MARS PATHFINDER LOOKS BACK

SCIENTIFIC AMERICAN

JULY 1998 $4.95

SPECIAL REPORT:
New Victories against HIV

The invisible charms of a winged Don Juan

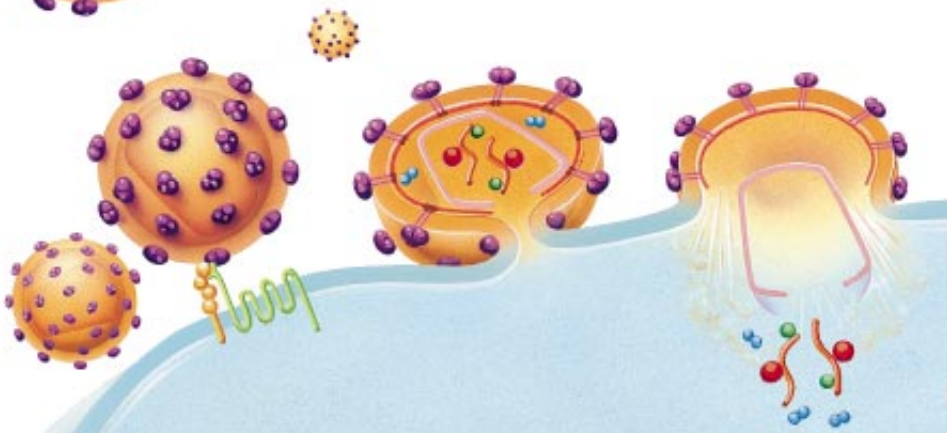
SCIENTIFIC AMERICAN

July 1998 Volume 279 Number 1

NEWS AND ANALYSIS

Plasma helix (page 37)

SPECIAL REPORT



About the Cover
Males of the Orange Sulphur butterfly *Colias eurytheme* are brightly colored, but unlike those of the females, their wings also strongly reflect attractive patterns in the ultraviolet end of the spectrum. Photograph by Dan Wagner.

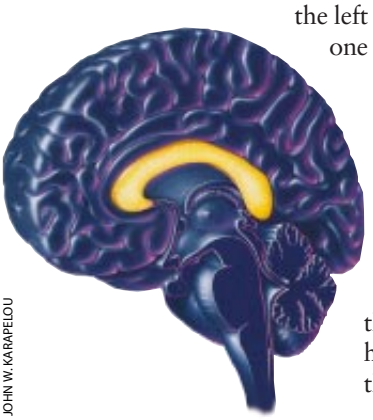
Visit the SCIENTIFIC AMERICAN Web site (http://www.sciam.com) for more information on articles and other on-line features.

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 (outside U.S. $49). Institutional price: one year $39.95 (outside U.S. $50.95). 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; 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.**

All for One

Michael S. Gazzaniga's original article for *Scientific American* on split-brain research, from 30 years ago, might count as one of the most widely influential papers written in modern times about the field of neurology. Not in the strict sense of scientific citation—after all, he was writing a review of experimental findings published long before in professional journals. Neurologists already knew. To the article's huge audience of lay readers, however, it was a revelation.

I first read the article as a student and was flabbergasted. Splitting an alarm clock down the middle would produce two piles of junk. Who would imagine, then, that a longitudinal fission of the brain's delicate higher centers would yield two distinct minds, as if gray matter were some mental amoeba? Equally unsettling, those minds were not identical twins: the left one was verbal and analytical; the right one was a visual and musical artist.



JOHN W. KARAPELOU

HALF A BRAIN is still a whole mind.

The research seemed to say that two different people lived inside everyone's head, and that idea took root in popular culture. Today references to "left-brain thinking" and "drawing with the right side of your brain" are commonplace. Gazzaniga's update on that work, beginning on page 50, shows that the true character of those divorced hemispheres is rather more complex, but the basic insight survives.

Split-brain research follows in the tradition of learning about the brain by seeing what happens when parts of it break down. Annals of neurology are filled with sad, informative cases like that of Phineas P. Gage, a quiet family man in 1848 until an accidental lobotomy by a flying steel rod turned him into a carousing brawler. The brain can survive all manner of assaults, but each can leave our skull with a different occupant.

So as Walt Whitman wrote, "I contain multitudes." In some sense, our heads are home to many potential minds, not just two. The question I've sometimes pondered is where those other people are before the injuries bring them to light. Are they created by the truncated circuitry? Or are they always there, murmuring voices in the chorus of our consciousness?

And yet this is probably a misleading way to understand minds and brains—whole, split or splintered. Our brains work as they do precisely because they are not naturally rent apart. Unlike the people in medical histories, we the uninjured enjoy the choice of finding the best or worst of those other voices within us. The orators, artists, beasts and angels of our nature await their chance.

JOHN RENNIE, *Editor in Chief*
editors@sciam.com

SCIENTIFIC AMERICAN®
Established 1845

John Rennie, EDITOR IN CHIEF

Board of Editors
Michelle Press, MANAGING EDITOR
Philip M. Yam, NEWS EDITOR
Ricki L. Rusting, ASSOCIATE EDITOR
Timothy M. Beardsley, ASSOCIATE EDITOR
Gary Stix, ASSOCIATE EDITOR
Carol Ezzell; W. Wayt Gibbs; Alden M. Hayashi;
Kristin Leutwyler; Madhusree Mