a permanent supply. They must reuse the precious little they have.

By the end of the 1960s, most underground aquifers and reservoirs on seasonal rivers near Windhoek had been tapped dry by the capital's burgeoning population, which has grown from 61,000 to more than 230,000 in the past 30 years. Transporting water from was based on the size of their property. Following a law passed in 1985, however, the city began keeping tabs on water use and charging accordingly. The law dictated that water meters be installed during building renovations, and the same requirement was applied to construction of new homes and apartments beginning in 1988. As of 1998, all properties in the city must be metered.

Homeowners who want to keep their water bills down under the new laws can request a free water-efficiency survey from Volt VIEWtech, the company that oversees the city's audit program. Inspectors check for leaky plumbing, offer advice on retrofitting with water-efficient fixtures and distribute free faucet aerators and low-flow showerheads. Low-flow showerheads use about half as much water as the old ones, and faucet aerators, which replace the screen in the faucet head and add air to the spray, can lower the flow of water from four gallons a minute to less than one gallon a minute. Volt VIEWtech has made several hundred thousand of these inspections, saving an estimated 11 million gallons of water a day in eliminated leaks and increased efficiency.

In efforts to save even more water, New York City has gone outside the home and into the streets. Water officials have installed magnetic locking caps on fire hydrants to keep people from turning them on in the summer. The city is also keeping an eye underground by using computerized sonar equipment to scan for leaks along all 32.6 million feet (6,174 miles) of its water mains.

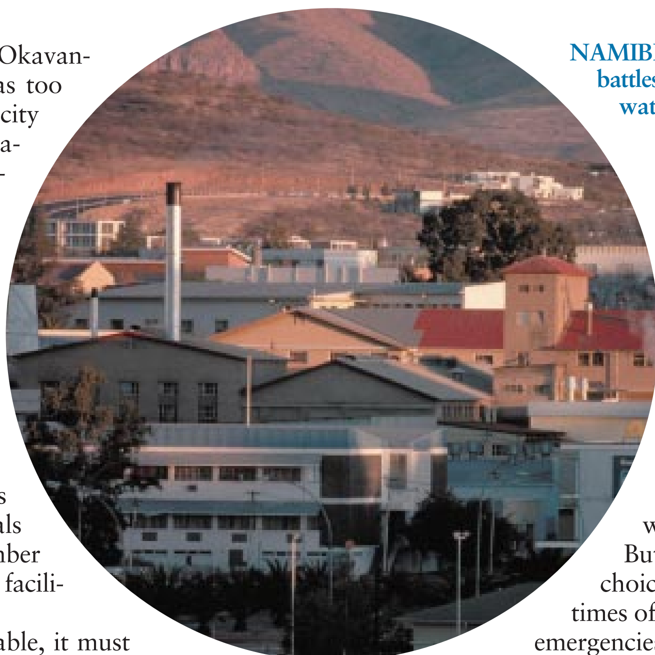
Although the city's population continues to grow, per person water use in New York dropped from 195 to 169 gallons a day between 1991 and 1999. From all indications, this trend is following its upward path. Water conservation works. And New Yorkers are proving that every flush makes a difference. —*D.M.*

For a list of the dos and don'ts about home water conservation, visit the New York City Department of Environmental Protection on the World Wide Web at www.ci.nyc.ny.us/dep

the closest permanent river, the Okavango—some 400 miles away—was too expensive. This crisis inspired city officials to implement a strict water conservation scheme that includes reclaiming domestic sewage and raising it once again to drinkable standards.

The city's first reclamation plant, initially capable of producing only 460 million gallons of clean water per year when it went on line in 1968, is now pumping out double that amount—enough to provide about 23 percent of the city's yearly water demands. Officials hope to boost that supply number to 51 percent with an upcoming facility now under construction.

To make wastewater drinkable, it must undergo a rigorous cleaning regimen. First, large solids are allowed to settle out while biofilters remove smaller organic particles. Advanced treatments remove ammonia, and carbon and sand filters ensure that the last traces of dissolved organic material are eliminated. The final step is to purify the water by adding chlorine and lime. To guarantee a safe drinking supply, the reclaimed water is tested once a week for the presence of harmful bacteria, viruses and heavy metals.



NAMIBIA'S CAPITAL CITY, Windhoek, battles water shortages by recycling wastewater for potable use.

PETER JOHNSON *Corbis*

(Industrial effluent laden with toxic chemicals is diverted to separate treatment plants.) Compared with local freshwater sources, the reclaimed water is equal or better in quality.

Despite 32 years of access to high-quality recycled water, the residents of Windhoek still doggedly oppose its use for personal consumption. For this reason, most of this purified wastewater irrigates parks and gardens. But sometimes people don't have a choice about their water source. In times of peak summer demand or during emergencies such as drought, local freshwater reservoirs are strained, and Windhoek relies heavily on treated effluent to boost supply. During the drought of 1995, for instance, reclaimed water accounted for more than 30 percent of the clean water piped into homes.

Officials hope to bolster support for the recycling program through enhanced public education—like letting the word slip that besides irrigating the city's greenery, treated wastewater is the secret ingredient in the prized local brew. —*D.M.*

Why the Y

Our X and Y chromosomes make an odd couple. The X resembles any other chromosome, but the Y—the source of maleness—is downright strange. How did the two come to differ so much?

by Karin Jegalian and Bruce T. Lahn

Y CHROMOSOME

The human chromosomes that determine sex—the X and Y—are a bizarre pair. The other 22 sets of chromosomes in our cells consist of well-matched partners, as alike as twin candlesticks. One chromosome in each duo comes from the mother and one from the father, but both are normally the same size and carry the same genes. (Genes are the DNA blueprints for proteins, which do most of the work in the body.) In stark contrast, the Y chromosome is much smaller than the X; in fact, it is positively puny. It harbors no more than several dozen genes, far fewer than the 2,000 to 3,000 on the X. A number of the Y genes have no kin at all on the X. And the Y is riddled with unusually high amounts of "junk" DNA: sequences of code letters, or nucleotides, that contain no instructions for making useful molecules.

Until recently, biologists had difficulty explaining how the Y fell into such disrepair. They had various theories but few ways to test their ideas. That situation has now changed, thanks in large part to the Human Genome Project and related efforts aimed at deciphering the complete sequence of DNA nucleotides in all 24 distinct chromosomes in humans—the X, the Y and the 22 autosomes (the chromosomes not involved in sex determination). Just as paleontologists can trace the evolution of a species by examining skeletons of living animals and fossils, molecular biologists have learned to track the evolution of chromosomes and genes by examining DNA sequences.

The new findings demonstrate that the history of the sex chromosomes has been strikingly dynamic, marked by a series of dramatic disruptions of the Y and by compensatory changes in the X. That interplay undoubtedly continues today.

Further, the Y chromosome—long regarded as a shambles, able to accomplish little beyond triggering the maleness program—turns out to do more than most biologists suspected. Over some 300 million years it has managed to preserve a handful of genes important for survival in males and to acquire others needed for fertility. Instead of being the Rodney Dangerfield of chromosomes (as some have called the chronically disrespected Y), the male chromosome is actually more like Woody Allen: despite its unassuming veneer, it wields unexpected power.

Sheer curiosity has driven much of the research into the evolution of the human sex chromosomes. But a more practical pursuit has informed the work as well: a desire to explain and reverse male infertility. Discoveries of Y genes that influence reproductive capacity could lead to innovative treatments for

Is So Weird

X CHROMOSOME

men who lack those genes or have defective versions [*see box on page 61*].

The recent advances have benefited from insights achieved beginning about 100 years ago. Before the 20th century, biologists thought that the environment determined sex in humans and other mammals, just as it does in modern reptiles. For reptiles, the temperature of an embryo at an early point in development tips some poorly understood system in favor of forming a male or female. In the early 1900s, though, investigators realized that chromosomes can arbitrate sex in certain species. About 20 years later mammals were shown to be among those using chromosomes—specifically the X and Y—to determine sex during embryonic development.

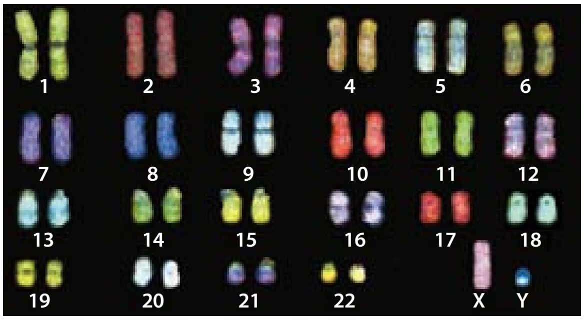
Clues Piled Up

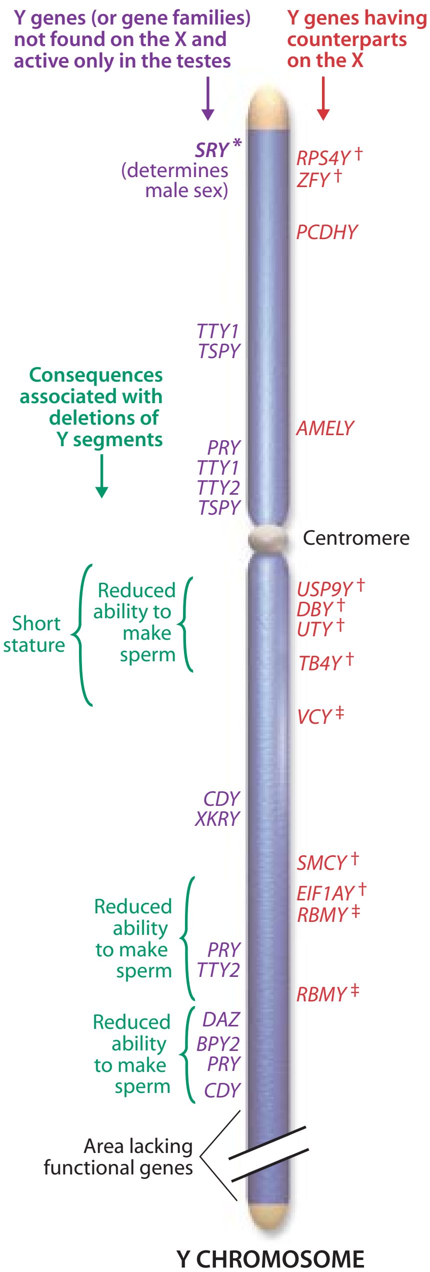
In the ensuing decades, researchers identified the Y as the male maker and deduced that the X and Y evolved from matching autosomes in an ancient ancestor. By chance, sometime shortly before or after mammals arose, a mutation in one small part of the autosome copy that would become the Y caused embryos inheriting that changed chromosome (along with its mate, the future X) to become males. Embryos inheriting two Xs became females.

In 1990 geneticists pinpointed the part of the Y that confers maleness. It is a single gene, named *SRY,* for "sex-determining region Y." The protein encoded by *SRY* triggers the formation of the

X AND Y CHROMOSOMES started off as a matched pair hundreds of millions of years ago. But the Y shrank to a nubbin, whereas the X maintained its integrity. How the pair came to diverge so strikingly is becoming clear. The micrographs show the chromosomes as they appear during the metaphase stage of cell division.

BIOPHOTO ASSOCIATES *Photo Researchers, Inc.*





"Pseudoautosomal" regions, able to swap DNA (recombine) with the X

* *SRY* causes the testes to form. It derived from the gene *SOX3* and resembles *SOX3* on the X, but the two have different functions

† Housekeeping genes

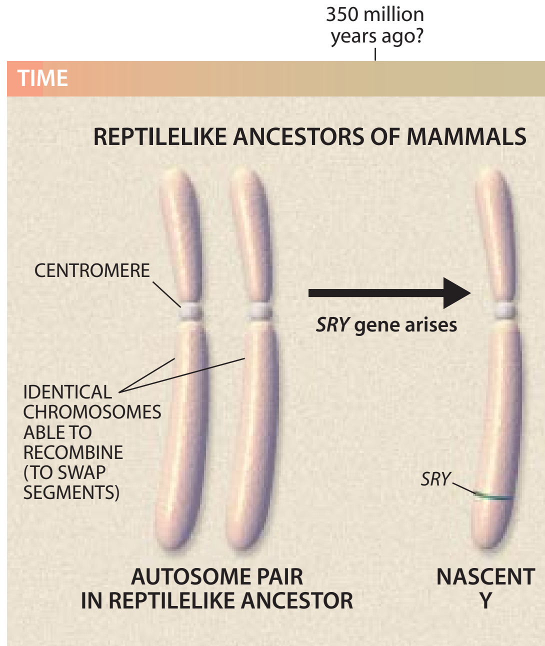
‡ Genes that have X counterparts but are active only in the testes

testes, apparently by activating genes on various chromosomes. Thereafter, testosterone and other substances made in the testes take over the molding of maleness.

Scientists concluded that the human sex chromosomes started life as a matched pair in part because the tips of the X and Y have remained twinlike and able to engage in a process called recombination. During meiosis (the cell division that yields sperm and eggs), matching chromosomes line up together and swap segments, after which one copy of every autosome plus a sex chromosome is distributed evenly to each reproductive cell. Even though most of the Y now bears little resemblance to the X, the tips of those chromosomes align during meiosis in males and exchange pieces as if the X and Y were still a matching set. (Such alignment is critical to the proper distribution of chromosomes to sperm.)

Other evidence that the X and Y were once alike came from the part of the Y that does not recombine with the X. Many of the genes scattered through this nonrecombining region still have counterparts on the X.

The existence of the nonrecombining region, which makes up 95 percent of the Y, offered a clue to how that chromosome became a shadow of its original self. In nature and in the laboratory, recombination helps to maintain the integrity of chromosomes. Conversely, a lack of it causes genes in nonrecombining regions to accumulate destructive mutations and to then decay or disappear. It seemed reasonable to think, therefore, that something caused DNA exchange between large parts of the X and Y to cease, after which genes in the nonrecombining region of the Y collapsed. But when and how recombination stopped after the Y emerged remained uncertain for decades.



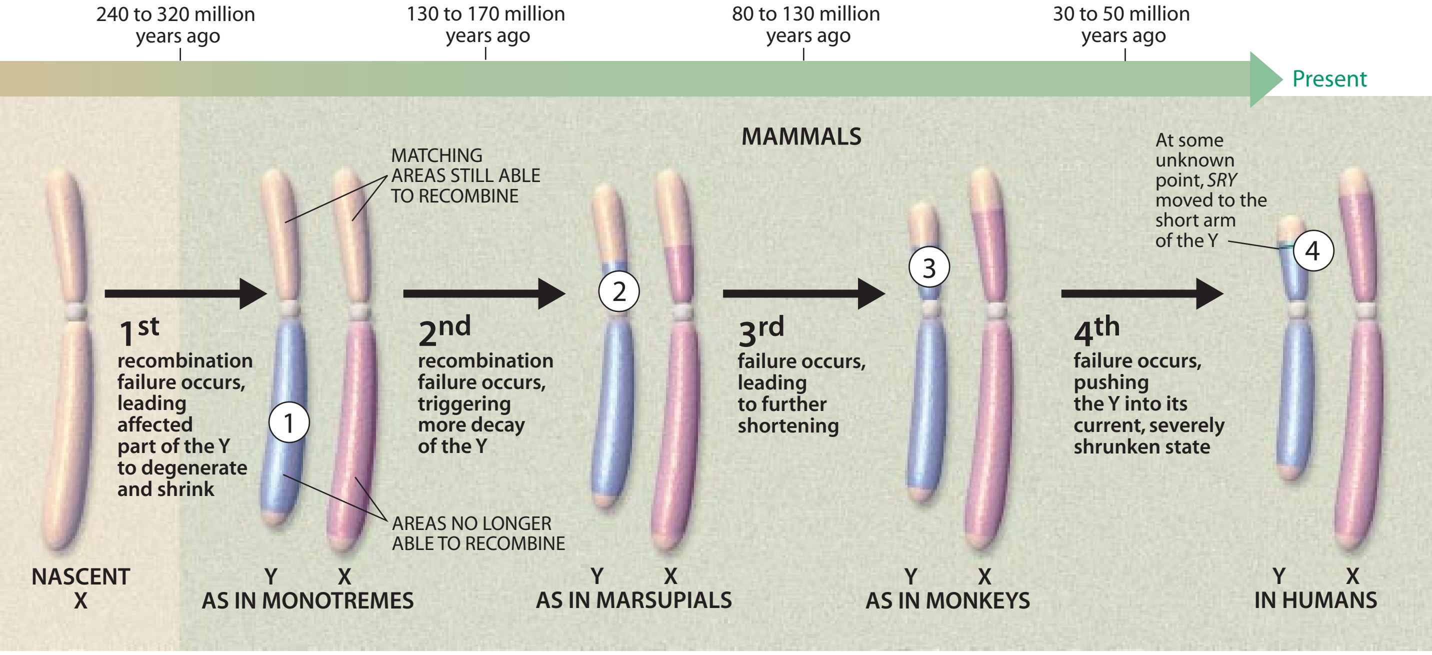
Shaped in Stages

Work completed in the past five years has filled in many of the gaps. For instance, in 1999 one of us (Lahn) and David C. Page of the Whitehead Institute for Biomedical Research in Cambridge, Mass., showed that the Y lost the ability to swap DNA with the X in an unexpected, stepwise fashion—first involving a swath of DNA surrounding the *SRY* gene and then spreading, in several discrete blocks, down almost the full length of the chromosome. Only the Y deteriorated in response to the loss of X-Y recombination, however; the X continued to undergo recombination when two copies met during meiosis in females.

What could account for the disruption of recombination between the X and the Y? As the early X and Y tried to trade segments during meiosis in some far-distant ancestor of modern mammals, a part of the DNA on the Y probably became inverted, or essentially flipped upside down, relative to the equivalent part on the X. Because recombination requires that two like sequences of DNA line up next to each other, an inversion would suppress fu-

CHROMOSOMES from a normal male cell (*photograph*) include 22 pairs of autosomes (those not involved in sex determination), plus an X and a Y; one member of each pair comes from the mother and one from the father. Genes in the NRY, or nonrecombining region of the Y (*blue in diagram*), have helped reveal the evolutionary history of the X and the Y. The region is so named because it cannot recombine, or exchange DNA, with the X. Only genes that still work are listed. About half have counterparts on the X (*red*); some of these are "housekeeping" genes, needed for the survival of most cells. Certain NRY genes act only in the testes (*purple*), where they likely participate in male fertility.

HESED PADILLA-NASH *National Institutes of Health (micrograph)*; ALFRED T. KAMAJIAN AND HEIDI NOLAND *(illustrations at left and above)*



DEGENERATION OF THE Y occurred in four discrete episodes starting about 300 million years ago, after a reptilelike ancestor of mammals acquired a new gene (*SRY*) on one of its autosomal chromosomes. Each of the episodes involved a failure of recombination (DNA exchange) between the X and the Y during meiosis, the cell division that yields sperm and eggs. If recombination is blocked, genes in the affected regions stop working and decay. The sequence shown is highly simplified. For instance, the Y actually expanded temporarily at times (by stitching autosomal DNA into areas still able to recombine), before failures of recombination led to a net shrinkage.

ture interaction between the formerly matching areas of the X and Y.

We discovered that recombination ceased in distinct episodes when we examined the nucleotide sequences of 19 genes that appear in the nonrecombining region of both the X and the Y. (Some of the Y copies no longer function.) In general, if paired copies of a gene have stopped recombining, their sequences will diverge increasingly as time goes by. A relatively small number of differences implies recombination stopped fairly recently; a large number means it halted long ago.

Most of the X-Y pairs fell into one of four groups. Within each group, the X and the Y copies differed by roughly the same amount, indicating that recombination stopped at about the same time. But the groups clearly varied from one another. The Y copies that began diverging from their counterparts on the X at about the time the *SRY* gene arose differed from their partners the most, and the other groups showed progressively less divergence between the X and Y copies.

By comparing DNA sequences across species, biologists can often calculate roughly when formerly matching genes (and hence the regions possessing those genes) began to go their separate ways. Such comparisons revealed that the autosomal precursors of the X and Y were still alike and intact in reptiles that existed before the mammalian lineage began branching extensively. But monotremes (such as the platypus and echidna), which were among the earliest to branch off from other mammals, possess both the *SRY* gene and an adjacent nonrecombining region. These differences implied that the *SRY* gene arose, and nearby recombination halted, close to when the mammalian lineage emerged, roughly 300 million years ago.

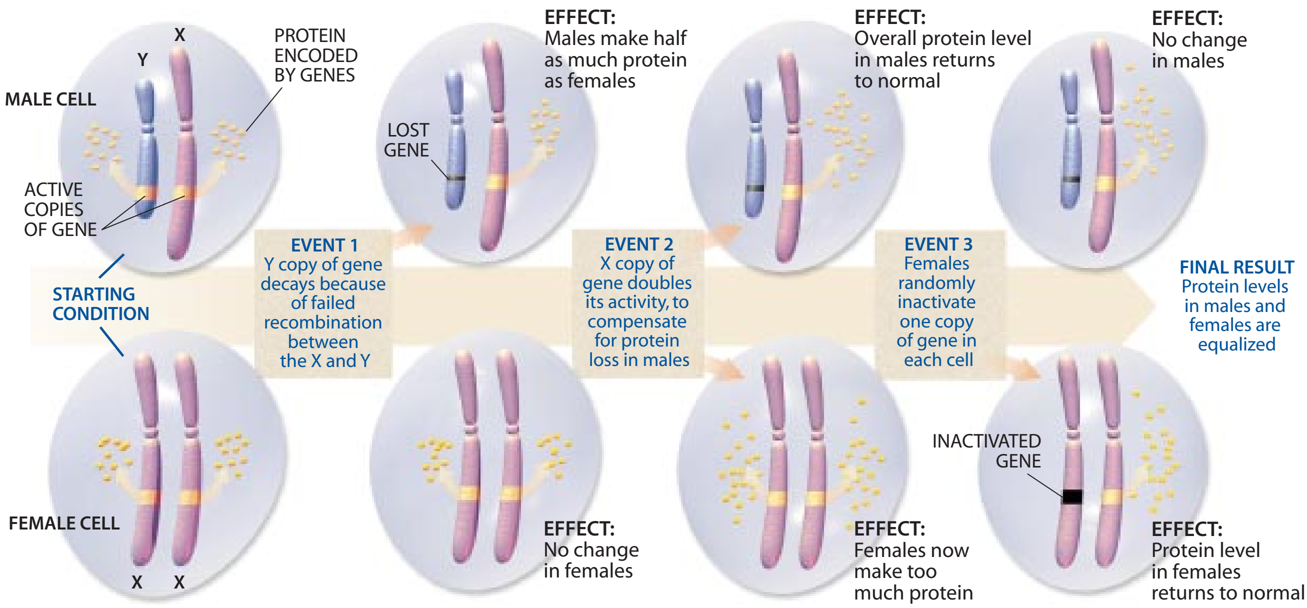
We gained more information about the timing by applying a "molecular clock" analysis. Biologists can estimate the background rate at which DNA sequences are likely to change if they are under no particular pressure to stay the same. By essentially multiplying the extent of sequence disparity in X-Y pairs by that estimated rate, we deduced that the first recombination-halting inversion took place between 240 million and 320 million years ago.

Similar analyses imply that the next inversion occurred 130 million to 170 million years ago, shortly before marsupials branched off from the lineage that gave rise to all placental mammals. The third struck 80 million to 130 million years ago, before placental mammals diversified. And the final inversion rocked the Y roughly 30 million to 50 million years ago, after monkeys set off on their own evolutionary path but before apes and hominids parted company.

Bucking the overall trend for X-Y pairs, some genes in the nonrecombining region of the Y code for proteins that differ remarkably little from the proteins encoded by their X counterparts, even in the regions that underwent inversion earliest. Their preservation is probably explained by a simple evolutionary law: if a gene is crucial to survival, it will tend to be conserved. Indeed, the Y genes that have changed the least are mainly "housekeeping" genes—ones critically required for the integrity of almost all cells in the body.

Making up for Losses

Logic—and a large body of research—indicates that the failure of recombination between the X and the Y, and the subsequent deterioration of many Y genes, must have been followed by a third process that compensated for the degeneration. The reasoning goes like this: Not all genes are active in every cell. But when a cell needs particular proteins, it typically switches on both the maternal and paternal copies of the corresponding genes. The amount of protein generated from each copy is fine-tuned for the optimal development and day-to-day operation of an organism. Therefore, as genes on the Y began



EVOLUTION OF X INACTIVATION, the silencing of most genes on one X chromosome in female cells, apparently occurred in a piecemeal fashion—one gene or a few genes at a time—to compensate for losses of genes on the Y chromosome (*diagram*). One effect of X inactivation can be seen in calico cats (*photograph*). The gene determining whether fur color is orange or black (that is, not orange) resides on the X. Females that carry the orange version on one X and the black version on the other X will end up with some orange areas and some black ones, depending on which X is shut down in each cell. A different gene accounts for the white areas.

RAZI SEARLES *Bruce Coleman Inc.* (cat); ALFRED T. KAMAJIAN

to disappear, the production of the associated proteins would have been halved disastrously in males unless the affected species evolved compensatory tricks.

Many animals, such as the fruit fly, handle this inequity by doubling the activity of the X versions of lost Y genes in males. Others employ a more complex strategy. First they increase the activity of X genes in both males and females, a maneuver that replenishes protein levels in males but creates an excess in females. Some animals, such as the nematode worm, then halve the activity of X genes in females. Others, including mammals, invoke a process called X inactivation, in which cells of early female embryos randomly shut off most of the genes in one of their two X chromosomes. Neighboring cells may silence different copies of the X, but all the descendants of a given cell will display its same X-inactivation pattern.

Although X inactivation has long been thought to be a response to the decay of Y genes, proof for that view was lacking. If degeneration of Y genes drove X inactivation, then X genes having functional counterparts in the nonrecombining region of the Y would be expected to keep working in females (that is, to evade inactivation)—so as to keep protein levels in females equivalent to those in males. In analyzing the activity levels of surviving X-Y pairs from two dozen mammalian species, one of us (Jegalian) and Page found a few years ago that the X copies of working Y genes do escape inactivation. Those analyses also revealed that X inactivation, although it happens in an instant during an animal's development today, did not evolve all at once. Instead it arose rather diffidently—patch by patch or perhaps gene by gene within a patch, not all at once down the chromosome.

Emerging Themes

Curiously, the nonrecombining region of the Y possesses not only a handful of valuable genes mirrored on the X but also perhaps a dozen genes that promote male fertility. The latter code for proteins made solely in the testes, presumably to participate in sperm production. Some seem to have jumped onto the Y from other chromosomes. Others have apparently been on the Y from the start but initially had a different purpose; they acquired new functions over time. Degeneration, then, is but one theme prominent in the evolution of the Y chromosome. A second theme, poorly recognized until lately, is the acquisition of fertility genes.

Theorists disagree on the forces that turned the Y into a magnet for such genes. The species as a whole may benefit from sequestering in males genes that could harm females or do nothing useful for them. It is also possible that being on the Y protects male fertility genes by ensuring that they go from male to male without having to detour through females (who could discard them without suffering any direct consequences).

Another mystery is how fertility genes can thrive in the absence of recombination, under conditions that corrupted most of the Y's other genes. An answer may lie in the observation that nearly every male fertility gene on the Y exists in multiple copies. Such amplification can buffer the effects of destructive mutations, which usually afflict just one copy at a time. As some copies accumulate mutations and eventually fail, the remaining ones continue to preserve a man's reproductive ability and to serve as seeds for their own multiplication.

The evolution of the sex chromosomes has been studied most thoroughly in humans. But together with cross-species comparisons, that research has identified general principles operating even in creatures that evolved sex chromosomes independently from mammals. Some of those animals, such as birds and butterflies, use the W-Z system of sex determination. When inheritance of a single copy of a specific chromosome leads to the formation of a male, that chromosome is termed the Y, and its partner is termed the X. When inheritance of one copy of a chromosome leads to the formation of a female, that chromosome is called the W, and its mate is called the Z.

One notable principle is that sex chromosomes derive from autosomes. The affected autosomes can vary, however. W and Z chromosomes in birds evolved, for example, from different autosomes than those that gave rise to the mammalian X and Y. And the X and Y in fruit flies derived from different autosomes than those enlisted by mammals.

In most sexually reproducing species, once sex chromosomes arose, they became increasingly dissimilar as they underwent one or more cycles of three sequential steps: suppression of recombination, degeneration of the nonrecombining parts of the sex-specific chromosome (the Y or W) and, finally, compensation by the other chromosome. At the same time, the sex-specific chromosome in many instances became important for fertility, as happened to the Y in humans and insects.

It is reasonable to wonder what the future holds for our own species. Might the cycle continue until it wipes out all recombination between the X and the Y and ultimately destroys the Y, perhaps thousands or millions of years from now? The new discoveries suggest males are able to protect Y genes that are critical for male survival and fertility. Nevertheless, total decay of the Y remains a theoretical possibility.

Research into genes is often undertaken with an eye to understanding and correcting human disorders. Some investigations into the Y chromosome began with just such a goal in mind—understanding male development and correcting infertility. But many studies were less focused on therapy. As more and more genes on the X and Y were identified by medical research and systematic sequencing efforts, evolutionary-minded scientists could not resist asking, on a more basic level, whether those genes had anything new to say about the distant past of the strangely mismatched X and Y chromosome. As it turns out, the genes had a rich tale to tell. SA

NEW ANSWERS FOR MEN

Y Am I Infertile?

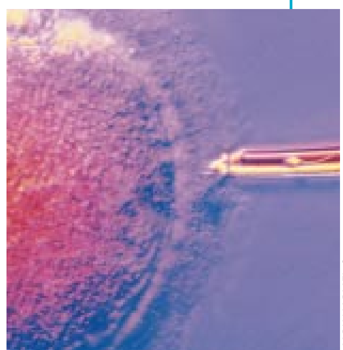
Beyond revealing the history of the sex chromosomes, genetic studies of the Y are helping to explain some cases of infertility. In about half of all affected couples, the problem rests fully or partly with the man, who occasionally produces insufficient numbers of sperm or even none at all. Often the roots of these abnormalities are obscure. New findings suggest, though, that the Y contains a number of fertility genes and that disruption of one or more of them accounts for about 10 percent of men who make little or no sperm.

Investigators first inferred a role for the Y in infertility in the1970s, when they saw through a microscope that many sterile men lacked small bits of the Y normally present in fertile men. Today scientists know that deletions in any of three specific regions on the Y can cause infertility, and they have learned that each of these regions—referred to as AZF (for azoospermia factor) a, b and c—contains multiple genes.

Most of those genes are highly active in the testes, where sperm is made (that is, the genes yield abundant amounts of the proteins they encode). This behavior strongly suggests that the genes in the AZF regions are important for sperm manufacture, although their exact contributions, and their interactions with fertility genes on other chromosomes, remain to be determined.

Some infertility specialists are now assessing Y chromosome deletions as part of their diagnostic workups. If men found to have such deletions produce at least some sperm, they might be offered a therapy called intracytoplasmic sperm injection (ICSI), in which sperm is retrieved from the testes and inserted into eggs in the laboratory. Regrettably, sons conceived in this way will inherit their fathers' defective Y chromosomes and so will probably face the same fertility challenges.

Once researchers decipher the exact functions of the proteins encoded by AZF-area genes, they may be able to reverse infertility in men possessing Y deletions by replacing the missing proteins, perhaps by restoring the lost genes themselves. On the flip side, such information should make it possible to devise drugs that purposely disrupt the sperm-production machinery—thereby providing new male contraceptives. *—K.J. and B.T.L.*



DELIVERING SPERM (*visible in microneedle*) directly into an egg may overcome infertility in some men afflicted by mutations of the Y chromosome.

MARK HARMEL *Stone*

The Authors

KARIN JEGALIAN and BRUCE T. LAHN completed their doctoral research in the laboratory of geneticist David C. Page at the Whitehead Institute for Biomedical Research in Cambridge, Mass., and the Massachusetts Institute of Technology. Jegalian is now a science writer at the National Institutes of Health. Lahn, a native of China, is a researcher of the Howard Hughes Medical Institute and professor in the department of human genetics at the University of Chicago.

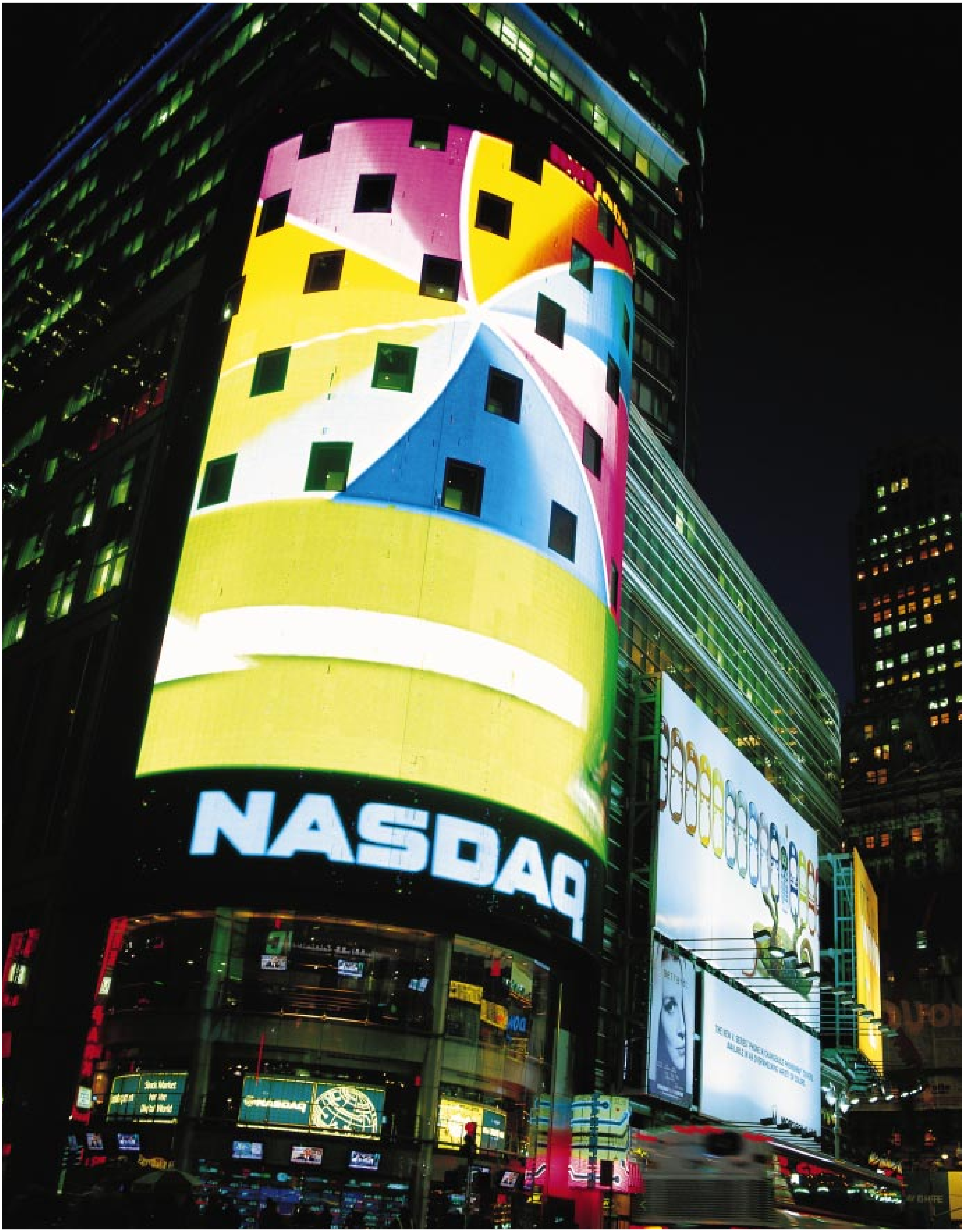
Further Information

Functional Coherence of the Human Y Chromosome. Bruce T. Lahn and David C. Page in *Science*, Vol. 278, pages 675–680; October 24, 1997.

A Proposed Path by Which Genes Common to Mammalian X and Y Chromosomes Evolve to Become X Inactivated. Karin Jegalian and David C. Page in *Nature*, Vol. 394, pages 776–780; August 20, 1998.

Four Evolutionary Strata on the Human X Chromosome. Bruce T. Lahn and David C. Page in *Science*, Vol. 286, pages 964–967; October 29, 1999.

The Human Y Chromosome, in Evolution's Light. Bruce T. Lahn, Nathaniel M. Pearson and Karin Jegalian in *Nature Reviews Genetics* (in press).



PETER MURPHY

Full-spectrum light-emitting diodes, or LEDs, are becoming widespread—and the race is on to develop white-light versions to replace Edison's century-old incandescent bulb

by M. George Craford, Nick Holonyak, Jr., and Frederick A. Kish, Jr.

In Pursuit of the ULTIMATE LAMP

In 1995 one of us (Holonyak) was honored to accept the Japan Prize for pioneering work in semiconductor light emitters and lasers. Asked to say a few words about tomorrow's technology, he simply pointed to the ceiling lights and said, "All of this is going."

A revolution is taking place, literally in front of our eyes, thanks to semiconductor devices known as light-emitting diodes, or LEDs. Most familiar as the little glowing red or green indicator lights on electronic equipment, LEDs are beginning to replace incandescent bulbs in many applications. The reason? LEDs convert electricity to colored light more efficiently than their incandescent cousins—for red light, their efficiency is 10 times greater. They are rugged and compact; some types last a phenomenal 100,000 hours, or about a decade of regular use. In contrast, the average incandescent bulb lasts about 1,000 hours. Moreover, the intensity and colors of LED light have improved so much that the diodes are now suitable for large displays—perhaps the most impressive example being the eight-story-tall Nasdaq billboard in New York City's Times Square.

Currently engineers are trying to lower the cost of manufacturing LEDs, improve their efficiency and extend their range of useful colors. In fact, it is possible to combine the output of red, green and blue LEDs to make white light, a cheap, mass-market form of which would be the brilliant prize of the industry. Such an LED could someday supplant, more than a century after Thomas Edison's invention, the incandescent lightbulb.

NASDAQ MARKETSITE TOWER, the world's largest video screen, uses 18,677,760 LEDs covering 10,736 square feet.

LEDs are already replacing lightbulbs in several instances, albeit in a fashion less dramatic than the giant Nasdaq display. The new applications are perhaps most noticeable to automobile drivers. In Europe 60 to 70 percent of the cars produced use LEDs for their high-mount brake lights, and the U.S. is beginning to move in that direction, too. LEDs are also being used for taillights and turn signals, as well as for side markers for trucks and buses. We expect that by the end of the decade LEDs will dominate the red and amber lighting on the exterior of vehicles. Larger and brighter LEDs are making their way into the red of traffic lights. About 10 percent of the nation's stoplights include LEDs.

Traditionally, traffic signals and other colored lamps use incandescent bulbs, which are then covered with a filter to produce the appropriate hue. Although filtering is cheap—the bulbs emit light that costs a mere fraction of a cent per lumen (the standard measurement unit of illumination)—it is a terribly inefficient way to produce light. A red filter, for example, blocks about 80 percent of the glow, so the amount of light that emerges drops from about 17 lumens per watt of power to about three to five lumens per watt.

In contrast, lumens cast by an LED stoplight may cost around 15 cents each to produce, but virtually all of them are of the right color. What is more, the LEDs in a stoplight consume only 10 to 25 watts, compared with the 50 to 150 watts used by an incandescent bulb of similar brightness. This energy savings pays for the higher cost of an LED in as little as one year. When this figure is considered with the reduced maintenance and labor costs of LEDs, it is easy to see why LEDs are becoming more popular with city planners.

Interior designers began to use LEDs a few years ago, when high-brightness models of all colors made their appearance. Because each LED gives off one distinct hue, users can have complete control of nearly the full spectrum. By putting differently colored LEDs together in an array, the user can adjust their combined light. For example, white light composed of red, green and blue LEDs can be made to feel "cooler" by turning off more of the red LEDs and turning on more of the blue ones.

This flexibility offers novel ways of using light. Rather than, say, putting up new wallpaper or applying fresh paint, one can manipulate the color of a room by adjusting the ratio of wavelengths in the emitted light (the wavelengths of light determine the colors). The Metropolitan Museum of Art in New York City relied on just this kind of LED lighting to illuminate its display of the Beatles' *Sgt. Pepper's* costumes in 1999. (The lighting is also cool and will not damage the fabric.) Photographers would have total control of artificial light sources without the need for cumbersome filters or gels.

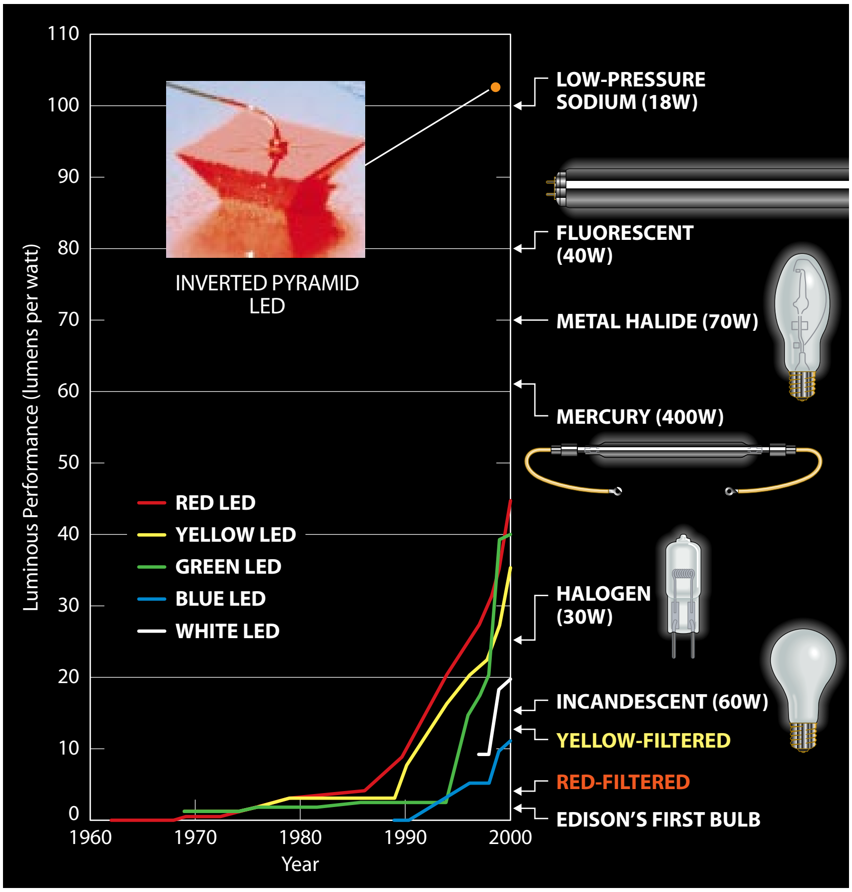
LEDs also offer intriguing possibilities for medical science. For example, LEDs' cool temperatures, precise wavelength control and broad-beam characteristics enable cancer researchers to study the photodynamic treatment of tumors more effectively. In this therapy, patients receive light-sensitive drugs that are preferentially absorbed by tumor cells. When light of the appropriate wavelength strikes these chemicals, they become excited and destroy the cells. By using an array of LEDs linked together, researchers can produce an even "sheet" of light that stimulates light-sensitive drugs without burning the patient's skin.

Up Close

Maybe the easiest way to examine light-emitting diodes in detail is to purchase a few of them, in the form of a $15 flashing-red LED bicycle light. Open up the casing, and you'll see a pair of AA batteries wired to a circuit board containing a series of clear, colorless, cylindrical knobs approximately 5/16 of an inch high and 3/16 of an inch in diameter. Each of these transparent knobs is a light-emitting diode. Press the "on" button, and the clear LED turns red, casting a color so brilliant that it can be painful to look at directly. If you turn it off and closely examine the LED, you will see what is the equivalent of a wire threading through its base—and what looks like a miniature cup about halfway up. This cup is a reflector, which holds a semiconductor chip about the size of a grain of sand. This chip is the LED's "heart."

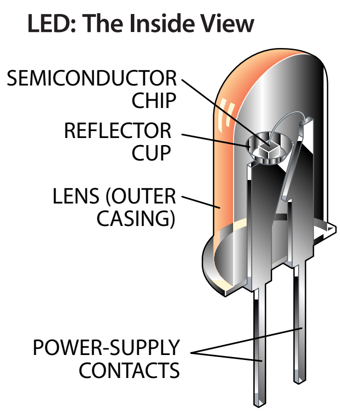
Inside the chip, there is a layer that has an excess of electrons; the substance is called *n*-type (for "negative"). Another layer rests on top and is made of a mate-

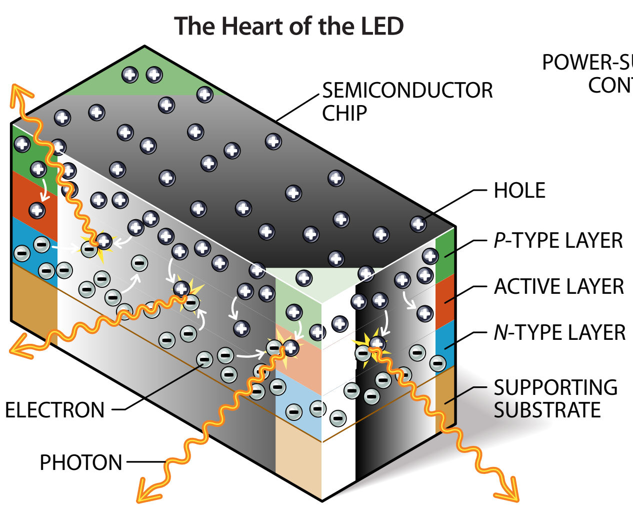
LED Performance

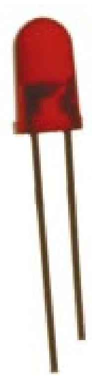


LIGHT-EMITTING DIODES have steadily improved and now outperform many other kinds of lights; the best is a prototype red-orange inverted pyramid LED.

SEMICONDUCTOR CHIP is the key to an LED's glow. An applied voltage drives "holes" (positive charges) from the *p*-type layer and electrons from the *n*-type layer into the active layer. When they meet, they give off photons. The color of the photons depends on the chemical makeup of the layers, although some manufacturers house LEDs in colored lenses as a means of identification (*photograph*).

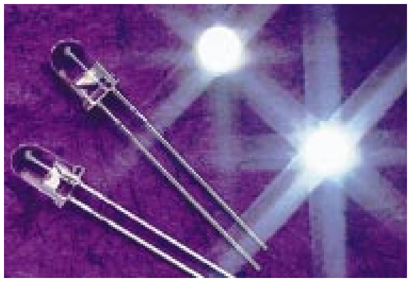






JARED SCHNEIDMAN DESIGN (*illustrations*); LUMILEDS LIGHTING (*photograph*) rial that has a dearth of electrons—or, as electrical engineers like to say, an excess of positively charged particles known as holes. This material is *p*-type (for "positive"). At the junction of the *n* and *p* layers is the so-called active layer, where light is emitted.

Applying a voltage drives electrons and holes into the active layer, where they meet. As they join, they emit photons—the basic units of light. The atomic structures of the active layer and adjoining materials on each side determine the number of photons produced and their wavelengths.



LUMILEDS LIGHTING

WHITE LEDs are possible, but affordable ones powerful enough to illuminate a room remain at least a decade away.

In early LEDs, made in the 1960s with a combination of gallium, arsenic and phosphorus to yield red light, electrons merged with holes relatively inefficiently: for every 1,000 electrons, only one red photon was produced. Such an LED generated less than one tenth the amount of light found in a comparably powered, red-filtered incandescent bulb.

Over time, however, dramatic improvements in output were realized, especially at the red end of the spectrum. In 1999 Michael Krames and his co-workers at Hewlett-Packard set an efficiency record, building LEDs that transform more than 55 percent of the incoming electrons into photons at the red wavelength. Chief among the reasons for these improvements has been the continued rise in material quality and the development of substances that allow the efficient transformation of electrons and holes into photons. One of the biggest boosts in efficiency came when scientists found that the materials do not have to be homogeneous. Instead each layer can have a different chemical makeup, so that when placed next to the active layer they can confine the electrons and holes better, thereby increasing the odds that an electron can combine with a hole to produce light.

Researchers have also learned to tailor the properties of the semiconducting layers. They can vary the material composition of the active layer and "dope" the layers with impurities. Doping alters the *n*- and *p*-type characteristics of the semiconductor layers. To take the simplest example, consider silicon. Silicon belongs to group IV of the periodic table: it has four electrons that are available for bonding with other atoms. In the crystal form, each atom shares electrons with neighboring atoms. When a small number of group III atoms—that is, those containing three electrons in their outermost energy level, such as boron—are incorporated into the crystal, the resulting structure has an insufficient number of bonds to share with the surrounding silicon atoms. Vacancies for electrons appear, creating holes and rendering the material *p*-type.

Conversely, elements that belong to group V of the periodic table, such as phosphorus, have an extra electron in their outermost energy level. When silicon is doped with phosphorus, the crystal gains electrons, making the material *n*-type.

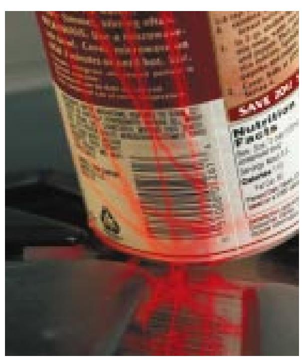
In LEDs, the crystal is not silicon but a mixture of group III and group V elements. By carefully controlling the concentration of aluminum, gallium, indium and phosphorus, for example, and by incorporating suitable dopants, typically tellurium and magnesium, researchers can control the formation of the *n*-side and the *p*-side, making LEDs that emit at the red, orange or yellow wavelengths. By the early 1970s red LEDs containing gallium arsenide phosphide were bright enough to illuminate the first calculators and digital clocks.

Another key to LED improvement lies in manufacturing techniques that reliably create viable, smooth crystals instead of lumpy, defect-riddled systems. The atomic lattices of the *p* and *n* materials must match up with those of the underlying supporting substrate and active layer. One such manufacturing

LEDs and Lasers

Although light-emitting diodes and laser diodes sound similar—in fact, both are made of semiconductor materials—they are very different beasts, designed to behave in different ways and to tackle different jobs.

Laser diodes take the form of a semiconductor material between what is essentially a pair of mirrors. The region between the mirrors is called the resonator cavity. When electricity goes through the semiconductor, it gives off photons, which then bounce back and forth inside the cavity, exciting other, nearby electron-hole pairs to release more photons at the same wavelength. The light increases continuously in intensity, with the photons marching in lockstep together as they oscillate between the two mirrors. If one of the mirrors allows just a small fraction of the light to escape, then some of the photons exit. All at the same wavelength and in phase, they produce an extremely narrow column of pure, bright light at a single wavelength. In physicists' terms, the photons are coherent.



BAR-CODE SCANNING relies on semiconductor lasers.

PHIL DEGGINGER *Stone*

This extremely well defined beam is one of the main characteristics of a laser. As such, it is something like a scalpel: sharp, thin and able, with proper optics, to do delicate work, such as reading the fine pits on a compact disc or scanning the bar codes in a checkout line.

By comparison, the widely scattered light of an LED is like the patter of raindrops. Because LEDs are not in a proper cavity (that is, not between mirrors), the photons they emit are, in a sense, incoherent. Light comes out not in a unidirectional column but in a broader, more diffuse pattern, composed of a spread of wavelengths from one area of the spectrum. The photons an LED produces may not be all at the exact same wavelength, but they are close enough so that they are perceived by the human eye as being the same color.

—M.G.C., N.H. and F.A.K.

method is vapor phase deposition, in which hot gases are channeled over a substrate to create a thin film. This technique was first incorporated into a high-volume LED manufacturing process at Monsanto in the late 1960s. In 1977 a different process of vapor phase deposition, one utilizing cool gases directed over a hot substrate, was demonstrated by Russell D. Dupuis, now at the University of Texas, to produce semiconductor lasers. This process, which enables the growth of a wider variety of materials, is now used to make high-quality LEDs. Shuji Nakamura, now at the University of California at Santa Barbara, used a variation of the technique to manufacture high-quality gallium nitride crystal capable of shining blue light. (For a profile of Nakamura, see "Blue Chip," by Glenn Zorpette; Scientific American, August 2000.)

In the mid-1990s a team at Hewlett-Packard found another way to enhance brightness—by reshaping the chip itself. Through careful manipulation, researchers can remove the original gallium arsenide wafer on which the active layer was grown, replace it with a transparent gallium phosphide wafer and sculpt an LED into the shape of an inverted pyramid. This shape decreases the number of internal reflections and thus boosts the amount of light escaping from the chip.

The Great White Hope

Thanks to these improvements in color and brightness, researchers have begun to zero in on making affordable, bright white LEDs. Low-power white LEDs with an efficiency somewhat better than incandescent bulbs are already available commercially, but high-power devices suitable for illumination are still far too expensive to be mass-marketed. The potential benefits of such a light, if made cheaply, are enormous. Instead of dealing with fragile, hot, gas-filled glass bulbs that burn out relatively quickly and waste most of their energy in the form of heat, consumers would own long-lasting, solid-state interior lights. In automobiles, for example, the LEDs would last the lifetime of the car. And the minimal power demands of LEDs mean that more energy is left in the automobile's battery for all the onboard electronic devices.

Society as a whole could benefit as well. Lighting represents about 20 to 30 percent of the U.S. electrical use, and even the best standard illumination systems convert no more than about 25 percent of electricity into light. If white LEDs could be made to match the efficiency of today's red LEDs, they could reduce energy needs and cut the amount of carbon dioxide pumped into the air by electrical generating plants by 300 megatons a year.

The first company to mass-produce affordable high-brightness white LEDs stands to capture an estimated $12-billion worldwide market for illumination lighting. That's why the big three players of lighting—Philips, Osram Sylvania and General Electric—are spending so much on LED research and development and why newer companies are springing up, such as LumiLeds, a joint venture between Philips and Agilent Technologies, for which one of us—Craford—is chief technology officer.

Low-power white LEDs are already used for cell-phone backlights and pedestrian walk signals. Second-generation,

LEDs Light the Deep

When marine biologist Greg Marshall of National Geographic Television wanted to film deep-diving animals such as sperm whales, he faced several problems. These creatures can plunge thousands of feet below the surface, to where it is virtually pitch-black and the pressures are enormous. Further compounding the situation is that any kind of visible lighting would affect his subjects' behavior, attracting or repelling them and their prey. By causing the whales to act abnormally, the standard underwater light would defeat the entire purpose of his project.

The solution: compact LEDs that emit light at the near-infrared wavelength, making for a light that the videotape can "see" but the animal cannot. When placed inside a hardy, torpedo-shaped metal cylinder containing an automatic camera, the devices act as invisible headlights, illuminating objects two or three meters away without altering the whales' behavior.

SPERM WHALES from Crittercam, in a view lit by LEDs.

GREG MARSHALL *National Geographic Television*

The small size of the LEDs meant that there was room to cram other equipment inside, such as devices to record audio, time of day, depth and duration of dive, direction, temperature and velocity. Dubbed Crittercam, the whole automatic camera package is small enough to be placed on the backs of whales, seals, dolphins, penguins, sea lions and other marine creatures, giving scientists a whale's-eye view of the sea. After filming, a time-release device in the harness lets go of the Crittercam, allowing it to bob to the surface where it can be retrieved, along with its precious footage.

Crittercam is the result of 14 years of experimentation, much of which was on suitable light sources. When the first blue LEDs came out commercially about four years ago—at $40 apiece—Marshall was one of the first to own one, which he purchased directly from the factory in Japan. (Unfortunately, he found that whereas blue-wavelength LEDs penetrated the deep ocean's gloom most efficiently, their light was too visible to his swimming subjects. Near-infrared became the choice instead. In other words, wavelength selection matters, a big advantage when it comes to LEDs.)

Marshall's effort paid off. Watching the footage from a Crittercam mounted to the dorsal fin of a shark is as exciting as watching the chase scene in a spy movie. And the camera has enabled marine biologists to observe behaviors never seen before, such as seals blowing bubbles and "singing" undersea as they perform courtship rituals. *—The Editors*

COLOR KINETICS (*top left*); LUMILEDS LIGHTING

CURRENT LED APPLICATIONS include (*clockwise, from top left*): exhibit lighting, such as that used on the Beatles' *Sgt. Pepper's* costumes in New York City's Metropolitan Museum of Art; traffic lights; auto taillights and turn signals; and outdoor signs.

higher-power LEDs suitable for, say, landscape and accent lighting are becoming available. But large-scale replacement of lamps for general-purpose illumination are not expected for a decade or two because of the difficulty in making white LEDs efficient and cost-competitive.

There are two main ways to generate white light from LEDs. One way is to combine the output of LEDs at the red, green and blue wavelengths, based on the additive principle of color theory. The problem with this technique is that it is difficult to mix the colors of the LEDs efficiently with good uniformity and control.

The second way relies on an LED photon to excite a phosphor. For example, one can package a yellow phosphor around a blue LED. When the energy of the LED strikes the phosphor, it becomes excited and gives off yellow light, which mixes with the blue light from the LED to give white light. Alternatively, one can use an ultraviolet LED to excite a mixture of red, green and blue phosphors to give white light. This process, similar to that in fluorescent tubes, is simpler than mixing three colors but is inherently less efficient, because energy is lost in converting ultraviolet or blue light into lower-energy light (that is, light toward the red end of the spectrum). Moreover, light is also lost because of scattering and absorption in the phosphor packaging.

In any case, the high cost of LED chips and packages currently makes both approaches prohibitive for illumination applications. The best commercial white LEDs now cost about 50 cents per lumen, compared with a fraction of a penny per lumen for a typical incandescent bulb.

Whichever method is chosen—and both will most likely be important for different applications—the key issues are to reduce production costs substantially and to improve performance. Still, it may be a while before consumers accept LEDs, which cost more up front but are cheaper over the span of a decade. As energy prices rise and the consequences of global warming become more urgent, LEDs should become more attractive. One solution to our energy and environmental problems, it seems, may soon come to light. SA

The Authors

M. GEORGE CRAFORD, NICK HOLONYAK, JR., and FREDERICK A. KISH, JR., have all won awards for their work on light-emitting diodes. Craford, who made the first yellow LED, managed the LED technology groups at Monsanto and Hewlett-Packard before becoming chief technology officer at LumiLeds Lighting in San Jose, Calif., a firm jointly created by Philips and Agilent Technologies to seek emerging LED applications. Holonyak, professor of electrical and computer engineering and physics at the University of Illinois, is credited as being the inventor of the first practical LED: the red gallium arsenide phosphide LED. He was also two-time Nobelist John Bardeen's first Ph.D. student. Kish is R&D and manufacturing department manager at Agilent Technologies. There he was one of the primary instigators of a new family of high-brightness red-orange-yellow LEDs, which were the first LEDs to exceed the efficiency of unfiltered incandescent bulbs and are now the dominant technology for traffic signals and exterior lights on automobiles. Both Craford and Kish were graduate students in Holonyak's laboratory.

Further Information

WE WERE BURNING: JAPANESE ENTREPRENEURS AND THE FORGING OF THE ELECTRONIC AGE. Bob Johnstone. Basic Books, 1999.

REPLACING THE LIGHT BULB. Charles T. Whipple in *Photonics Spectra,* Vol. 33, No. 12, pages 104–109; December 1999.

LEDS LIGHT THE FUTURE. Neil Savage in *Technology Review*, Vol. 103, No. 5, pages 38–44; September/October 2000.

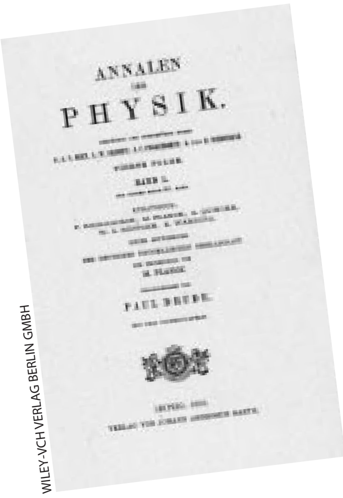
Advanced Illumination Web site: www.advancedillumination.com

Color Kinetics Web site: www.colorkinetics.com

LumiLeds Lighting Web site: www.lumileds.com

100 Years of
QUANTUM MYSTERIES

by Max Tegmark and John Archibald Wheeler

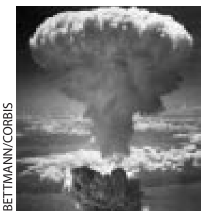


WILEY-VCH VERLAG BERLIN GMBH

- Planck Explains Blackbody Radiation (1900)
- Einstein Explains Photoelectric Effect (1905)
- Discovery of Superconductivity (1911)
- Bohr's Theory of Atomic Spectra (1913)
- Bose-Einstein Condensation Predicted (1924)
- Pauli Exclusion Principle (1925)
- Schrödinger Equation; Copenhagen Interpretation (1926)
- Heisenberg Uncertainty Principle (1927)
- Dirac Equation for the Electron (1928)
- Anti-Electron Discovered (1932)
- Schrödinger's Cat Paper; Einstein-Podolsky-Rosen Paper about Local Realism (1935)
- Superfluidity Discovered (1938)
- Atomic Bomb (1945)
- Transistor (1947)
- Quantum Electrodynamics and Renormalization (1948)

1900s | 1910s | 1920s | 1930s | 1940s

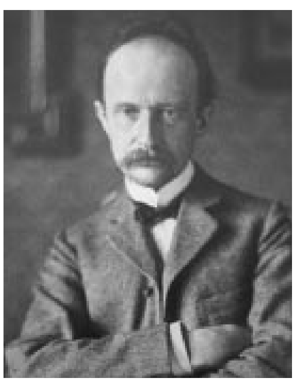
TEXAS CENTER FOR SUPERCONDUCTIVITY, UNIVERSITY OF HOUSTON



BETTMANN/CORBIS

AT&T ARCHIVES

FOUNDATIONS of quantum mechanics were laid in the period 1900–1926, including seminal contributions from the seven physicists shown at the right. Over its century of development, quantum mechanics has not only profoundly advanced our understanding of nature but has also provided the basis of numerous technologies. Yet some fundamental enigmas of quantum theory remain unresolved.

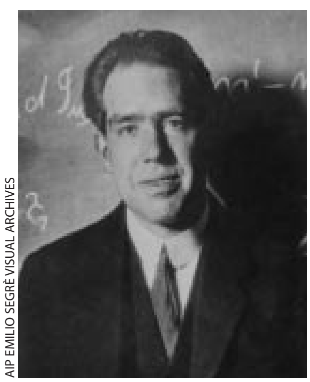


BETTMANN/CORBIS

MAX PLANCK
(1858–1947)

THE ALBERT EINSTEIN ARCHIVES, HEBREW UNIVERSITY OF JERUSALEM

ALBERT EINSTEIN
(1879–1955)



AIP EMILIO SEGRÈ VISUAL ARCHIVES

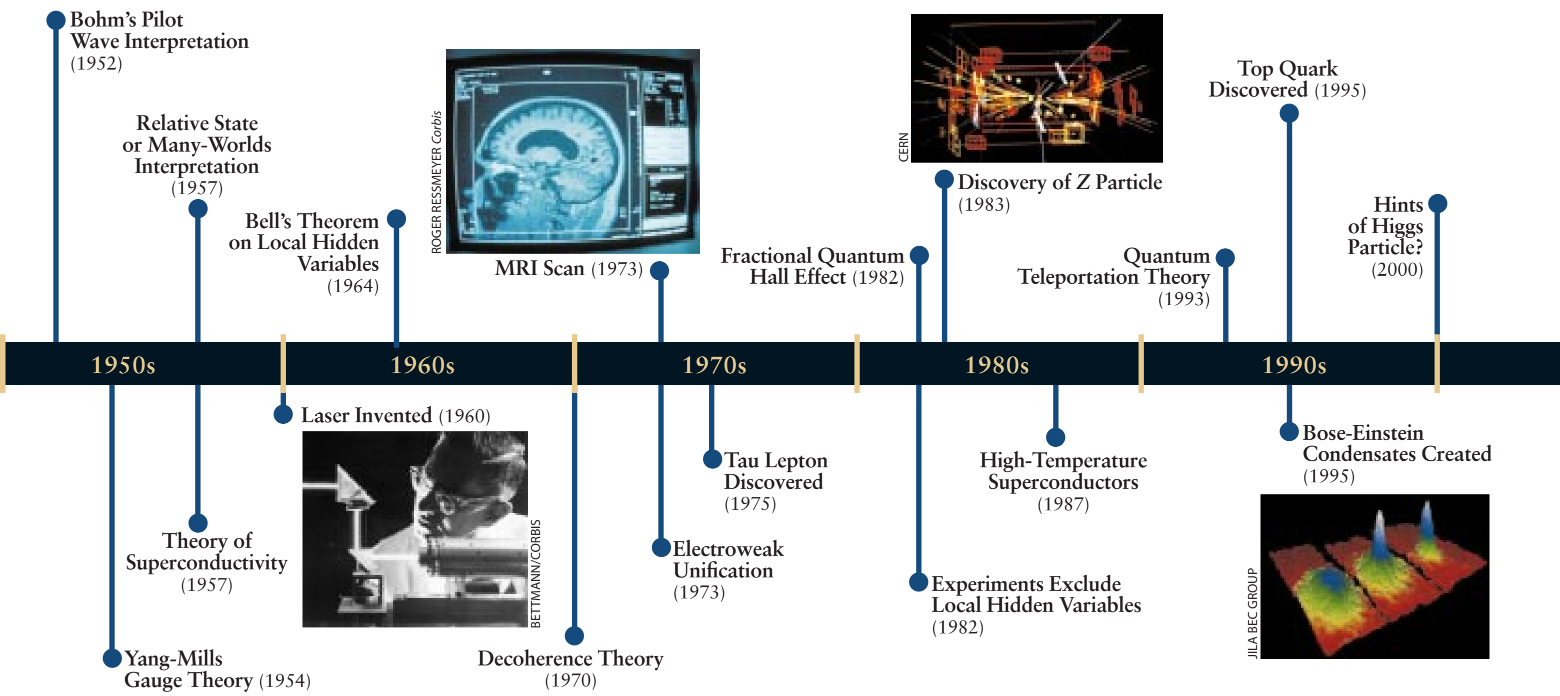
NIELS BOHR
(1885–1962)

As quantum theory celebrates its 100th birthday, spectacular successes are mixed with persistent puzzles

"In a few years, all the great physical constants will have been approximately estimated, and . . . the only occupation which will then be left to the men of science will be to carry these measurements to another place of decimals." As we enter the 21st century amid much brouhaha about past achievements, this sentiment may sound familiar. Yet the quote is from James Clerk Maxwell and dates from his 1871 University of Cambridge inaugural lecture expressing the mood prevalent at the time (albeit a mood he disagreed with). Three decades later, on December 14, 1900, Max Planck announced his formula for the blackbody spectrum, the first shot of the quantum revolution.

This article reviews the first 100 years of quantum mechanics, with particular focus on its mysterious side, culminating in the ongoing debate about its consequences for issues ranging from quantum computation to consciousness, parallel universes and the very nature of physical reality. We virtually ignore the astonishing range of scientific and practical applications that quantum mechanics undergirds: today an estimated 30 percent of the U.S. gross national product is based on inventions made possible by quantum mechanics, from semiconductors in computer chips to lasers in compact-disc players, magnetic resonance imaging in hospitals, and much more.

In 1871 scientists had good reason for their optimism. Classical mechanics and electrodynamics had powered the industrial revolution, and it appeared as though



LOUIS DE BROGLIE
(1892–1987)

ERWIN SCHRÖDINGER
(1887–1961)

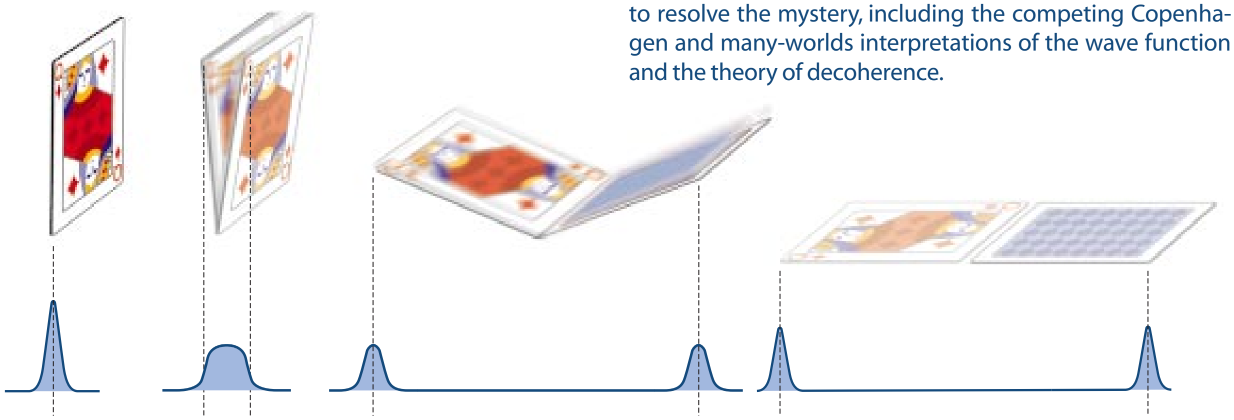
MAX BORN
(1882–1970)

WERNER HEISENBERG
(1901–1976)

QUANTUM CARDS

A SIMPLE FALLING CARD IN PRINCIPLE LEADS TO A QUANTUM MYSTERY

According to quantum physics, an ideal card perfectly balanced on its edge will fall down in both directions at once, in what is known as a superposition. The card's quantum wave function (*blue*) changes smoothly and continuously from the balanced state (*left*) to the mysterious final state (*right*) that seems to have the card in two places at once. In practice, this experiment is impossible with a real card, but the analogous situation has been demonstrated innumerable times with electrons, atoms and larger objects. Understanding the meaning of such superpositions, and why we never see them in the everyday world around us, has been an enduring mystery at the very heart of quantum mechanics. Over the decades, physicists have developed several ideas to resolve the mystery, including the competing Copenhagen and many-worlds interpretations of the wave function and the theory of decoherence.



LAURIE GRACE

their basic equations could describe essentially all physical systems. But a few annoying details tarnished this picture. For example, the calculated spectrum of light emitted by a glowing hot object did not come out right. In fact, the classical prediction was called the ultraviolet catastrophe, according to which intense ultraviolet radiation and x-rays should blind you when you look at the heating element on a stove.

The Hydrogen Disaster

In his 1900 paper Planck succeeded in deriving the correct spectrum. His derivation, however, involved an assumption so bizarre that he distanced himself from it for many years afterward: that energy was emitted only in certain finite chunks, or "quanta." Yet this strange assumption proved extremely successful. In 1905 Albert Einstein took the idea one step further. By assuming that radiation could transport energy only in such lumps, or "photons," he explained the photoelectric effect, which is related to the processes used in present-day solar cells and the image sensors used in digital cameras.

Physics faced another great embarrassment in 1911. Ernest Rutherford had convincingly argued that atoms consist of electrons orbiting a positively charged nucleus, much like a miniature solar system. Electromagnetic theory, though, predicted that orbiting electrons would continuously radiate away their energy and spiral into the nucleus in about a trillionth of a second. Of course, hydrogen atoms were known to be eminently stable. Indeed, this discrepancy was the worst quantitative failure in the history of physics—underpredicting the lifetime of hydrogen by some 40 orders of magnitude.

In 1913 Niels Bohr, who had come to the University of Manchester in England to work with Rutherford, provided an explanation that again used quanta. He postulated that the electrons' angular momentum came only in specific amounts, which would confine them to a discrete set of orbits. The electrons could radiate energy only by jumping from one such orbit to a lower one and sending off an individual photon. Because an electron in the innermost orbit had no orbits with less energy to jump to, it formed a stable atom.

Bohr's theory also explained many of hydrogen's spectral lines—the specific frequencies of light emitted by excited atoms. It worked for the helium atom as well, but only if the atom was deprived of one of its two electrons. Back in Copenhagen, Bohr got a letter from Rutherford telling him he had to publish his results. Bohr wrote back that nobody would believe him unless he explained the spectra of all the elements. Rutherford replied: Bohr, you explain hydrogen and you explain helium, and everyone will believe all the rest.

Despite the early successes of the quantum idea, physicists still did not know what to make of its strange and seemingly ad hoc rules. There appeared to be no guiding principle. In 1923 Louis de Broglie proposed an answer in his doctoral thesis: electrons and other particles act like standing waves. Such waves, like vibrations of a guitar string, can occur only with certain discrete (quantized) frequencies. The idea was so unusual that the examining committee went outside its circle for advice. Einstein, when queried, gave a favorable opinion, and the thesis was accepted.

In November 1925 Erwin Schrödinger gave a seminar on de Broglie's work in Zurich. When he was finished, Peter Debye asked, You speak about waves, but where is the wave equation? Schrödinger went on to produce his equation, the master key for so much of modern physics. An equivalent formulation using matrices was provided by Max Born, Pascual Jordan and Werner Heisenberg around the same time. With this powerful mathematical underpinning, quantum theory made explosive progress. Within a few years, physicists had explained a host of measurements, including spectra of more complicated atoms and properties of chemical reactions.

But what did it all mean? What was this quantity, the "wave function," that Schrödinger's equation described? This central puzzle of quantum mechanics remains a potent and controversial issue to this day.

Born had the insight that the wave function should be interpreted in terms of probabilities. When experimenters measure the location of an electron, the probability of finding it in each region depends on the magnitude of its wave function there. This interpretation suggested that a fundamental randomness was built into the laws of nature. Einstein was deeply unhappy with this conclusion and expressed his preference for a deterministic universe with the oft-quoted remark, "I can't believe that God plays dice."

Curious Cats and Quantum Cards

Schrödinger was also uneasy. Wave functions could describe combinations of different states, so-called superpositions. For example, an electron could be in a superposition of several different locations. Schrödinger pointed out that if microscopic objects such as atoms could be in strange superpositions, so could macroscopic objects, because they are made of atoms. As a baroque example, he described the now well-known thought experiment in which a nasty contraption kills a cat if a radioactive atom decays. Because the radioactive atom enters a superposition of decayed and not decayed, it produces a cat that is both dead and alive in superposition.

The illustration on the opposite page shows a simpler variant of this thought experiment. You take a card with a perfectly sharp edge and balance it on its edge on a table. According to classical physics, it will in principle stay balanced forever. According to the Schrödinger equation, the card will fall down in a few seconds even if you do the best possible job of balancing it, and it will fall down in both directions—to the left and the right—in superposition.

If you could perform this idealized thought experiment with an actual card, you would undoubtedly find that classical physics is wrong and that the card falls down. But you would always see it fall down to the left *or* to the right, seemingly at random, never to the left and to the right simultaneously, as the Schrödinger equation might have you believe. This seeming contradiction goes to the very heart of one of the original and enduring mysteries of quantum mechanics.

The Copenhagen interpretation of quantum mechanics, which grew from discussions between Bohr and Heisenberg in the late 1920s, addresses the mystery by asserting that observations, or measurements, are special. So long as the balanced card is unobserved, its wave function evolves by obeying the Schrödinger equation—a continuous and smooth evolution that is called "unitary" in mathematics and has several very attractive properties. Unitary evolution produces the superposition in which the card has fallen down both to the left and to the right. The act of observing the card, however, triggers an abrupt change in its wave function, commonly called a collapse: the observer sees the card in one definite classical state (face up or face down), and from then onward only that part of the wave function survives. Nature supposedly selects one state at random, with the probabilities determined by the wave function.

The Copenhagen interpretation provided a strikingly successful recipe for doing calculations that accurately described the outcomes of experiments, but the suspicion lingered that some equation ought to describe when and how this collapse occurred. Many physicists took this lack of an equation to mean that something was intrinsically wrong with quantum mechanics and that it would soon be replaced by a more fundamental theory that would provide such an equation. So rather than dwell on ontological implications of the equations, most physicists forged ahead to work out their many exciting applications and to tackle pressing unsolved problems of nuclear physics.

That pragmatic approach proved stunningly successful. Quantum mechanics was instrumental in predicting antimatter, understanding radioactivity (leading to nuclear power), accounting for the behavior of materials such as semiconductors, explaining superconductivity, and describing interactions such as those between light and matter (leading to the invention of the laser) and of radio waves and nuclei (leading to magnetic resonance imaging). Many successes of quantum mechanics involve its extension, quantum field theory, which forms the foundations of elementary particle physics all the way to the present-day experimental frontiers of neutrino oscillations and the search for the Higgs particle and supersymmetry.

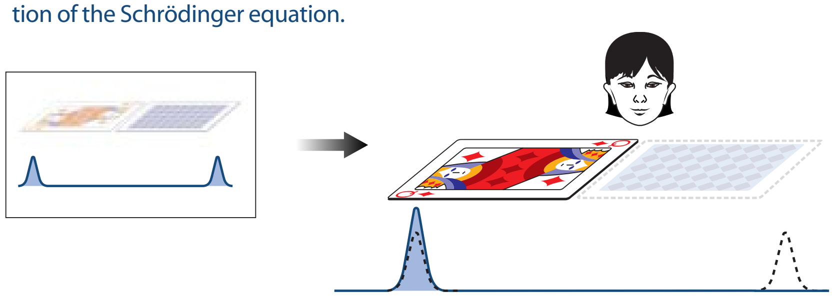
Many Worlds

By the 1950s this ongoing parade of successes had made it abundantly clear that quantum theory was far more than a short-lived temporary fix. And so, in the mid-1950s, a Princeton University student named Hugh Everett III decided to revisit the collapse postulate in his doctoral thesis. Everett pushed the

COPENHAGEN INTERPRETATION

IDEA: Observers see a random outcome; probability given by the wave function.
ADVANTAGE: A single outcome occurs, matching what we observe.
PROBLEM: Requires wave functions to "collapse," but no equation specifies when.

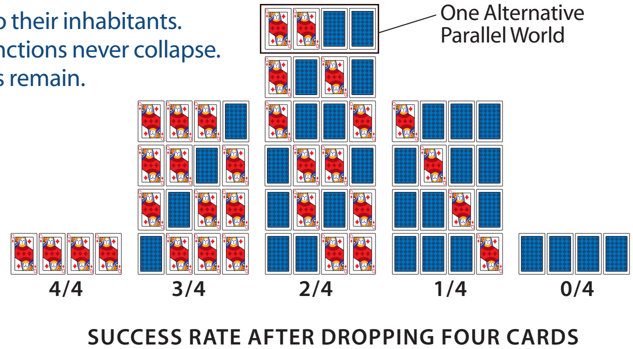
When a quantum superposition is observed or measured, we see one or the other of the alternatives at random, with probabilities controlled by the wave function. If a person has bet that the card will fall face up, when she first looks at the card she has a 50 percent chance of happily seeing that she has won her bet. This interpretation has long been pragmatically accepted by physicists even though it requires the wave function to change abruptly, or collapse, in violation of the Schrödinger equation.



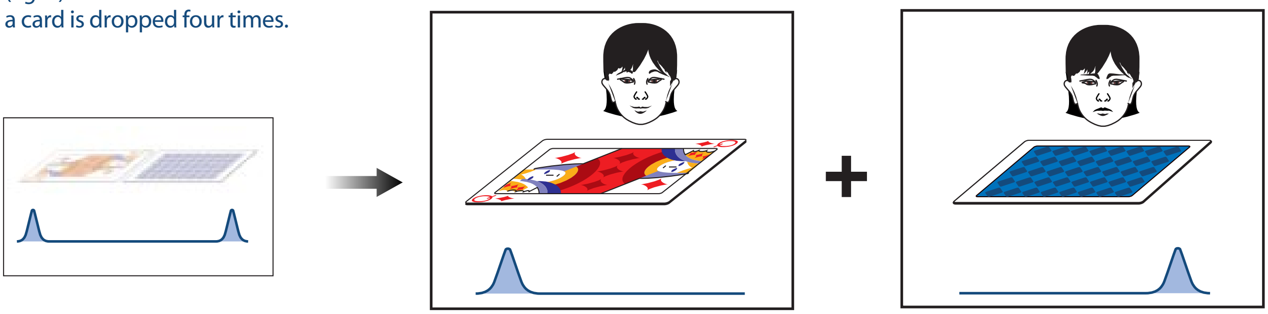
LAURIE GRACE

MANY-WORLDS INTERPRETATION

IDEA: Superpositions will seem like alternative parallel worlds to their inhabitants.
ADVANTAGE: The Schrödinger equation always works: wave functions never collapse.
PROBLEMS: The bizarreness of the idea. Some technical puzzles remain.



If wave functions never collapse, the Schrödinger equation predicts that the person looking at the card's superposition will herself enter a superposition of two possible outcomes: happily winning the bet or sadly losing. These two parts of the total wave function (of person plus card) carry on completely independently, like two parallel worlds. If the experiment is repeated many times, people in most of the parallel worlds will see the card falling face up about half the time. Stacked cards (*right*) show 16 worlds that result when a card is dropped four times.



LAURIE GRACE

quantum idea to its extreme by asking the following question: What if the time evolution of the entire universe is always unitary? After all, if quantum mechanics suffices to describe the universe, then the present state of the universe is described by a wave function (an extraordinarily complicated one). In Everett's scenario, that wave function would always evolve in a deterministic way, leaving no room for mysterious nonunitary collapse or God playing dice.

Instead of being collapsed by measurements, microscopic superpositions would rapidly get amplified into byzantine macroscopic superpositions. Our quantum card would really be in two places at once. Moreover, a person looking at the card would enter a superposition of two different mental states, each perceiving one of the two outcomes. If you had bet money on the queen's landing face up, you would end up in a superposition of smiling and frowning. Everett's brilliant insight was that the observers in such a deterministic but schizophrenic quantum world could perceive the plain old reality that we are familiar with. Most important, they could perceive an apparent randomness obeying the correct probability rules [*see illustration above*].

Everett's viewpoint, formally called the relative-state formulation, became popularly known as the many-worlds interpretation of quantum mechanics, because each component of one's superposition perceives its own world. This viewpoint simplifies the underlying theory by removing the collapse postulate. But the price it pays for this simplicity is the conclusion that these parallel perceptions of reality are all equally real.

Everett's work was largely disregarded for about two decades. Many physicists still hoped that a deeper theory would be discovered, showing that the world was in some sense classical after all, free from oddities like big objects being in two places at once. But such hopes were shattered by a series of new experiments.

Could the seeming quantum randomness be replaced by some kind of unknown quantity carried about inside particles—so-called hidden variables? CERN theorist John S. Bell showed that in this case quantities that could be measured in certain difficult experiments would inevitably disagree with the standard quantum predictions. After many years, technology allowed researchers to conduct the experiments and to eliminate hidden variables as a possibility.

A "delayed choice" experiment proposed by one of us (Wheeler) in 1978 was successfully carried out in 1984, showing another quantum feature of the world that defies classical descriptions: not only can a photon be in two places at once, but experimenters can choose, after the fact, whether the photon was in both places or just one.

The simple double-slit interference experiment, in which light or electrons pass through two slits and produce an interference pattern, hailed by Richard Feynman as the mother of all quantum effects, was successfully repeated for ever larger objects: atoms, small molecules and, most recently, 60-atom buckyballs. After this last feat, Anton Zeilinger's group in Vienna even started discussing conducting the experiment with a virus. In short, the experimental verdict is in: the weirdness of the quantum world is real, whether we like it or not.

Quantum Censorship—Decoherence

The experimental progress of the past few decades was paralleled by great advances in theoretical understanding. Everett's work had left two crucial questions unanswered. First, if the world actually contains bizarre macroscopic superpositions, why don't we perceive them?

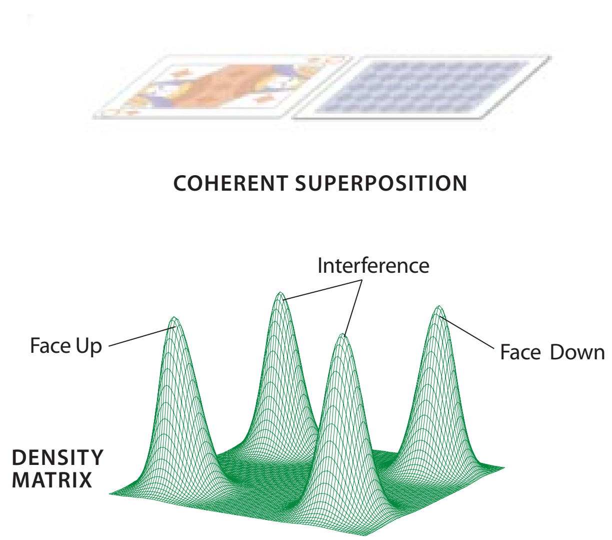
The answer came in 1970 with a seminal paper by H. Dieter Zeh of the University of Heidelberg, who showed that the Schrödinger equation itself gives rise

DECOHERENCE: HOW THE QUANTUM GETS CLASSICAL

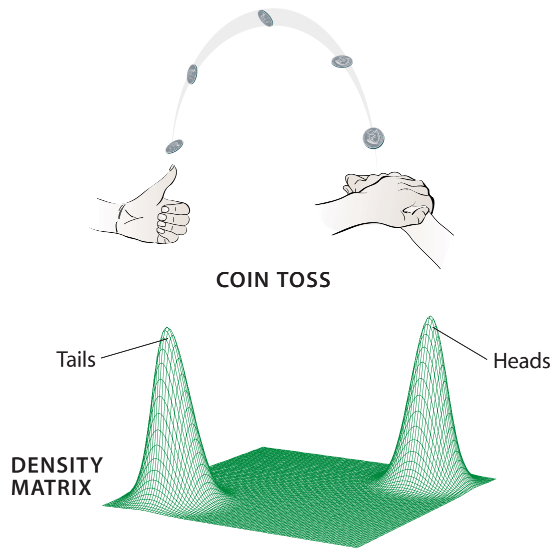
IDEA: Tiny interactions with the surrounding environment rapidly dissipate the peculiar quantumness of superpositions.
ADVANTAGES: Experimentally testable. Explains why the everyday world looks "classical" instead of quantum.
CAVEAT: Decoherence does not completely eliminate the need for an interpretation such as many-worlds or Copenhagen.

The uncertainty of a quantum superposition (*left*) is different from the uncertainty of classical probability, as occurs after a coin toss (*right*). A mathematical object called a density matrix illustrates the distinction. The wave function of the quantum card corresponds to a density matrix with four peaks. Two of these peaks represent the 50 percent probability of each outcome, face up or face down. The other two indicate that these two outcomes can still, in principle, interfere with each other. The quantum state is still "coherent." The density matrix of a coin toss has only the first two peaks, which conventionally means that the coin is really either face up or face down but that we just haven't looked at it yet.

QUANTUM UNCERTAINTY

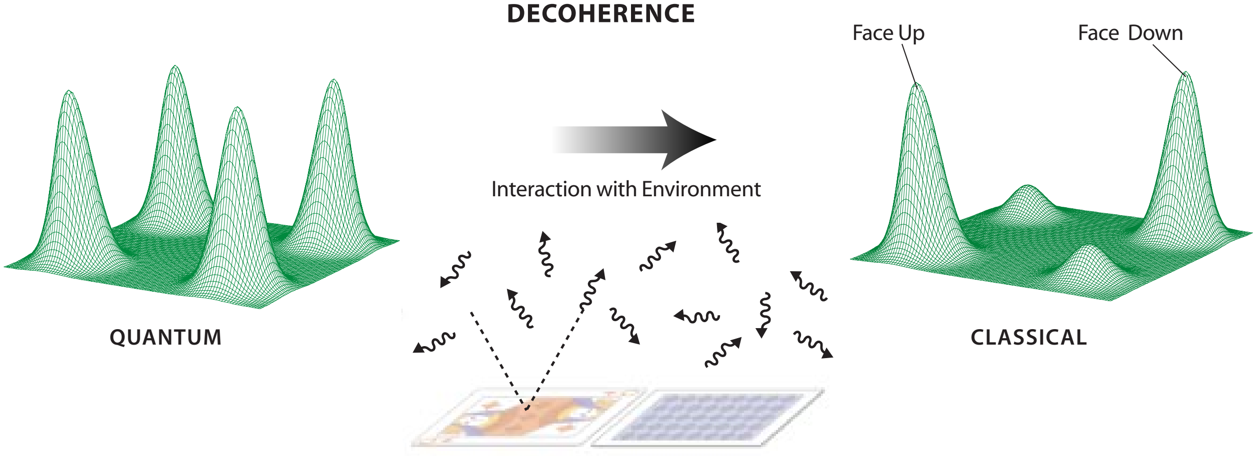


CLASSICAL UNCERTAINTY



Decoherence theory reveals that the tiniest interaction with the environment, such as a single photon or gas molecule bouncing off the fallen card, transforms a coherent density matrix very rapidly into one that, for all practical purposes, represents classical probabilities such as those in a coin toss. The Schrödinger equation controls the entire process.

DECOHERENCE



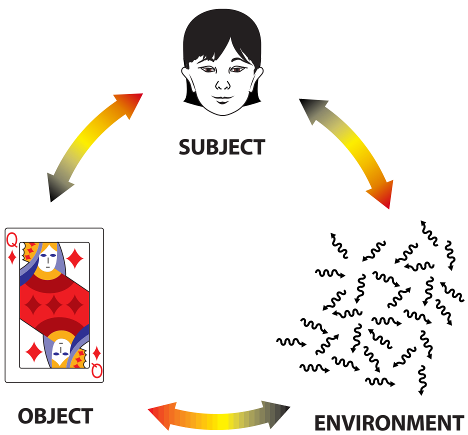
LAURIE GRACE

SPLITTING REALITY

It is instructive to split the universe into three parts: the object under consideration, the environment, and the quantum state of the observer, or subject. The Schrödinger equation that governs the universe as a whole can be divided into terms that describe the internal dynamics of each of these three subsystems and terms that describe interactions among them. These terms have qualitatively very different effects.

The term giving the object's dynamics is typically the most important one, so to figure out what the object will do, theorists can usually begin by ignoring all the other terms. For our quantum card, its dynamics predict that it will fall both left and right in superposition. When our observer looks at the card, the subject-object interaction extends the superposition to her mental state, producing a superposition of joy and disappointment over winning and losing her bet. She can never perceive this superposition, however, because the interaction between the object and the environment (such as air molecules and photons bouncing off the card) causes rapid decoherence that makes this superposition unobservable.

Even if she could completely isolate the card from the environment (for example, by doing the experiment in a dark vacuum chamber at absolute zero), it would not make any difference. At least one neuron in her optical nerves would enter a superposition of firing and not firing when she looked at the card, and this superposition would decohere in about 10^{-20} second, according to recent calculations. If the complex patterns of neuron firing in our brains have anything to do with consciousness and how we form our thoughts and perceptions, then decoherence of our neurons ensures that we never perceive quantum superpositions of mental states. In essence, our brains inextricably interweave the subject and the environment, forcing decoherence on us. *—M.T. and J.A.W.*



to a type of censorship. This effect became known as decoherence, because an ideal pristine superposition is said to be coherent. Decoherence was worked out in great detail by Los Alamos scientist Wojciech H. Zurek, Zeh and others over the following decades. They found that coherent superpositions persist only as long as they remain secret from the rest of the world. Our fallen quantum card is constantly bumped by snooping air molecules and photons, which thereby find out whether it has fallen to the left or to the right, destroying ("decohering") the superposition and making it unobservable [*see box on preceding page*].

It is almost as if the environment acts as an observer, collapsing the wave function. Suppose that your friend looked at the card without telling you the outcome. According to the Copenhagen interpretation, her measurement collapses the superposition into a definite outcome, and your best description of the card changes from a quantum superposition to a classical representation of your ignorance of what she saw. Loosely speaking, decoherence calculations show that you do not need a human observer (or explicit wave-function collapse) to get much the same effect—even an air molecule bouncing off the fallen card will suffice. That tiny interaction rapidly changes the superposition to a classical situation for all practical purposes.

Decoherence explains why we do not routinely see quantum superpositions in the world around us. It is not because quantum mechanics intrinsically stops working for objects larger than some magic size. Instead macroscopic objects such as cats and cards are almost impossible to keep isolated to the extent needed to prevent decoherence. Microscopic objects, in contrast, are more easily isolated from their surroundings so that they retain their quantum behavior.

The second unanswered question in the Everett picture was more subtle but equally important: What mechanism picks out the classical states—face up and face down for our card—as special? Considered as abstract quantum states, there is nothing special about these states as compared to the innumerable possible superpositions of up and down in various proportions. Why do the many worlds split strictly along the up/down lines that we are familiar with and never any of the other alternatives? Decoherence answered this question as well. The calculations showed that classical states such as face up and face down were precisely the ones that are robust against decoherence. That is, interactions with the surrounding environment would leave face-up and face-down cards unharmed but would drive any superposition of up and down into classical face-up/face-down alternatives.

Decoherence and the Brain

Physicists have a tradition of analyzing the universe by splitting it into two parts. For example, in thermodynamics, theorists may separate a body of matter from everything else around it (the "environment"), which may supply prevailing conditions of temperature and pressure. Quantum physics traditionally separates the quantum system from the classical measuring apparatus. If unitarity and decoherence are taken seriously, then it is instructive to split the universe into three parts, each described by quantum states: the object under consideration, the environment, and the observer, or subject [*see box at left*].

Decoherence caused by the environment interacting with the object or the subject ensures that we never perceive quantum superpositions of mental states. Furthermore, our brains are inextricably interwoven with the environment, and decoherence of our firing neurons is unavoidable and essentially instantaneous. As Zeh has emphasized, these conclusions justify the long tradition of using the textbook postulate of wave-function collapse as a pragmatic "shut up and calculate" recipe: compute probabilities as if the wave function collapses when the object is observed. Even though in the Everett view the wave function technically never collapses, decoherence researchers generally agree that decoherence produces an effect that looks and smells like a collapse.

LAURIE GRACE

The discovery of decoherence, combined with the ever more elaborate experimental demonstrations of quantum weirdness, has caused a noticeable shift in the views of physicists. The main motivation for introducing the notion of wave-function collapse had been to explain why experiments produced specific outcomes and not strange superpositions of outcomes. Now much of that motivation is gone. Moreover, it is embarrassing that nobody has provided a testable deterministic equation specifying precisely when the mysterious collapse is supposed to occur.

An informal poll taken in July 1999 at a conference on quantum computation at the Isaac Newton Institute in Cambridge, England, suggests that the prevailing viewpoint is shifting. Out of 90 physicists polled, only eight declared that their view involved explicit wave-function collapse. Thirty chose "many worlds or consistent histories (with no collapse)." (Roughly speaking, the consistent-histories approach analyzes sequences of measurements and collects together bundles of alternative results that would form a consistent "history" to an observer.)

But the picture is not clear: 50 of the researchers chose "none of the above or undecided." Rampant linguistic confusion may contribute to that large number. It is not uncommon for two physicists who say that they subscribe to the Copenhagen interpretation, for example, to find themselves disagreeing about what they mean.

This said, the poll clearly suggests that it is time to update the quantum textbooks: although these books, in an early chapter, infallibly list explicit nonunitary collapse as a fundamental postulate, the poll indicates that today many physicists—at least in the burgeoning field of quantum computation—do not take this seriously. The notion of collapse will undoubtedly retain great utility as a calculational recipe, but an added caveat clarifying that it is probably not a fundamental process violating the Schrödinger equation could save astute students many hours of confusion.

Looking Ahead

After 100 years of quantum ideas, what lies ahead? What mysteries remain? How come the quantum? Although basic issues of ontology and the ultimate nature of reality often crop up in discussions about how to interpret quantum mechanics, the theory is probably just a piece in a larger puzzle. Theories can be crudely organized in a family tree where each might, at least in principle, be derived from more fundamental ones above it. Almost at the top of the tree lie general relativity and quantum field theory. The first level of descendants includes special relativity and quantum mechanics, which in turn spawn electromagnetism, classical mechanics, atomic physics, and so on. Disciplines such as computer science, psychology and medicine appear far down in the lineage.

All these theories have two components: mathematical equations and words that explain how the equations are connected to what is observed in experiments. Quantum mechanics as usually presented in textbooks has both components: some equations and three fundamental postulates written out in plain English. At each level in the hierarchy of theories, new concepts (for example, protons, atoms, cells, organisms, cultures) are introduced because they are convenient, capturing the essence of what is going on without recourse to the theories above it. Crudely speaking, the ratio of equations to words decreases as one moves down the tree, dropping near zero for very applied fields such as medicine and sociology. In contrast, theories near the top are highly mathematical, and physicists are still struggling to comprehend the concepts that are encoded in the mathematics.

The ultimate goal of physics is to find what is jocularly referred to as a theory of everything, from which all else can be derived. If such a theory exists, it would take the top spot in the family tree, indicating that both general relativity and quantum field theory could be derived from it. Physicists know something is missing at the top of the tree, because we lack a consistent theory that includes both gravity and quantum mechanics, yet the universe contains both phenomena.

A theory of everything would probably have to contain no concepts at all. Otherwise one would very likely seek an explanation of its concepts in terms of a still more fundamental theory, and so on in an infinite regress. In other words, the theory would have to be purely mathematical, with no explanations or postulates. Rather an infinitely intelligent mathematician should be able to derive the entire theory tree from the equations alone, by deriving the properties of the universe that they describe and the properties of its inhabitants and their perceptions of the world.

The first 100 years of quantum mechanics have provided powerful technologies and answered many questions. But physics has raised new questions that are just as important as those outstanding at the time of Maxwell's inaugural speech—questions regarding both quantum gravity and the ultimate nature of reality. If history is anything to go by, the coming century should be full of exciting surprises. SA

The Authors

MAX TEGMARK and JOHN ARCHIBALD WHEELER discussed quantum mechanics extensively during Tegmark's three and a half years as a postdoc at the Institute for Advanced Studies in Princeton, N.J. Tegmark is now an assistant professor of physics at the University of Pennsylvania. Wheeler is professor emeritus of physics at Princeton, where his graduate students included Richard Feynman and Hugh Everett III (inventor of the many-worlds interpretation). He received the 1997 Wolf Prize in physics for his work on nuclear reactions, quantum mechanics and black holes. In 1934 and 1935 Wheeler had the privilege of working on nuclear physics in Niels Bohr's group in Copenhagen. On arrival at the institute he asked a workman who was trimming vines running up a wall where he could find Bohr. "I'm Niels Bohr," the man replied.

The authors wish to thank Emily Bennett and Ken Ford for their help with an earlier manuscript on this topic and Jeff Klein, Dieter Zeh and Wojciech H. Zurek for their helpful comments.

Further Information

ONE HUNDRED YEARS OF QUANTUM PHYSICS. Daniel Kleppner and Roman Jackiw in *Science,* Vol. 289, pages 893–898; August 11, 2000.

BEAM LINE. Special quantum century issue. Vol. 30, No. 2; Summer/Fall 2000. Available online at www.slac.stanford.edu/pubs/beamline/pdf/00ii.pdf

MAX PLANCK: THE RELUCTANT REVOLUTIONARY. Helge Kragh in *Physics World,* Vol. 13, No. 12, pages 31–35; December 2000.

THE QUANTUM CENTENNIAL. A. Zeilinger in *Nature,* Vol. 408, pages 639–641; December 7, 2000.

For more on decoherence, see www.decoherence.de

Salespeople, politicians, friends and family all have a stake in getting you to agree to their requests. Social psychology has determined the basic principles that govern getting to "yes"

The Science of Persu

by Robert B. Cialdini

Hello there.

I hope you've enjoyed the magazine so far. Now I'd like to let you in on something of great importance to you personally. Have you ever been tricked into saying yes? Ever felt trapped into buying something you didn't really want or contributing to some suspicious-sounding cause? And have you ever wished you understood why you acted in this way so that you could withstand these clever ploys in the future?

Yes? Then clearly this article is just right for you. It contains valuable information on the most powerful psychological pressures that get you to say yes to requests. And it's chock-full of NEW, IMPROVED *research showing exactly how and why these techniques work. So don't delay, just settle in and get the information that, after all, you've already agreed you want.*

The scientific study of the process of social influence has been under way for well over half a century, beginning in earnest with the propaganda, public information and persuasion programs of World War II. Since that time, numerous social scientists have investigated the ways in which one individual can influence another's attitudes and actions. For the past 30 years, I have participated in that endeavor, concentrating primarily on the major factors that bring about a specific form of behavior change—compliance with a request. Six basic tendencies of human behavior come into play in generating a positive response: reciprocation, consistency, social validation, liking, authority and scarcity. As these six tendencies help to govern our business dealings, our societal involvements and our personal relationships, knowledge of the rules of persuasion can truly be thought of as empowerment.

Reciprocation

When the Disabled American Veterans organization mails out requests for contributions, the appeal succeeds only about 18 percent of the time. But when the mailing includes a set of free personalized address labels, the success rate almost doubles, to 35 percent. To understand the effect of the unsolicited gift, we must recognize the reach and power of an essential rule of human conduct: the code of reciprocity.

All societies subscribe to a norm that obligates individuals to repay in kind what they have received. Evolutionary selection pressure has probably entrenched the behavior in social animals such as ourselves. The demands of reciprocity begin to explain the boost in donations to the veterans group. Receiving a gift—unsolicited and perhaps even unwanted—

STEVEN ADAMS *AP Photo*

FREE SAMPLES carry a subtle price tag; they psychologically indebt the consumer to reciprocate. Here shoppers get complimentary tastes of a new product, green ketchup. The samples prime the consumer to return the favor with a purchase. The novel color may also make the product seem scarce, an attractive attribute.

convinced significant numbers of potential donors to return the favor.

Charitable organizations are far from alone in taking this approach: food stores offer free samples, exterminators offer free in-home inspections, health clubs offer free workouts. Customers are thus exposed to the product or service, but they are also indebted. Consumers are not the only ones who fall under the sway of reciprocity. Pharmaceutical companies spend millions of dollars every year to support medical researchers and to provide gifts to individual physicians—activities that may subtly influence researchers' findings and physicians' recommendations. A 1998 study in the *New England Journal of Medicine* found that only 37 percent of researchers who published conclusions critical of the safety of calcium channel blockers had received prior drug company support. Among researchers whose conclusions supported the drugs' safety, however, the number of those who had received free trips, research funding or employment skyrocketed—to 100 percent.

Reciprocity includes more than gifts and favors; it also applies to concessions that people make to one another. For example, assume that you reject my large request, and I then make a concession to you by retreating to a smaller request. You may very well then reciprocate with a concession of your own: agreement with my lesser request. In the mid-1970s my colleagues and I conducted an experiment that clearly illustrates the dynamics of reciprocal concessions. We stopped a random sample of passersby on public walkways and asked if they would volunteer to chaperone juvenile detention center inmates on a day trip to the zoo. As expected, very few complied, only 17 percent.

For another random sample of passersby, however, we began with an even larger request: to serve as an unpaid counselor at the center for two hours per week for the next two years. Everyone in this second sampling rejected the extreme appeal. At that point we offered them a concession. "If you can't do that," we asked, "would you chaperone a group of juvenile detention center inmates on a day trip to the zoo?" Our concession powerfully stimulated return concessions. The compliance rate nearly tripled, to 50 percent, compared with the straightforward zoo-trip request.

Consistency

In 1998 Gordon Sinclair, the owner of a well-known Chicago restaurant, was struggling with a problem that afflicts all restaurateurs. Patrons frequently reserve a table but, without notice, fail to appear. Sinclair solved the problem by asking his receptionist to change two words of what she said to callers requesting reservations. The change dropped his no-call, no-show rate from 30 to 10 percent immediately.

The two words were effective because they commissioned the force of another potent human motivation: the desire to be, and to appear, consistent. The receptionist merely modified her request from "Please call if you have to change your plans" to "Will you please call if you have to change your plans?" At that point, she politely paused and waited for a response. The wait was pivotal because it induced customers to fill the pause with a public commitment. And public commitments, even seemingly minor ones, direct future action.

In another example, Joseph Schwarzwald of Bar-Ilan University in Israel and his co-workers nearly doubled monetary contributions for the handicapped in certain neighborhoods. The key factor: two weeks before asking for contributions, they got residents to sign a petition supporting the handicapped, thus making a public commitment to that same cause.

Social Validation

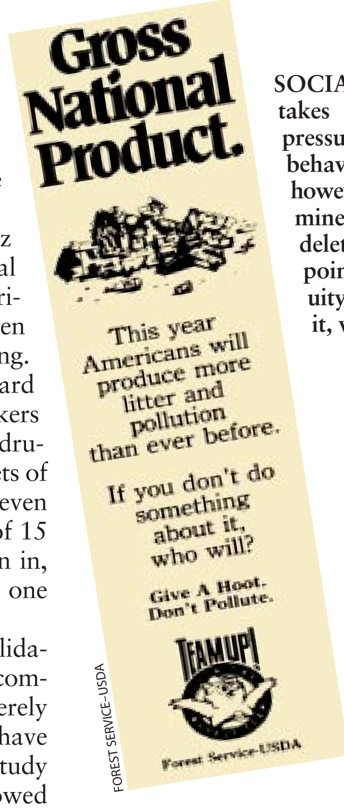
On a wintry morning in the late 1960s, a man stopped on a busy New York City sidewalk and gazed skyward for 60 seconds, at nothing in particular. He did so as part of an experiment by City University of New York social psychologists Stanley Milgram, Leonard Bickman and Lawrence Berkowitz that was designed to find out what effect this action would have on passersby. Most simply detoured or brushed by; 4 percent joined the man in looking up. The experiment was then repeated with a slight change. With the modification, large numbers of pedestrians were induced to come to a halt, crowd together and peer upward.

The single alteration in the experiment incorporated the phenomenon of social validation. One fundamental way that we decide what to do in a situation is to look to what others are doing or have done there. If many individuals have decided in favor of a particular idea, we are more likely to follow, because we perceive the idea to be more correct, more valid.

Milgram, Bickman and Berkowitz introduced the influence of social validation into their street experiment simply by having five men rather than one look up at nothing. With the larger initial set of upward gazers, the percentage of New Yorkers who followed suit more than quadrupled, to 18 percent. Bigger initial sets of planted up-lookers generated an even greater response: a starter group of 15 led 40 percent of passersby to join in, nearly stopping traffic within one minute.

Taking advantage of social validation, requesters can stimulate our compliance by demonstrating (or merely implying) that others just like us have already complied. For example, a study found that a fund-raiser who showed homeowners a list of neighbors who had donated to a local charity significantly increased the frequency of contributions; the longer the list, the greater the effect. Marketers, therefore, go out of their way to inform us when their product is the largest-selling or fastest-growing of its kind, and television commercials regularly depict crowds rushing to stores to acquire the advertised item.

Less obvious, however, are the circumstances under which social validation can backfire to produce the opposite of what a requester intends. An example is the understandable but potentially misguided tendency of health educators to call attention to a problem by depicting it as regrettably frequent. Information campaigns stress that alcohol and drug use is intolerably high, that adolescent suicide rates are alarming and that polluters are spoiling the environment. Although the claims are both true and well intentioned, the creators of these campaigns have missed something basic about the compliance process. Within the statement "Look at all the people who are doing this *undesirable* thing" lurks the powerful and undercutting message "Look at all the people who *are* doing this undesirable thing." Research shows that, as a consequence, many such programs boomerang, generating even more of the undesirable behavior.



FOREST SERVICE–USDA

SOCIAL VALIDATION takes advantage of peer pressure to drive human behavior. Poorly applied, however, it can also undermine attempts to curtail deleterious activities, by pointing out their ubiquity: If everyone's doing it, why shouldn't I?

For instance, a suicide intervention program administered to New Jersey teenagers informed them of the high number of teenage suicides. Health researcher David Shaffer and his colleagues at Columbia University found that participants became significantly more likely to see suicide as a potential solution to their problems. Of greater effectiveness are campaigns that honestly depict the unwanted activity as damaging despite the fact that relatively few individuals engage in it.

Liking

"Affinity," "rapport" and "affection" all describe a feeling of connection between people. But the simple word "liking" most faithfully captures the concept and has become the standard designation in the social science literature. People prefer to say yes to those

PUBLIC COMMITMENT of signing a petition influences the signer to behave consistently with that position in the future.

LUIS M. ALVAREZ *AP Photo*

TUPPERWARE (1958) *AP Photo*

FAMILIAR FACES sell products. Friends (who are already liked) are powerful salespeople, as Tupperware Corporation discovered. Strangers can co-opt the trappings of friendship to encourage compliance.

they like. Consider the worldwide success of the Tupperware Corporation and its "home party" program. Through the in-home demonstration get-together, the company arranges for its customers to buy from a liked friend, the host, rather than from an unknown salesperson. So favorable has been the effect on proceeds that, according to company literature, a Tupperware party begins somewhere in the world every 2.7 seconds. In fact, 75 percent of all Tupperware parties today occur outside the individualistic U.S., in countries where group social bonding is even more important than it is here.

Of course, most commercial transactions take place beyond the homes of friends. Under these much more typical circumstances, those who wish to commission the power of liking employ tactics clustered around certain factors that research has shown to work.

Physical attractiveness can be such a tool. In a 1993 study conducted by Peter H. Reingen of Arizona State University and Jerome B. Kernan of the University of Cincinnati, good-looking fund-raisers for the American Heart Association generated nearly twice as many donations (42 versus 23 percent) as did other requesters. In the 1970s researchers Michael G. Efran and E.W.J. Patterson of the University of Toronto found that voters in Canadian federal elections gave physically attractive candidates several times as many votes as unattractive ones. Yet such voters insisted that their choices would never be influenced by something as superficial as appearance.

Similarity also can expedite a rapport. Salespeople often search for, or outright fabricate, a connection between themselves and their customers: "Well, no kidding, you're from Minneapolis? I went to school in Minnesota!" Fundraisers do the same, with good results. In 1994 psychologists R. Kelly Aune of the University of Hawaii at Manoa and Michael D. Basil of the University of Denver reported research in which solicitors canvassed a college campus asking for contributions to a charity. When the phrase "I'm a student, too" was added to the requests, donations more than doubled.

Compliments also stimulate liking, and direct salespeople are trained in the use of praise. Indeed, even inaccurate praise may be effective. Research at the University of North Carolina at Chapel Hill found that compliments produced just as much liking for the flatterer when they were untrue as when they were genuine.

Cooperation is another factor that has been shown to enhance positive feelings and behavior. Salespeople, for example, often strive to be perceived by their prospects as cooperating partners. Automobile sales managers frequently cast themselves as "villains" so the salesperson can "do battle" on the customer's behalf. The gambit naturally leads to a desirable form of liking by the customer for the salesperson, which promotes sales.

Authority

Recall the man who used social validation to get large numbers of passersby to stop and stare at the sky. He might achieve the opposite effect and spur stationary strangers into motion by assuming the mantle of authority. In 1955 University of Texas at Austin researchers Monroe Lefkowitz, Robert R. Blake and Jane S. Mouton

RIC FIELD *AP Photo*

BEHOLD THE POWER of authority. Certainly not lost on the National Rifle Association is that the authority inherent in such heroic figures as Moses, El Cid and Ben-Hur is linked to the actor who portrayed them, Charlton Heston.

LIMITED OFFER of toys available for a short time often creates a figurative feeding frenzy at local fast-food establishments. Scarcity can be manufactured to make a commodity appear more desirable.

PETER BARRERAS *AP Photo*

discovered that a man could increase by 350 percent the number of pedestrians who would follow him across the street against the light by changing one simple thing. Instead of casual dress, he donned markers of authority: a suit and tie.

Those touting their experience, expertise or scientific credentials may be trying to harness the power of authority: "Babies are our business, our only business," "Four out of five doctors recommend," and so on. (The author's biography at the end of this article in part serves such a purpose.) There is nothing wrong with such claims when they are real, because we usually want the opinions of true authorities. Their insights help us choose quickly and well.

The problem comes when we are subjected to phony claims. If we fail to think, as is often the case when confronted by authority symbols, we can easily be steered in the wrong direction by ersatz experts—those who merely present the aura of legitimacy. That Texas jaywalker in a suit and tie was no more an authority on crossing the street than the rest of the pedestrians who nonetheless followed him. A highly successful ad campaign in the 1970s featured actor Robert Young proclaiming the health benefits of decaffeinated coffee. Young seems to have been able to dispense this medical opinion effectively because he represented, at the time, the nation's most famous physician. That Marcus Welby, M.D., was only a character on a TV show was less important than the appearance of authority.

Scarcity

While at Florida State University in the 1970s, psychologist Stephen West noted an odd occurrence after surveying students about the campus cafeteria cuisine: ratings of the food rose significantly from the week before, even though there had been no change in the menu, food quality or preparation. Instead the shift resulted from an announcement that because of a fire, cafeteria meals would not be available for several weeks.

This account highlights the effect of perceived scarcity on human judgment. A great deal of evidence shows that items and opportunities become more desirable to us as they become less available. For this reason, marketers trumpet the unique benefits or the one-of-a-kind character of their offerings. It is also for this reason that they consistently engage in "limited time only" promotions or put us into competition with one another using sales campaigns based on "limited supply."

Less widely recognized is that scarcity affects the value not only of commodities but of information as well. Information that is exclusive is more persuasive. Take as evidence the dissertation data of a former student of mine, Amram Knishinsky, who owns a company that imports beef into the U.S. and sells it to supermarkets. To examine the effects of scarcity and exclusivity on compliance, he instructed his telephone salespeople to call a randomly selected sample of customers and to make a standard request of them to purchase beef. He also instructed the salespeople to do the same with a second random sample of customers but to add that a shortage of Australian beef was anticipated, which was true, because of certain weather conditions there. The added information that Australian beef was soon to be scarce more than doubled purchases.

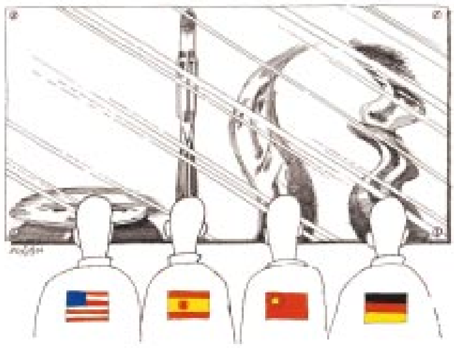
Finally, he had his staff call a third sample of customers, to tell them (1) about the impending shortage of Australian beef and (2) that this information came from his company's *exclusive* sources in the Australian National Weather Service. These customers increased their orders by more than 600 percent. They were influenced by a scarcity double whammy: not only was the beef scarce, but the information that the beef was scarce was itself scarce.

Knowledge Is Power

I think it noteworthy that many of the data presented in this article have come from studies of the practices of persuasion professionals—the marketers, advertisers, salespeople, fund-raisers and their comrades whose financial well-being depends on their ability to get others to say yes. A kind of natural selection operates on these people, as those who use unsuccessful tactics soon go out of business. In contrast, those using procedures that work well will survive, flourish and pass on these successful strategies [see "The Power of Memes," by Susan Blackmore; Scientific American, October 2000]. Thus, over time, the most effective principles of social influence will appear in the repertoires of long-standing persuasion professions. My own work indicates that those principles embody the six fundamental human tendencies examined in this article: reciprocation, consistency, social validation, liking, authority and scarcity.

From an evolutionary point of view, each of the behaviors presented would appear to have been selected for in animals, such as ourselves, that must find the best ways to survive while living in social groups. And in the vast majority of cases, these principles counsel us correctly. It usually makes great sense to repay favors, behave consistently, follow the lead of similar others, favor the

Influence across Cultures



Do the six key factors in the social influence process operate similarly across national boundaries? Yes, but with a wrinkle. The citizens of the world are human, after all, and susceptible to the fundamental tendencies that characterize all members of our species. Cultural norms, traditions and experiences can, however, modify the weight brought to bear by each factor.

Consider the results of a report published this year by Stanford University's Michael W. Morris, Joel M. Podolny and Sheira Ariel, who studied employees of Citibank, a multinational financial corporation. The researchers selected four societies for examination: the U.S., China, Spain and Germany. They surveyed Citibank branches within each country and measured employees' willingness to comply voluntarily with a request from a co-worker for assistance with a task. Although multiple key factors could come into play, the main reason employees felt obligated to comply differed in the four nations. Each of these reasons incorporated a different fundamental principle of social influence.

Employees in the U.S. took a reciprocation-based approach to the decision to comply. They asked the question, "What has this person done for me recently?" and felt obligated to volunteer if they owed the requester a favor. Chinese employees responded primarily to authority, in the form of loyalties to those of high status within their small group. They asked, "Is this requester connected to someone in my unit, especially someone who is high-ranking?" If the answer was yes, they felt required to yield.

Spanish Citibank personnel based the decision mostly on liking/friendship. They were willing to help on the basis of friendship norms that encourage faithfulness to one's friends, regardless of position or status. They asked, "Is this requester connected to my friends?" If the answer was yes, they were especially likely to want to comply.

German employees were most compelled by consistency, offering assistance in order to be consistent with the rules of the organization. They decided whether to comply by asking, "According to official regulations and categories, am I supposed to assist this requester?" If the answer was yes, they felt a strong obligation to grant the request.

In sum, although all human societies seem to play by the same set of influence rules, the weights assigned to the various rules can differ across cultures. Persuasive appeals to audiences in distinct cultures need to take such differences into account. *—R.B.C.*

requests of those we like, heed legitimate authorities and value scarce resources. Consequently, influence agents who use these principles honestly do us a favor. If an advertising agency, for instance, focused an ad campaign on the genuine weight of authoritative, scientific evidence favoring its client's headache product, all the right people would profit—the agency, the manufacturer *and* the audience. Not so, however, if the agency, finding no particular scientific merit in the pain reliever, "smuggles" the authority principle into the situation through ads featuring actors wearing lab coats.

Are we then doomed to be helplessly manipulated by these principles? No. By understanding persuasion techniques, we can begin to recognize strategies and thus truly analyze requests and offerings. Our task must be to hold persuasion professionals accountable for the use of the six powerful motivators and to purchase their products and services, support their political proposals or donate to their causes only when they have acted truthfully in the process.

If we make this vital distinction in our dealings with practitioners of the persuasive arts, we will rarely allow ourselves be tricked into assent. Instead we will give ourselves a much better option: to be informed into saying yes. Moreover, as long as we apply the same distinction to our own attempts to influence others, we can legitimately commission the six principles. In seeking to persuade by pointing to the presence of genuine expertise, growing social validation, pertinent commitments or real opportunities for cooperation, and so on, we serve the interests of both parties and enhance the quality of the social fabric in the bargain.

Surely, someone with your splendid intellect can see the unique benefits of this article. And because you look like a helpful person who would want to share such useful information, let me make a request. Would you buy this issue of the magazine for 10 of your friends? Well, if you can't do that, would you show it to just one friend? Wait, don't answer yet. Because I genuinely like you, I'm going to throw in—at absolutely no extra cost—a set of references that you can consult to learn more about this little-known topic.

Now, will you voice your commitment to help? ... Please recognize that I am pausing politely here. But while I'm waiting, I want you to feel totally assured that many others just like you will certainly consent. And I love *that shirt you're wearing.* SA

The Author

ROBERT B. CIALDINI is Regents' Professor of Psychology at Arizona State University, where he has also been named Distinguished Graduate Research Professor. He has been elected president of the Society of Personality and Social Psychology. Cialdini's book *Influence,* which was the result of a three-year study of the reasons why people comply with requests in everyday settings, has appeared in numerous editions and been published in nine languages. He attributes his long-standing interest in the intricacies of influence to the fact that he was raised in an entirely Italian family, in a predominantly Polish neighborhood, in a historically German city (Milwaukee), in an otherwise rural state.

Further Information

Psychology of Attitude Change and Social Influence. Phillip G. Zimbardo and Michael R. Leippe. Temple University Press, 1991.

Bargaining for Advantage. G. Richard Shell. Viking, 1999.

Age of Propaganda: The Everyday Use and Abuse of Persuasion. Anthony J. Pratkanis. W. H. Freeman and Company, 2000.

Influence: Science and Practice. Fourth edition. Robert B. Cialdini. Allyn & Bacon, 2001.

For regularly updated information about the social influence process, visit www.influenceatwork.com

DUSAN PETRICIC

FLU_VACCINE

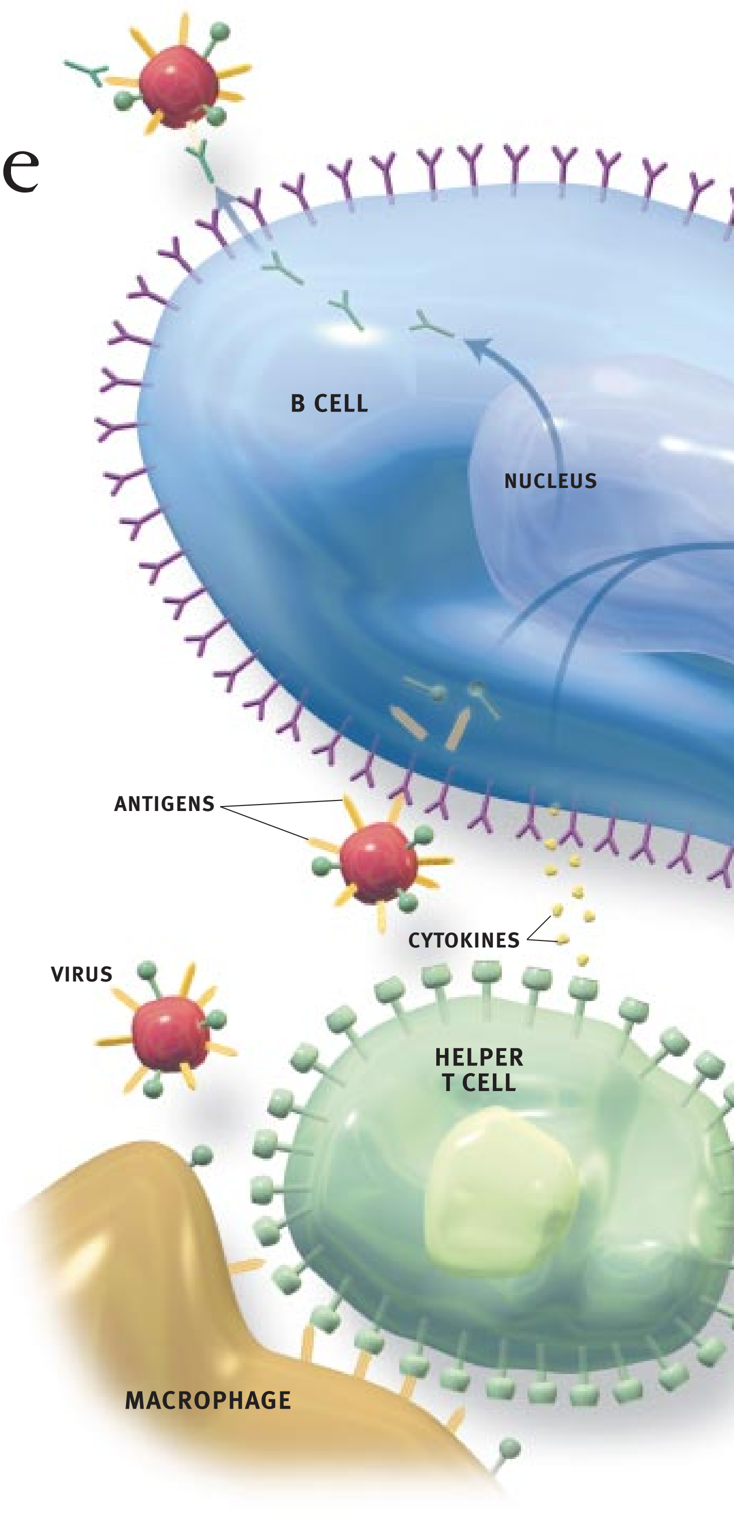
Preparing for Battle

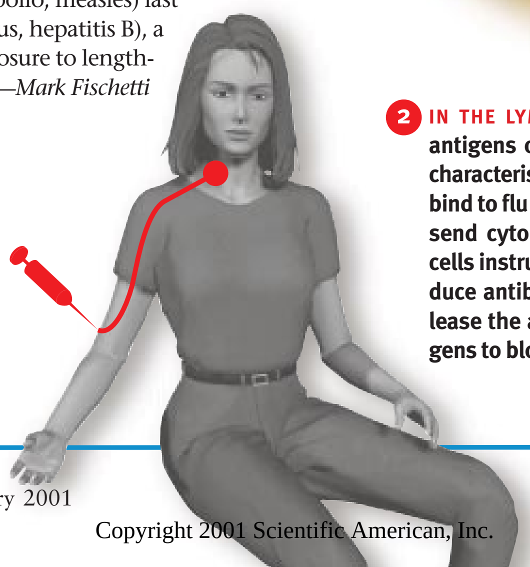
Every year influenza contributes to the death of 20,000 people in the U.S. and perhaps millions worldwide. The virus rides into your body on an inhaled water droplet, then tunnels into your cells, replicates, and invades other cells. Your immune system can hunt down and kill the organisms, but it takes a week or more. The spreading virus can overwhelm a person whose immune system does not respond strongly or quickly enough, leading to complications such as pneumonia.

Vaccination provides a training exercise that teaches the immune system how to muster a swift counterattack. Because the virus mutates regularly, a new vaccine must be mixed for each flu season, which in the U.S. usually begins by November and peaks by February. During the previous winter, the World Health Organization recommends three flu strains for the upcoming year's cocktail for the Northern Hemisphere, based on which strains are surging. The 2000–2001 U.S. recipe includes antigens (immunological targets) from A/Panama, A/New Caledonia and B/Yamanashi strains. By February the Food and Drug Administration or the Centers for Disease Control and Prevention provide viral stock to U.S. pharmaceutical companies. The firms inject the stock into fertilized chicken eggs, where it reproduces. They grow each strain separately and then draw off allantoic fluid (egg white) to harvest the virus, purify it, inactivate it, blend the strains with a carrier fluid, and dispense that into vials. Production is largely finished by August, and shipment to health organizations is completed by October.

In 2000 Aventis-Pasteur, Wyeth-Ayerst and Medeva produced 75 million doses, requiring millions of eggs. They sold doses in bulk at $2 to $3 apiece. But delivery was very late—only 70 percent complete by the end of November. Low production yield of the A/Panama strain caused the delay (neither the FDA nor the manufacturers would elaborate). The situation forced health care providers to give immunization priority to the elderly, children, and individuals with compromised immune systems. The National Institute of Allergy and Infectious Diseases has since awarded grants to Aventis, Aviron and Novavax to investigate ways to speed the annual cycle by using DNA production, nasal-mist vaccines and non-egg-grown vaccines, respectively.

Flu vaccine is only 70 to 90 percent effective in healthy adults and somewhat less so in others. Whereas the flu vaccine protects for only 12 months, other vaccines (polio, measles) last many years. For some diseases (tetanus, hepatitis B), a booster shot provides additional exposure to lengthen immunity. *—Mark Fischetti*

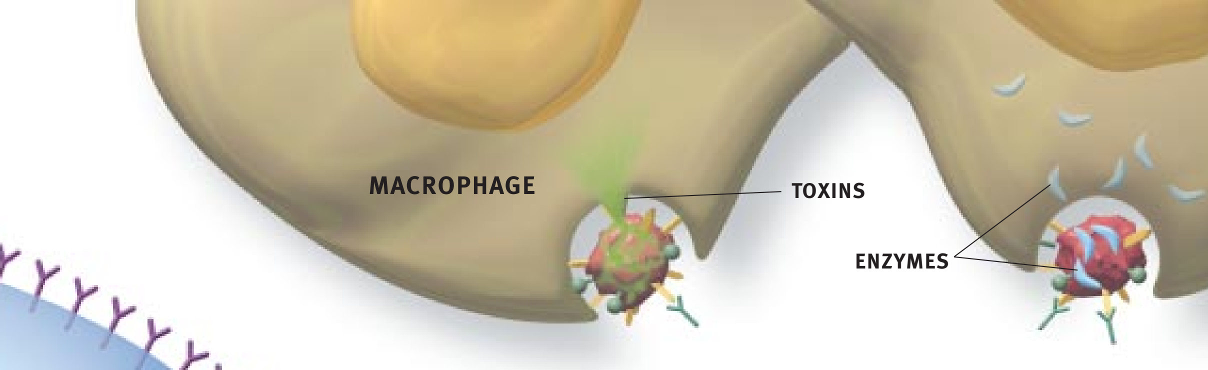




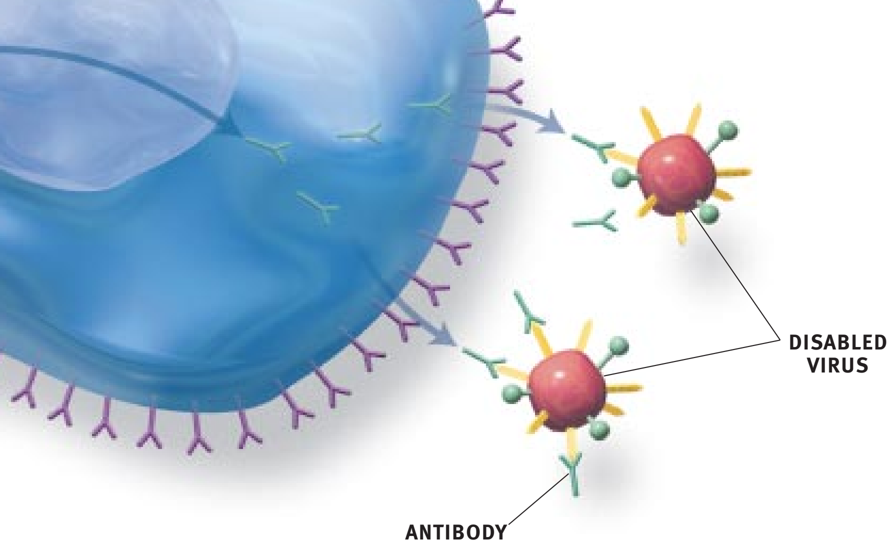
1 INACTIVATED FORM of the flu virus is injected into the body. Macrophages and dendritic cells, the immune system's front-line guards, carry the intruders' antigens to the lymph nodes.

2 IN THE LYMPH NODES, B lymphocyte cells bind to antigens on the virus and determine the invaders' characteristics. Meanwhile helper T lymphocyte cells bind to flu antigens displayed on macrophages. They send cytokines (chemical signals) that help the B cells instruct their nucleus to turn on genes that produce antibodies unique to the virus. The B cells release the antibodies, which attach to a virus's antigens to block it from infecting cells and replicating.

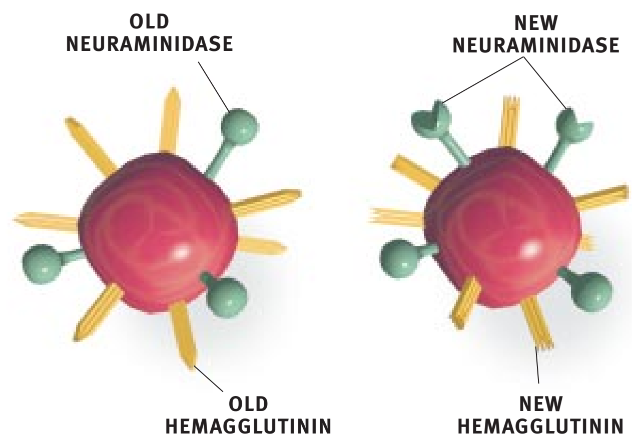
ILLUSTRATIONS BY BRYAN CHRISTIE



3 MACROPHAGE locates an antibody-coated virus and engulfs it. It then produces toxins to kill the invader or digests it with enzymes. Because the injected flu virus was disabled, it poses no health threat, but its presence is enough to cause the immune system to figure out how to generate the right antibodies. If you later inhale a live virus, the remaining antibodies will limit invaders, and the trained B and T cells work together to crank out more antibodies.



FLU VIRUS presents an annual threat because it regularly alters its antigens, called hemagglutinin (HA) and neuraminidase (NA). Therefore, B and T cells trained by the previous vaccine cannot bind to it well. In most years, gradual mutations in the virus's RNA create modest changes in HA or NA, a process called antigenic "drift." Immunologists can tweak vaccines to respond. Occasionally, however, different viruses swap genetic material. This antigenic "shift" dramatically alters the RNA, creating a new flu subtype with radically different HA or NA that requires an entirely new vaccine formula.



DID YOU KNOW ...

- **KILLER GENES:** The flu accounts for 110,000 hospitalizations and 70 million work-loss days annually in the U.S. Three times in the past century sudden, extreme changes in the virus's genes have caught health officials by surprise, causing global pandemics: The 1918 "Spanish flu" killed 500,000 in the U.S. and 20 million worldwide. The 1957 "Asian flu" killed 70,000 Americans, and the 1968 "Hong Kong flu" killed 34,000. Hong Kong strains, still not fully understood, have since resulted in more than 400,000 U.S. deaths, 90 percent in people age 65 and older.

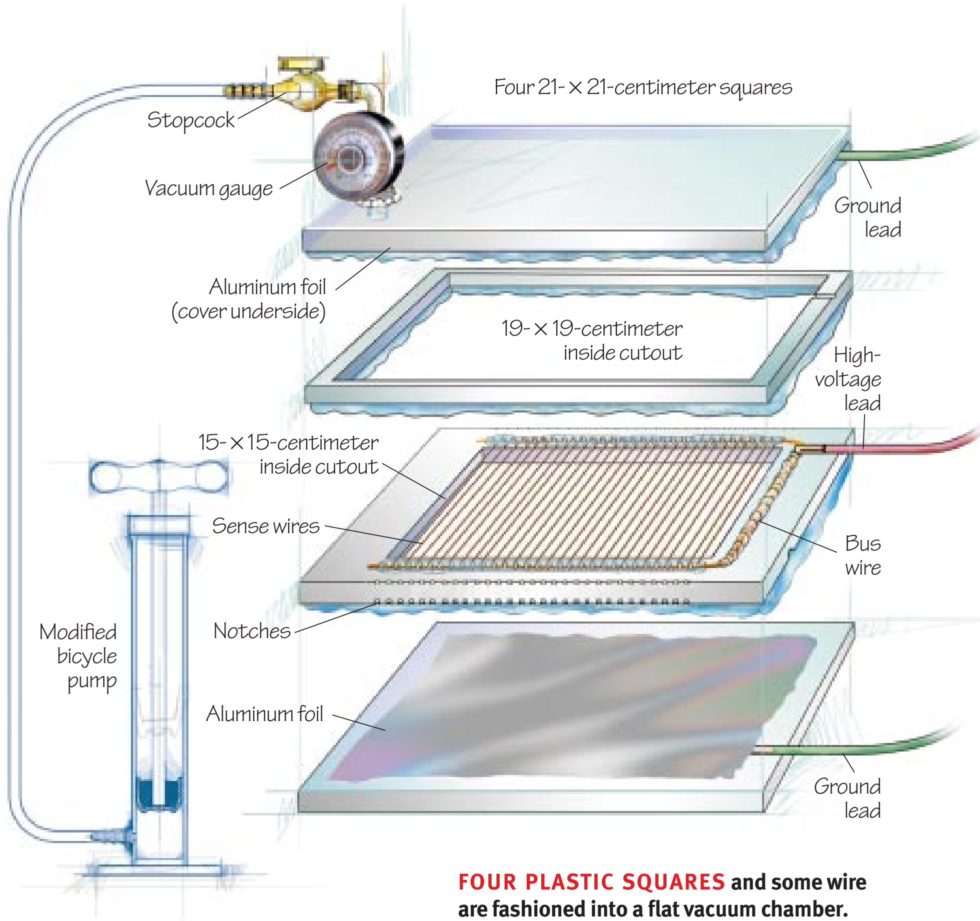
- **COMEBACK:** Supposedly eradicated childhood diseases are returning. More parents, overly concerned that vaccines are dangerous, are refusing to have their children immunized. Others have no insurance or fail to take advantage of government aid. Federal studies indicate that fewer than 50 percent of U.S. children now receive the complete regimen of recommended vaccines. In 1998 there were more than 7,400 U.S. cases of whooping cough, 660 cases of mumps, and 460 cases of measles.

- **ALTERNATIVES:** Biotech firm Aviron in Mountain View, Calif., has finished phase III trials of an attenuated live-virus vaccine, a mist squirted into the nose that would be easier to administer than shots. Aviron has asked the FDA for commercial approval in time for the 2001–2002 flu season. Also, in late November the FDA reported that during an outbreak, daily doses of Tamiflu, a Hoffman–La Roche pill prescribed to lessen flu symptoms in adults, could prevent an individual from getting the illness almost as well as a vaccine. Still, the FDA says that a vaccine is the best overall prevention.

COSMIC_RAYS

Counting Particles from Space

Shawn Carlson explains how to build a cosmic-ray telescope



FOUR PLASTIC SQUARES and some wire are fashioned into a flat vacuum chamber.

Energetic protons from deep space continuously bombard our planet and strike atoms in the upper layers of the atmosphere. These collisions produce tiny nuclear explosions, which in turn give rise to every species in the particle zoo—protons, neutrons, electrons, muons, lambda particles, you name it. Of these, only the muons have enough penetrating power to reach the ground unscathed. Still, the flux of these subatomic particles, known as cosmic rays, is surprisingly high: about 200 rain down on each square meter of Earth every second.

With the instrument described here, ambitious amateurs can monitor the intensity of cosmic rays throughout the day, chart their distribution in the sky and learn something about their energies. The detector consists of two large, flat Geiger counters linked together with a simple electronic circuit. Here's how they work. A set of fine wires carries about 1,000 volts or so. This potential creates an enormous electric field (more than one million volts per meter) near each wire. When a cosmic ray enters this space, it strips some of the atoms in the surrounding gas of a few electrons, which then move toward the nearest positively charged wire. On the way, these electrons gain enough energy from the huge electric field to knock more electrons from other gas molecules. These charges also accelerate and collide to release still more electrons, and so forth.

Within just millionths of a second, the few electrons originally liberated by the passage of the cosmic ray trigger an electric avalanche, causing more than a billion negative charges to cascade down onto the wire. This current flows into a capacitor (C1 on the diagram on page 87), which in turn generates a voltage pulse that feeds into the counting circuitry.

Most Geiger counters are filled with a noble gas, usually helium or argon. Both can be found at welder-supply shops. Helium, so useful for filling balloons, can also be obtained cheaply at any party-goods dealer. Ordinary air also works, albeit at a higher operating voltage.

No matter what gas you're using, you must reduce the pressure in the chamber to about seven centimeters of mercury—about 10 percent of atmospheric pressure. The March 1960 and October 1996 installments of this column describe homemade vacuum pumps that should serve nicely. But you can also reduce the chamber pressure with a bicycle pump if you modify it appropriately (consult the Web site of the Society for Amateur Scientists for details).

Begin construction by cutting four pieces (as shown above) from a rigid sheet of plastic that is 3/8 of an inch, or about one centimeter, thick. Using the edge of a small file, carve a series of small notches spaced precisely half a centimeter apart on opposite sides of the piece indicated in the diagram. Next, arrange a length of hefty "bus wire" (solid copper wire without insulation) as shown. Secure it with tape at the corners and apply tiny dollops of five-minute epoxy between the notches. Also add a liberal amount of epoxy to the wire along the side you've not filed, making sure to leave at least one centimeter around the perimeter untouched to accommodate the piece that fits above.

Use the notches to position the "sense wire," bare copper wire that is only 10 thousandths of an inch (about 250 microns) thick. Wrap this fine wire around the square plastic frame, using a steady hand to maintain tension, and hold the ends in place temporarily with duct tape.

Now you must delicately solder the sense wire to the bus wire everywhere they touch. Use a hot soldering iron and plenty of flux. Then attach the sense wire to the frame with a liberal coating of slow-setting (24-hour) epoxy. Once it sets, carefully snip the excess wire just where it

DANIELS & DANIELS

emerges from the epoxy to yield a single plane of 29 sense wires. Solder a high-voltage lead to the bus wire.

Use epoxy to attach aluminum foil to the top and bottom plastic squares, as shown in the diagram, and install the stopcock and low-pressure gauge to the top piece. Solder ground wires to the aluminum foil. Carve three narrow channels in the middle plastic pieces for the high-voltage and ground wires. (A Dremel tool will work well.)

Now you're ready to assemble the chamber. The unit has to be airtight, so run continuous beads of silicone aquarium cement where the layers join and put a heavy weight on top or clamp things while the adhesive sets. Make sure also that the channels that hold the lead wires are especially well sealed.

The speaker provides an audible output, but for more exacting work I count my cosmic rays on a digital pedometer that I bought for $15 from Radio Shack (catalogue no. 63-618). When this gadget is jostled enough to swing its tiny magnetic pendulum to one side, the magnet pulls together two fine metal strips inside an encapsulated switch, completing a circuit. By bypassing the switch with one of your own, the pedometer can be made to count almost anything that is not producing events too often. The limit seems to be about five times a second, which is just fine for counting cosmic rays.

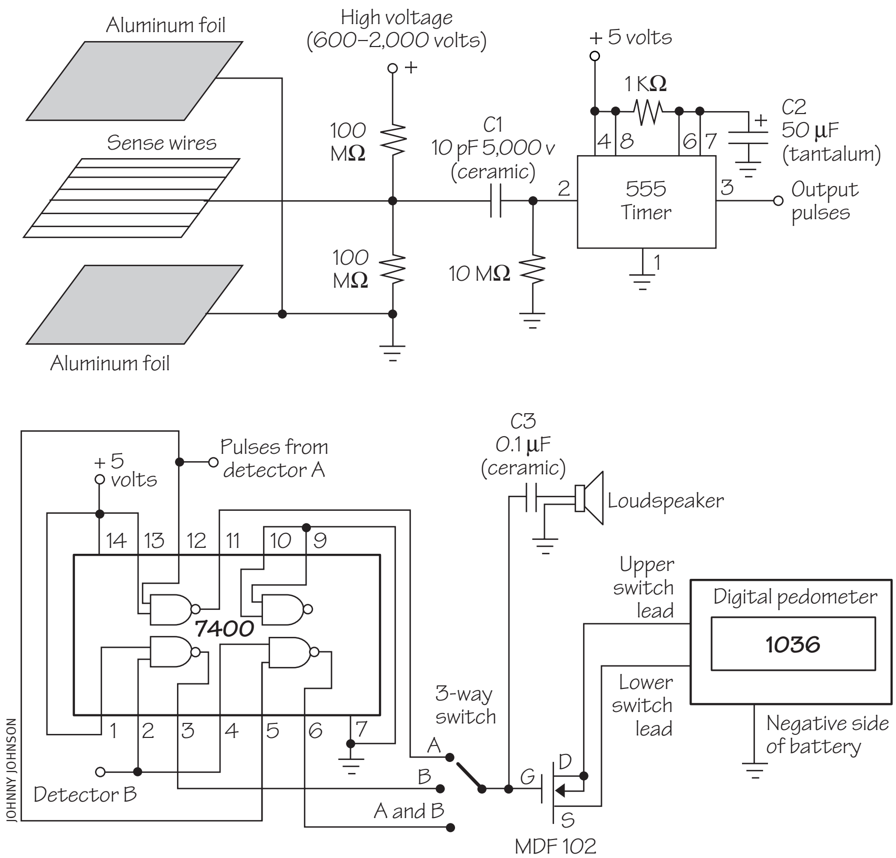
To convert the pedometer, remove the battery cover and nip away at the extreme right side of its plastic case using a small pair of pliers. Then cut off the exposed switch and solder on leads from the detector circuit.

You'll need a variable high-voltage supply to operate the apparatus. Before you power things up for the first time, be absolutely certain that no high-voltage wires are exposed and be extremely careful to avoid any possibility of a dangerous shock. When you're sure that everything is safe, apply 600 volts to start and slowly raise the potential until you just begin to register counts. This setting is your chamber's threshold voltage. The count rate will rise with the applied potential until essentially all the ionizing particles that enter the chamber are detected. At that point (about 1,200 volts for my detector), the count rate levels off. This "plateau" should extend for several hundred volts. As you raise the voltage even higher, secondary effects generate spurious counts, and so the rate rises again. Set your operating voltage at the center of the plateau.

Once you've built and tested two identical chambers, it's easy to construct a cosmic-ray telescope. Just align the two chambers and flip the switch to the A-and-B position, which counts just the events that trigger both detectors. Because particles produced by radioactive decay don't have enough energy to pass through both plastic boxes, your telescope will now show only cosmic rays.

This equipment affords many opportunities for research. Position the chambers close together to detect daily and seasonal variations in the flux of cosmic rays. Or place the detectors farther apart to restrict the angular acceptance of the telescope. This maneuver allows you to measure the flux coming from a given direction and to observe how the rate depends on elevation angle and azimuth.

By placing material between the two chambers, you can screen out low-energy cosmic rays. Muons lose about two million electron volts (MeV) of energy for each centimeter of water they pass through. A brick, which is about two times as dense as water, will extract about 4 MeV for each centimeter of thickness. You can use

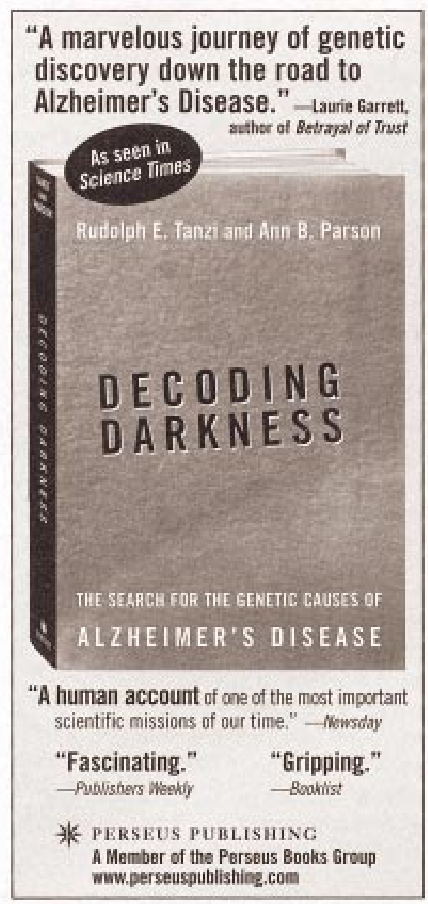


HIGH-VOLTAGE SOURCE and simple circuitry detect the passing cosmic-ray particles.

this effect to investigate the energy spectrum of the more feeble muons impinging on your detector. And you can detect the immense "air showers" that very energetic protons spawn by comparing results from two telescopes situated about 100 yards (or meters) apart. With a little imagination and effort, you will surely make some fascinating discoveries. SA

The Society for Amateur Scientists will offer a kit for this project until January 2002. The package contains only the various electronic components required (apart from the pedometer) and a spool of fine sense wire. The cost is $30. To order, call the society at 401-823-7800. For an ongoing discussion about this project, surf over to www.sas.org and click on the Forum button. You can write the society at 5600 Post Road, #114-341, East Greenwich, RI 02818. To purchase Scientific American*'s CD-ROM containing every article published in this department through the end of 1999 (more than 1,000 projects in all), consult www.tinkersguild.com or dial toll-free: 888-875-4255.*

Erratum: *Mercury's freezing temperature was incorrectly given in the Amateur Scientist for December 2000. The correct value is −38.9 degrees; the corresponding output voltage in the table on page 104 should read −0.365 volt.*

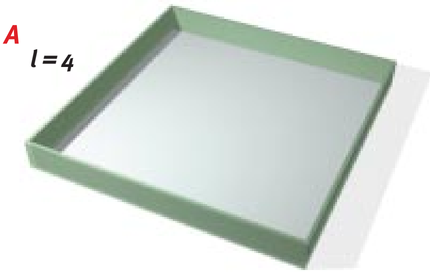


GEOMETRY_OPAQUE FENCES

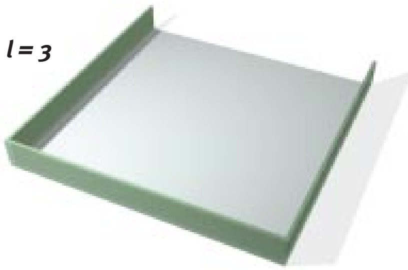
Pursuing Polygonal Privacy

Ian Stewart proves that good fences make good neighbors

A
$l = 4$

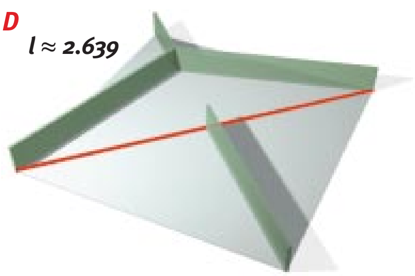


B
$l = 3$

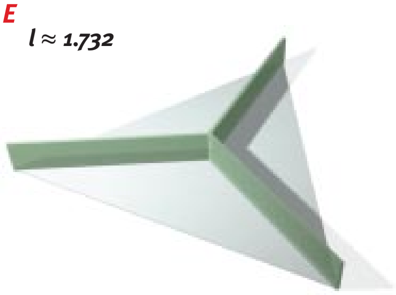


C
$l \approx 2.732$

D
$l \approx 2.639$



E
$l \approx 1.732$



Combinatorial geometry is one of the most appealing areas of mathematics, full of simple problems whose solutions are unknown. The aim of these problems is to find arrangements of lines, curves or other geometric shapes that achieve some objective in the most efficient manner. This month I want to concentrate on a puzzle known as the Opaque Square Problem, along with several fascinating variations. Bernd Kawohl of the University of Cologne in Germany brought the puzzle to my attention, and my discussion is based on an article he sent me.

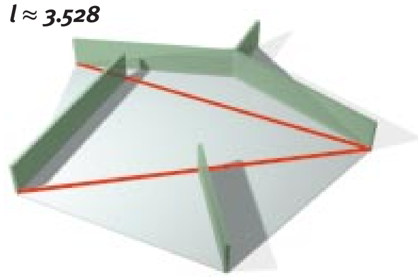
Suppose you own a square plot of land whose sides, for the sake of simplicity, are each one mile long. To ensure your privacy, you want to build an opaque fence—a barrier that will block any straight line of sight passing through the square plot. Moreover, to save money, you want the fence to be as short as possible. How should you build it? The fence can be as complicated as you like, with lots of different pieces that can be curved or straight.

Perhaps the most obvious solution is to build a fence around the perimeter of the square plot, with a total length of four miles [*see illustration A at left*]. A few moments' thought reveals an improvement: leave out one side to create a square-cornered U shape [*see illustration B*]. Now the length reduces to three miles. This is, in fact, the shortest fence possible if we impose the additional condition that the fence must be a single polygonal or curved line. Why? Because every opaque fence must contain all four corners of the square, and the three-sided U is the shortest single curve that contains all the corners.

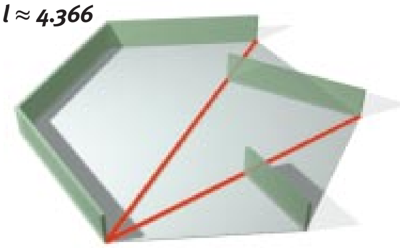
We can build a shorter fence, however, that consists of more than one curve. Illustration C shows a fence with a length of $1 + \sqrt{3}$ (about 2.732) miles. The angles between the lines are all 120 degrees. Arrangements of this kind are called Steiner trees; the 120-degree angles minimize the length of the tree. This is the shortest fence in which the curves are connected. If we allow the fence to have several disconnected pieces, the total length can be reduced to about 2.639 miles [*see illustration D*]. The three lines in the upper half of the diagram also meet at angles of 120 degrees. This last example is widely believed to be the shortest opaque fence for a square plot, but nobody has proved this yet.

Indeed, mathematicians are not sure whether a shortest opaque fence exists. It may be possible to keep shortening the length by making the fence more and more complicated. For any given number of connected components, it has been proved that a shortest opaque fence does exist. What is not known is whether the minimal length keeps shrinking as the number of components increases without limit or whether a fence with an infi-

F
$l \approx 3.528$

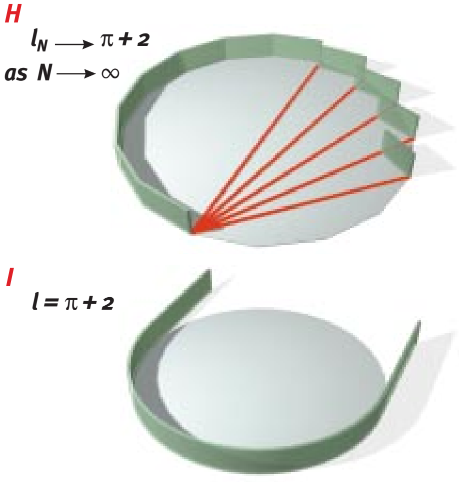


G
$l \approx 4.366$



OPAQUE FENCES are barriers that block any straight line of sight passing through a given figure. For a square, a perimeter fence (*A*) and a three-sided U shape (*B*) are opaque, but a Steiner tree (*C*) and a two-component fence (*D*) are shorter. The shortest opaque fence for an equilateral triangle is also a Steiner tree (*E*). The best-known opaque fences for the regular pentagon (*F*) and hexagon (*G*) each have three components. All fence lengths (*l*) are approximate except those for *A* and *B*.

ALL ILLUSTRATIONS BY BRYAN CHRISTIE



EVEN-SIDED POLYGON with many sides has an opaque fence with many components (*H*). Their combined length approximates the length of the shortest single-curve fence for a circle (*I*).

nite number of components can outperform all fences with a finite number of components. These possibilities seem unlikely, but neither has been ruled out.

Kawohl has provided a lovely proof that illustration D on the opposite page is the shortest fence having exactly two components. He shows that one component must contain three corners of the square and that the other must contain the remaining corner. The first component must therefore be the shortest Steiner tree linking three corners, which is the shape shown in the upper part of the figure. The convex hull of this shape—the smallest convex region that contains it—is the triangle formed by cutting the square in two along a diagonal. The second component must be the shortest curve that joins the fourth corner to this triangle: the diagonal line from that corner to the center of the square.

What about shapes other than the square? If the plot of land is an equilateral triangle, the shortest opaque fence is a Steiner tree formed by joining each corner to the center along a straight line [*see illustration E*]. If the plot is a regular pentagon, the best-known opaque fence comes in three pieces [*see illustration F*]. One piece of the fence is a Steiner tree linking three adjacent corners of the pentagon. The second piece is a straight line joining the fourth corner to the convex hull of the Steiner tree. The third piece is a straight line joining the fifth corner to the convex hull of the four other corners. Nobody has proved that this fence has a minimal length, but no shorter opaque fence has been found.

The best-known fence for the regular hexagon is similar [*see illustration G*]. Because the corner angles of the hexagon are 120 degrees, the Steiner tree consists of three consecutive sides of the figure itself, linking four adjacent corners. The second component of the fence is the shortest line joining a fifth corner to the convex hull of the Steiner tree, and the third component is the shortest line joining the sixth corner to the convex hull of the five other corners. Again, no one has proved that this fence has a minimal length.

You can use the same type of construction to draw a conjectured minimal fence for any regular polygon with an even number of sides [*see illustration H at left*]. Simply divide the polygon in two by a diameter joining two opposite corners. The first component of the fence is formed from all the edges that lie in that half, forming the polygonal analogue of a semicircle. The second component is the shortest line linking the next corner to the convex hull of the first component. The third component is the shortest line linking the next corner to the convex hull of the first two components, and so on.

A regular polygon with a large number of sides is very close to a circle. What is the shortest fence that makes a circle opaque? For simplicity, suppose that the circle has a radius of one mile. The simplest fence that comes to mind is the circumference of the circle, which has a length of 2π (about 6.283) miles. We can do better, however, if the fence is permitted to lie outside the circular plot. Run the fence along half the circumference, creating a semicircle, and extend it by adding two one-mile lines that are tangent to the circle at the ends of the semicircle [*see illustration I*]. The resulting U shape is an opaque fence for the circle, with a length of $\pi + 2$ (about 5.142) miles.

It can be proved that this figure is the shortest opaque fence if we insist that it be a single curve—all in one piece and with no branching points. Another way to describe the problem is to think of trenches instead of fences. Imagine that a straight underground pipe is known to pass within a mile of some specific point. What is the shortest trench we can dig that is guaranteed to find the pipe? We know that the pipe must cross a circle with a one-mile radius centered at that point and must therefore hit any opaque fence for that circle. So we should dig a trench in the form of an opaque fence.

In this version of the puzzle, it is natural to allow the trench to go outside the circle, but fences are typically built on the owner's land rather than on the neighbors'. Kawohl shows that the shortest opaque fence lying entirely inside the circle also cannot be longer than $\pi + 2$ miles. He does this by considering the conjectured fence for an even-sided polygon with a very large number of sides, thus closely approximating the circle. A trigonometric calculation proves that the length of the fence shown in illustration H approaches $\pi + 2$ as the number of sides increases without limit.

But are the conjectured fences truly the shortest, or is there a way to shorten them further? What about other shapes, such as irregular polygons (convex or not), ellipses and semicircles? And what about the same problem in three dimensions (the opaque cube and sphere)? Recreational mathematicians have much to investigate. SA

READER_FEEDBACK

Several readers objected to a calculation I did in the column on logical fractals ["A Fractal Guide to Tic-Tac-Toe," August 2000]. I stated that the number of possible games of tic-tac-toe is 362,880. I should have made it clear that this number is correct only under the assumption that the game continues until all the squares in the grid are filled, rather than stopping when someone wins. The total number of sequences leading to a completed grid is $9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ (denoted as 9!), which equals 362,880.

But what is the number of actual games? John Stewart of Rockledge, Fla., pointed out that the number can be expressed as:

$$9! - 24M - 6N - 2P - Q + (M + N + P + Q)$$

where M, N, P and Q are the number of games completed after the fifth, sixth, seventh and eighth moves, respectively. The precise values of M, N, P and Q remain to be calculated. Any takers? John Stewart (no relation to myself, by the way) suggests that M might be 1,440. *—I.S.*

Books

REVIEWER_NANCY PADIAN

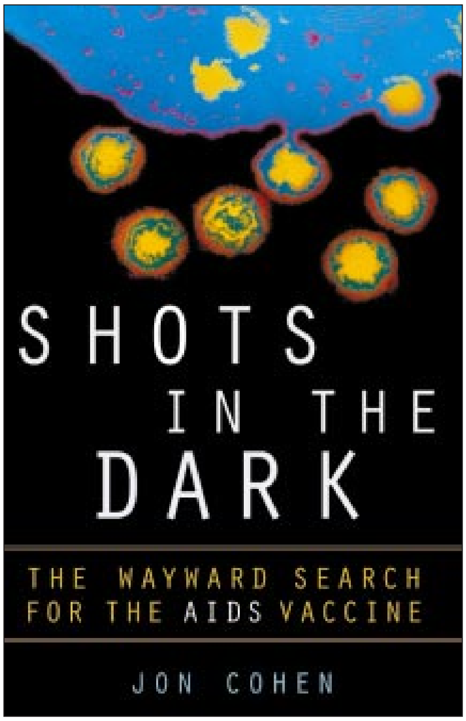
Why Haven't We Found an AIDS Vaccine?

Jon Cohen argues that the obstacles may be more human than viral

Vaccines are among the greatest public health achievements of the past century. Since 1900, vaccines have controlled smallpox, poliomyelitis, measles, rubella, tetanus, diphtheria and many other infectious diseases. In contrast, for the past 20 years we have been faced with HIV/AIDS, a deadly new infectious disease that continues to elude effective vaccines. HIV is now the primary cause of death in Africa and the fourth worldwide. More than 15,000 new HIV infections occur every day, most in developing countries, and over 34 million people now live with HIV or AIDS—including over 13 million children orphaned by fatally infected parents.

One might think that more than a decade's search for an AIDS vaccine would end with success. In *Shots in the Dark,* however, journalist Jon Cohen brilliantly describes the inextricable weave of science, politics, legalities, ethics and business that, like a dysfunctional family, seems to have repelled the very cooperation that a successful vaccine effort needs most. The biology of the HIV virus—numerous strains, rapid rates of mutation and replication, and its habit of attacking and exploiting the very cells that are designed to fend off infection—hinders the development of an effective vaccine. As the story unwinds, though, Cohen makes it clear that science presents fewer obstacles than other forces.

Rugged scientific individualism has been one impediment. The culture of government-funded biomedical science favors investigator-initiated, basic laboratory research over applied research, such as clinical trials. Cohen casts this as the struggle between reductionism and empiricism. Biotech and pharmaceutical industries, in contrast, tend to prize applied research, but their measure of success is as much financial as scientific. For them,



Shots in the Dark: The Wayward Search for an AIDS Vaccine
by Jon Cohen
W. W. Norton, New York, 2001 ($27.95)

fiscal and legal risks present other impediments. The market for vaccines is mainly in the developing world. In the industrial world, individuals and governments can afford treatment because their resources are greater and fewer individuals are infected, so the need for vaccines is limited. Potential liability resulting from "breakthrough" infections (no vaccine is 100 percent effective) also discourages private industry.

Cohen provides a cogent example of how these forces play out. Limited data on a recent AIDS trial vaccine showed that the vaccine protected chimps and so seemed safe for humans. Yet endless debate ensued as to whether to move it into large-scale efficacy trials in humans. Much of the debate centered on basic research: how it might work and its likelihood for success. Milestones by which to gauge success had not been established; thus, different advisory committees gave different advice. Empiricists argued that because no in vitro or animal testing could ever ultimately determine potential, trials should move forward. Antagonists (reductionists?) argued that existing data were not sufficient or compelling enough to merit efficacy trials. Pharmaceutical companies moved in and out of the debate depending on their read of the tea leaves. Who, Cohen asks, is in charge?

Ethical issues further impede the ability to test vaccines. Developing countries have higher infection rates, so the effectiveness of vaccines can be tested more quickly. Could black Africa become an experimental testing ground? Is it ethical to test a vaccine in one country when it was developed in another? If the control (placebo) group receives intensive counseling about behavior change or other preventive strategies, is it possible to detect an effect of the vaccine? Should individuals who are infected during the course of a trial be offered standard care in that country or the best available treatment—which might not be sustainable once the trial is over?

Is There a Solution?

Cohen's description of events is gripping, even when he lays out the intricacies of molecular genetics, but his most valuable contribution is his prescription for advancing the effort to develop a vaccine. He proposes an AIDS March of Dollars, along the lines of the March of Dimes for a polio vaccine. Funding, he suggests, should come mainly from philanthropists, as an adjunct to the National Institutes of Health. The program would provide a central authority to create a targeted, strategic plan that could define gaps in knowledge, assess the relative merits of numerous candidate vaccines, and foster coordination. He suggests possible leaders, an ethical review board and scientific advisory boards. Funds would be allocated to those willing to participate and share data. Government would facilitate the effort and perhaps provide legal protection for inventors and researchers, who would receive limited royalties.

Cohen clearly has a bead on the scene in the U.S., and so his focus is mainly on coordinating the vaccine effort here. The global epidemiology of HIV/AIDS, however, requires a global response. In February 2000 a joint WHO-UNAIDS HIV Vaccine Advisory Committee was created to deal more visibly with the coordination of vaccine development. Last June, 40 leading African scientists pledged to use their "personal and collective commitment and expertise in the development and implementation of an HIV vaccine strategy specific to Africa." To be successful, the March of Dollars will most likely have to be an international effort.

In principle, Cohen's prescription for the AIDS vaccine search could be applied on a global level. Promising signs indicate that such an effort may now be possible. These include the pledge by leading African scientists and the creation of organizations such as the International AIDS Vaccine Initiative, whose mission is to test a variety of candidate vaccines as rapidly as possible, including plans for distribution. Even so, could this model be successful? Cohen himself says that such an effort "may well not lead to the day where ... the world has realized its hopes and found an effective vaccine against HIV. But the world at least could declare that it did everything in its collective power to develop an AIDS vaccine as quickly as possible, which is not something it can now say."

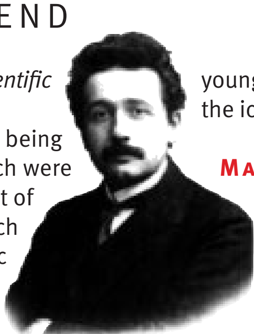
> ***Cohen proposes an AIDS March of Dollars,*** *along the lines of the March of Dimes for a polio vaccine.*

Cohen's proposal, an elaboration of the model of the March of Dimes, could clearly benefit the search for a vaccine. But perhaps more important, it is also the kind of coordination necessary for other types of public health prevention efforts besides vaccines. For example, the search for a microbicide that can protect against HIV is a research movement that suffers from a lack of leadership and organization. There are still no clear criteria for selecting products to move from animal and in vitro studies to large-scale efficacy trials in humans. Because fiscal returns are uncertain, pharmaceutical companies have yet to mount large-scale attempts to develop products. The same ethical issues described above hold here as well. The reader of *Shots in the Dark* needs to consider its subtitle, "The Wayward Search for an AIDS Vaccine," as emblematic of the devastating global consequences of not forging stronger cooperation among governments, affected communities, industry and scientists in all HIV/AIDS prevention efforts. SA

NANCY PADIAN is a professor of obstetrics and gynecology at the University of California, San Francisco, with a joint appointment at the U.C. Berkeley School of Public Health. She is director of international research at the U.C.S.F.-U.C.B. AIDS Research Institute. For the past 15 years she has studied prevention of HIV and other STDs and currently has a large research program in Zimbabwe focused on HIV prevention in women.

THE EDITORS_RECOMMEND

DENNIS OVERBYE'S *Einstein in Love: A Scientific Romance*. Viking, New York, 2000 ($27.95).



Albert Einstein

One might be misled by the title. Far from being only a chronicle of Einstein's romances, which were in fact more numerous than one might expect of such a cerebral man, Overbye's book is a rich and absorbing account of Einstein's scientific work in the first two decades of the 20th century and his family life from childhood onward. Overbye, deputy science editor of the *New York Times,* is excellent at describing the great man's work—on, among other things, electrodynamics, thermodynamics, statistical mechanics and relativity. "Few so-called revolutions in science are truly revolutions," Overbye says, "but relativity was one." And having visited many of the places that were important in Einstein's life and read "hundreds upon hundreds of published and unpublished letters" in "Einstein's cramped handwriting," he paints a vivid portrait of the man. Overbye says that his "goal has been to bring the youthful Einstein to life, to illuminate the young man who performed the deeds for which the old man, the icon, is revered." He has done that admirably.

Mileva Maric, his first wife

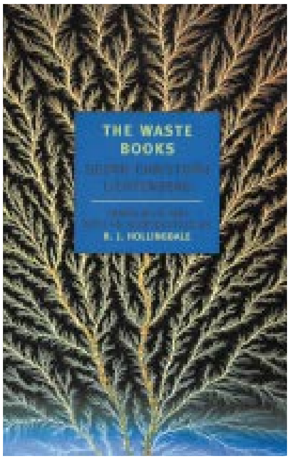
FROM *EINSTEIN IN LOVE*

MARTIN GARDNER'S *From the Wandering Jew to William F. Buckley, Jr.: On Science, Literature, and Religion*. Prometheus Books, Amherst, N.Y., 2000 ($27).

Once again the reader gets to see what a broad range of things Gardner thinks about and how crisply he writes. Already renowned as a mathematical gamesman and a steely critic of pseudoscience, Gardner extends his reach in this collection of nine essays and 20 book reviews. In "The Wandering Jew and the Second Coming," he touches on a biblical message that is, he says, "for Bible fundamentalists one of the most troublesome of all New Testament passages." Reviewing *Demon-Haunted World,* in which astronomer Carl Sagan attacked the "dumbing down" of science, Gardner calls the book "a powerful indictment of today's miserable science teaching, the upsurge of Protestant fundamentalism and the roles of greedy book publishers, abetted by the print and electronic media, in accelerating America's dumbing down." He also considers a "question that troubles all the parents of chess prodigies," namely, what direction the prodigy will take. "Will he become an honored grandmaster, happy and well adjusted as the Russian Boris Spassky, or will the game turn him into a miserable misfit like Bobby Fischer?"

Georg Christoph Lichtenberg's *The Waste Books.* Translated by R. J. Hollingdale. New York Review Books, New York, 2000 ($12.95).

Odd title, unusual book. Lichtenberg (1742–1799) was a German polymath: astronomer, experimental physicist, mathematician and critic of art and literature. In his student days he began the lifelong practice of recording his thoughts, observations and reminders in notebooks that he called *Sudelbücher* after the "waste books" in which English business houses of the time entered transactions temporarily until they could be recorded in formal account books. By the end of his life he had accumulated 11 *Sudelbücher,* which he labeled as volumes A through L (skipping I). Hollingdale, a translator of Nietzsche, Goethe and Schopenhauer, has translated the notebooks. Here he presents excerpts, focusing on what he says are best called aphorisms. Lichtenberg turns out to be quite an aphorist, repeatedly surprising and entertaining the modern reader. Examples: "Whenever he was required to use his reason he felt like someone who had always used his right hand but was now required to do something with his left." "You can make a good living from soothsaying but not from truthsaying." "The book which most deserved to be banned would be a catalog of banned books." "Astronomy is perhaps the science whose discoveries owe least to chance, in which human understanding appears in its whole magnitude, and through which man can best learn how small he is."



Paul R. Ehrlich's *Human Natures: Genes, Cultures, and the Human Prospect.* Island Press, Washington, D.C., 2000 ($29.95).

The idea that human nature is a unitary, unchanging thing, Ehrlich says, "has become a major roadblock to understanding ourselves." And so he argues for the concept of human natures, plural. "The universals that bind people together at any point in our evolution are covered in the word *human*. The word *natures* emphasizes the differences that give us our individuality, our cultural variety, and our potential for future genetic and—especially—cultural evolution." To understand the concept, Ehrlich writes, one must trace the course of human evolution. And that is what he does, emphasizing human cultural evolution, "the super-rapid kind of evolution in which our species excels." With the result that the nature of a great musician is not identical with that of a fine soccer player and the nature of an inner-city gang member differs from that of a child raised in an affluent suburb. "We need to learn how to direct that cultural process in ways more beneficial for the human future," he says. Ehrlich, professor of population studies and of biological sciences at Stanford University, has an extraordinary range of interests and mines a rich lode of knowledge in laying out his argument.

FROM *HUMAN NATURES*

Marjorie Shostak's *Return to Nisa.* Harvard University Press, Cambridge, Mass., 2000 ($24.95).

Twenty years ago Marjorie Shostak published the story of her relationship with Nisa, a rural tribeswoman in the Kalahari Desert of Botswana. The book, *Nisa: The Life and Words of a !Kung Woman*, became one of the classics of anthropological literature. (Harvard University Press has simultaneously brought out a new paperback edition of this book.)

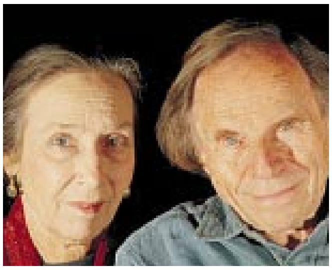
In *Return to Nisa,* Shostak tells of her travels back to Botswana to see what has become of Nisa and of the !Kung as they have moved from hunting and gathering toward a more sedentary way of life. The book would be poignant for this tale alone—the changes in the life of the tribe and its political travails, Shostak's surprise and hurt at her ambivalent reception, especially the flood of queries of "What have you brought me?" But it is made even more so because one reason Shostak returned was that she had breast cancer. She died before she finished the book (which her husband and two friends completed from manuscript drafts). It is, understandably, a much more personal story than *Nisa,* a search for healing and for a less complicated past and the record of a friendship that surmounted time, distance, and cultural boundaries.

FROM *RETURN TO NISA*

Steven Levy's *Crypto: When the Code Rebels Beat the Government—Saving Privacy in the Digital Age.* Viking, New York, 2000 ($25.95).

The government's argument, doggedly pressed mainly by such security-obsessed arms as the Department of Defense, the Justice Department and the National Security Agency: cryptography should be under firm government control, with strong codes to protect national security and weak ones for the public so that the government can break them to catch criminals and terrorists. The counterargument, pressed with equal determination by a mixed group that Levy rather unflatteringly calls the Cypherpunks: secure codes are vital to business transactions in the digital age and to people wanting privacy in electronic communications.

Levy, chief technology writer for *Newsweek,* goes deeply into the 30-year battle over which side would prevail. He tours the landmarks of the battlefield, among them the government's Data Encryption Standard, public-key cryptography, the key escrow plan and the Clipper Chip. And he vividly portrays the leading actors on both sides. In the end, it was the burgeoning of the Internet and the necessities of e-commerce that won the day for the Cypherpunks. As Judge Betty B. Fletcher of the U.S. Court of Appeals for the Ninth Circuit put it in a decision handed down last year: "Government attempts to control encryption ... may well implicate not only First Amendment rights of cryptographers but also the constitutional rights of each of us as potential recipients of encryption's bounty."



The Big Bang: Wit or Wisdom?

Philip and Phylis Morrison review the way the universe came into existence—and how it continues to evolve

No cosmological concept is as widely known as the big bang: from a state without physical order, lacking even space and time, matter appeared. How could so flippant a term denote so profound an idea?

A friend of ours at M.I.T., a skeptical experimenter, often finds himself at work amid that dreamier and indulgent society on our Pacific Coast. Last summer he was in a Caltech audience, his peers in celebratory mode. The stage was held by a performance artist who entertained with original songs. She describes herself as “Bette Midler meets Carl Sagan, with a touch of Tom Lehrer and Mae West.” A spotlight lit the tall performer, her gleaming gown ornamented with patches that, though colorless, dispersed the white beam into rich spectral hues. Another performer might regard such visual effects as arcane stagecraft, but not this artist, whose day job is based on years of graduate studies in physics. For Lynda Williams, instructor in physics and astronomy at San Francisco State University, “physics is such a lyrical subject.”

Professional dancer and singer, this physics chanteuse entertains fellow scientists with her “Cosmic Cabaret,” her apt talents certified by our M.I.T. eyewitness and publicly praised in the *New York Times* by an interviewer of repute. We expand here on one song in her repertoire. “In the beginning, there was nothing,” she offers gently, and then, “BIG BANG!” She has captured the essence of the widespread belief and displayed its wide acceptance.

The term arose in the early months of 1950. A young cosmologist, original, articulate—now Sir Fred Hoyle—completed his hastily organized series of Saturday evening radio broadcasts over the BBC Third Program. These exciting talks had “hit the top of the annual national ratings,” he recalls. “In the press of that last lecture I coined the term ‘big bang.’ ” His intent was in no way to accept the concept but rather to bury it under an ironic name. Instead his put-down has lived on; its pith and drama have become the public description of a unique moment. Even the experts were drawn in.

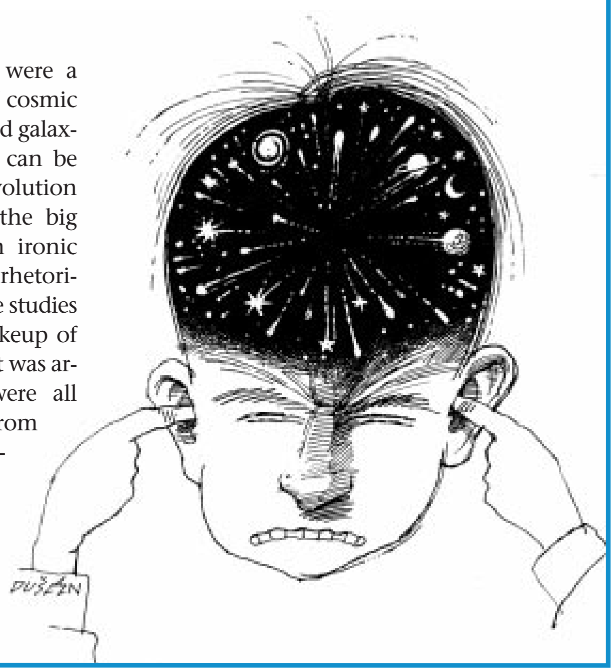
A couple of years earlier Hoyle and two like-minded colleagues had proposed a startlingly new alternative to the equations in early 1917 of Albert Einstein’s enduring cosmological theory. Their steady state view required a minute leakage of new matter, particle by particle, into empty space. Hoyle contended that the classical Einsteinian view taken literally implied an even more striking postulate of origin: all matter-energy created at one moment. Of course, no big bang had ever been seen; it followed only by smooth extrapolation of the equations backward into a visibly ever hotter and ever denser past. At the peak of the cosmic inferno, there was an abrupt failure of physical theory.

At the peak of the cosmic inferno, there was an abrupt failure of physical theory.

By the mid-1960s there were a number of clear signs of cosmic evolution—real changes in old galaxies of some properties that can be measured from far away. Evolution became a better bet, and the big bang seemed no longer an ironic put-down but a memorable rhetorical rubric. The richest of these studies was about the chemical makeup of the stars. After World War II it was argued that the elements were all made by nuclear reactions from the gas of protons and electrons, raw material for the rest. In open space, as hot then as the beam of a cyclotron, the elements were to be built. But the allowed reactions failed to build up elements heavier than beryllium. By the late 1950s the site of element “cooking” was firmly set within the dense and enduring stars, where most starlight is fueled. That proved wrong, too; the stars do make all the heavies but cannot make enough elements below beryllium. It takes both processes: in hot space, only protons and a few light products are made; all the rest comes from the stars.

By the 1970s we knew the density and temperature in space, minutes after the start of expansion, that would in fact generate the observed mix of the lightest, simplest atoms, the bulk of all atomic matter. Farther out in space, back in time, we expect a much hastier and hotter domain, one that makes protons and electrons, the raw material for what was to come. We see a background “glow” of microwaves filling the dark sky between the galaxies. This is a glimpse of early radiation, much cooled now by expansion, an aftermath of prior hotness. The unbroken process invited

Continued on page 95



DUSAN PETRICIC

Home from Home

James Burke employs body snatching, mastodons, Great White Explorers, journalism, war, raincoats and malaria to literary ends

DAVE PAGE

Some time ago I wrote that I'd revisit the case of Anne Home, who married John Hunter, the 18th-century carpenter turned (what else?) surgeon-pathologist and patron of body snatchers. Anne was a minor poet who wrote stuff Haydn set to music (one notable piece: "My Mother Bids Me Bind My Hair") and was known for her talkative drawing-room lit-crit parties (on one occasion an exasperated Hunter chucked them all out).

In 1792 a 17-year-old nice young man, recommended by an old school chum of hers, fetched up at Anne and John's home in London and was given a freebie apprenticeship to sketch Hunter's work (bits of pieces, so to speak) and to look after the growing confusion of anatomically related bric-a-brac Hunter had amassed in his back-room museum. This pile eventually became famous as the Hunterian Collection, and after Hunter's death in 1793 the nice young man, William Clift, looked after it for more than 50 years. Night and day, they said. Clift became the indefatigable research resource on Hunteriana for such luminaries as Cuvier, Lyell, Davy and Banks and a walking encyclopedia on anything anatomical.

Surprisingly, in view of all this, he also had time to marry and have a daughter, Caroline Amelia. When, in turn, another nice young man arrived to become Clift's assistant, in 1835 Caroline married him. The new hubby, anatomy whiz Richard Owen, had already burst upon the scientific scene three years earlier with his boffo "Memoir on the Pearly Nautilus," thus establishing himself in an area of shell study too arcane for me to appreciate. But not others. Soon after taking over all matters Hunterian from Clift, Owen had catapulted to fame and the natural history department of the British Museum. Where his dynamism went over like a lead balloon with the local snoozers. Which is why there isn't one anymore (nat. hist. dept. at B.M.).

Owen got his own place, after designing it himself and making such a fuss that the government handed over some of the profits from the 1851 Crystal Palace Exhibition to fund his purpose-built Natural History Museum of South Kensington, which finally opened its doors to the public in 1881. Multicolored-Victorian-pseudo-Baroque-Rhineland-Romanesque might hardly be what you would expect "purpose-built" to look like, but the place is fine if you're into architectural mishmash. As you can tell, Owen ended up a real mover and shaker. He knew everybody, dissected everything (from kinkajou to wombat), anonymously attacked Darwin's theory, and became so harrumph the queen gave him a noble title and a place to live.

Just the guy you consulted when you wanted to write a best-seller on "Missionary Travels and Researches in South Africa." Which in 1857 was the been-there-done-that effort of a rags-to-riches missionary pal of Owen's who had spent 15 years getting mauled by lions, gutted by fevers, and going where they had never seen a white man before to take the Gospel message to darkest (that is to say, non–Church of England) Africa. He returned, the nation's hero, in 1856.

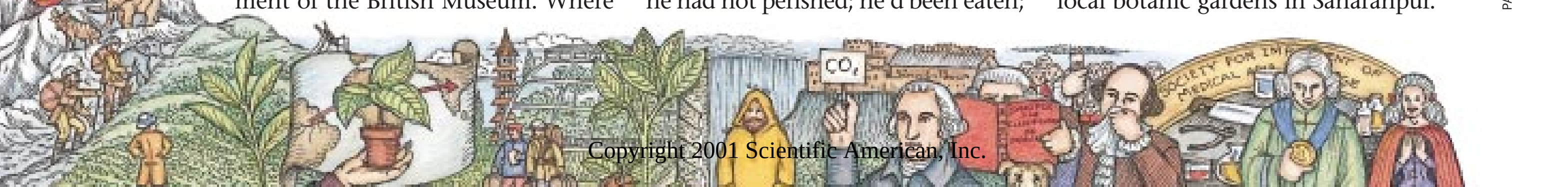
Before his second trip up the jungle (1858 to 1864, and more exploration than evangelism this time), David Livingstone was to develop severe hemorrhoids and ignore the fact. Probably because of which an eventual third trip (1866) was to prove fatal. During this last venture, back home, after three years without news of him, rumors flew: he had perished; he had not perished; he'd been eaten; he was never coming back. Dead or alive, he had boldly gone where no European had gone before, and the public couldn't wait to find out how the story would end. Just the stuff to sell a lot of newspapers.

So in 1869 the down-market, sensation-seeking, highly successful *New York Herald* hired somebody to go find Livingstone. "Somebody" was Henry Morton Stanley: real name John Rowlands, a Welsh drifter (be-

Who else was likely to be sunburnt in Ujiji, on Lake Tanganyika, in October 1871?

friended by the Stanley family of New Orleans), then Civil War turncoat, then U.S. Navy deserter, then (logically enough) journalist. Stanley found his man. "Dr. Livingstone, I presume," he famously said (who *else* was likely to be sunburnt in Ujiji, on Lake Tanganyika, in October 1871?).

This was Stanley's second trip to Africa. In 1868 he'd made his name covering events in Magdala, Ethiopia, where the locals had had the temerity to throw some Englishmen in jail, triggering the arrival of a 13,000-troop British army detachment, complete with artillery, under Gen. Robert Napier (freed the captives, razed Magdala, left), to whom a grateful nation then erected a statue around the corner from my dentist. Early in his career while still a young lieutenant in the Bengal Engineers, in 1831, Napier found himself in the Siwalik Hills on the edge of the Himalayas, with plenty of spare time, which he filled by digging up and sketching fossils, together with Hugh Falconer, who was about to be the boss of the local botanic gardens in Saharanpur.



PATRICIA J. WYNNE

While also becoming a hotshot paleontologist (discovering mastodon, sivatheria, giant tortoises et al.), Falconer was the guy who saved the British Empire by running the top-secret program that finally determined that tea would grow in India (British at the time) and would be every bit as good as that grown in China (not British at the time). Pip-pip. The chap Falconer replaced at Saharanpur Botanic Gardens (who'd left for England) was J. F. Royle, known for his 1839 report recommending that the Brits bring cinchona seedlings from South America to India and also save the British Empire with the other thing it badly needed: quinine, produced from cinchona bark and used to treat malaria.

Back in London it was almost certainly the Royal Botanical Gardens at Kew's obsession with Royle's report on cinchona that scuppered the grand plan presented by Thomas Hancock and his partner Charles Macintosh for an expedition (also to South America) to do the same kind of transfer job on rubber-tree seedlings. At the time, imported rubber came in "bottles," and Hancock had developed a rubber-shredding process to deal with the spare bits you get when using bottle-shaped pieces of rubber. Macintosh had found (1823) that a by-product of coal tar, naphtha, would dissolve these shreds, so you got a rubber paste you could spread between two sheets of cloth and call a raincoat. Given the British weather, this was going to be a winner. The two men did eventually make it to fame and fortune, but only well after the cinchona problem had been licked.

Macintosh had learned his chemistry in Edinburgh at the feet of the great Joseph Black (latent heat, carbon dioxide, Watt's patron and adviser on steam power). Black's mentor (and prof at the same establishment) was William Cullen (synopsis and classification of diseases). One of Cullen's other favorite pupils was George Fordyce, who had 19 aunts and uncles, graduated in 1758 with a paper on the chemistry of catarrh, and ended up at St. Thomas's in London. Where a few years later he co-founded the Society for Improvement of Medical and Chirurgical Knowledge.

About which I know nothing, except that in 1788 its gold medal went to an ex–naval surgeon for his stylish "Dissertation on the Properties of Pus." Maybe the author (Everard Home) had picked up some writing tips from his poetical sister, Anne. SA

Wonders, *continued from page 93*

ambitious extrapolation to the limits. The evolving universe had won out. In its undetected but plausible start, the picture resonated with ancient teaching: "In the beginning, nothing, and then BIG BANG!" Hoyle's put-down had changed into a big bang—before which we could point out no process at all.

As the 1980s opened, a young particle physicist, Alan Guth, showed us the new power of an old result. He used the 1917 work of Willem de Sitter of Leiden, the first theorist to take up Einstein's ideas, even the strange repulsive gravity. De Sitter's solution was exact: a spacetime filled with a repulsive field could expand to huge dimensions in almost no time. Guth called it "inflation," and he showed how it had intervened on the way back to the hot big bang. The enormous inflationary expansion stretched out all chance wrinkles to uniformity, newly filled regions decayed to energized normal matter, and at last attractive gravity could clump the cosmic gas into a myriad of galaxies.

By the mid-1990s the COBE space probe had confirmed these amazing simplicities. The background radiation is uniform to astonishing precision: one part in 100,000 over the entire sky. Today we see many small parts-per-million flaws in the sky map, embryos of our present lumpy world. Headlong inflation had intervened grossly, its energy giving the outward push to the sedate expansion among starry galaxies.

Even the pros still use the "big bang" to allude to the Einsteinian end point, now not to be reached. The term remained in vogue but came to mean an evolving cosmos. We simply do not know our cosmic origins; intriguing alternatives abound, but none yet compel. We do not know the details of inflation, nor what came before, nor the nature of the dark, unseen material, nor the nature of the repulsive forces that dilute gravity. The book of the cosmos is still open. Note carefully: we no longer see a big bang as a direct solution. Inflation erases evidence of past space, time and matter. The beginning—if any—is still unread.

It is deceptive to maintain so long the very term that stood for a beginning out of nothing. The chanteuse will compose a clever new song once the case is clear. Witty Sir Fred, who authored the bang, will not be sorry. SA


DAN WAGNER

Life Savers

A small book contains the wit and wisdom to make even the worst situations just awful, says **Steve Mirsky**

During my last trip to Florida, which took place during that lovely time of year when the recounts begin to bloom, I took a break from television coverage of electoral-college chaos and chanced on a documentary about the less violent world of alligator wrestlers. One grizzled veteran sagely said of his necessarily undefeated record, "There's no such thing as a *pretty good* alligator wrestler."

That quote came to mind shortly after a friend handed me a Christmas present: a thin Florida-orange-colored volume called *The Worst-Case Scenario Survival Handbook,* by Joshua Piven and David Borgenicht. Think helpful hints from Heloise, if Heloise were a U.S. Navy SEAL.

I was deeply touched by my friend's gift of instructions on how to cope with medical and other assorted emergencies and vowed never again to be a passenger in her car. The first item I turned to, at random, was entitled "How to Fend Off a Shark." A confirmed land mammal, I pish-toshed the whole business, as I'm sure not stupid enough to get myself into a position in which I'd need to fend off a shark. I figured the book was a waste of time but riffled further to land quickly at "How to Wrestle Free from an Alligator." Whoa, I thought, now this is information I *am* stupid enough to need.

For a New Yorker who doesn't work in the sewers (as such), I'm near alligators a lot. Once while dreamily strolling through the Loxahatchee National Wildlife Refuge just south of ballot-bouncing West Palm Beach, I stepped within inches of a gator sunning on a levee edge. Yes, you can come dangerously close to a six-foot-long reptile without seeing it if you fully commit to mental vacuity. He exploded down the bank into the water, and I resolved to pay more expletive attention.

Five years ago, researching an Everglades story, I waded through waist-deep, gator-gorged waters with a park ranger named Bob Hicks. As we wandered, Hicks shared the self-knowledge he gained when, backpacking through the 'glades, he put his submerged foot down on top of a slightly more deeply submerged gator. The major part of the insight concerned a discovery about the quality of his screams. "I now know what I sound like when I'm scared," Hicks recalled. "It's not a high-pitched *aaahhhhhhh!!!* It's more of an *uuuhhhh, uuuhhhh, uuuhhhh*. It's like a Moe, Larry, Curly thing." (Oddly enough, I learned the same Stoogey self-truth the night I fell face-first down a dark flight of stairs.)

Anyway, the handbook's unassailable instructions for getting away from an alligator that takes more of an offensive attitude than Hicks's or mine did are fairly straightforward. They include, "If its jaws are closed on something you want to remove (for example, a limb), tap or punch it on the snout." The authors' counsel concludes with "Seek medical attention immediately."

Worst-Case Scenario, which I now rate as a fine addition to my library, covers a variety of such topics, often citing scientific facts. For example, in the section "How to Escape from Quicksand," the authors offer this grainy guideline: "The viscosity of quicksand increases with shearing—move slowly so the viscosity is as low as possible." In "How to Jump from a Bridge or Cliff into a River," they note the importance of going in feetfirst: "If your legs hit the bottom, they will break. If your head hits, your skull will break." And in "How to Perform a Tracheotomy," they advise not to waste time sterilizing whatever instruments you're lucky enough to have: "Infection is the least of your worries at this point."

Unfortunately, in addition to neglecting instructions on how to fall safely face-first down a dark flight of stairs, the handbook fails to address such worstest-case scenarios as "How to Unhang a Chad," "How to Complete an Overseas Absentee Ballot" and "How to Revive Dick Cheney." Perhaps in volume two. SA

"Once and for all I want to know* what *I'm paying for. When the electric company tells me whether light is a wave or a particle I'll write my check."

SIDNEY HARRIS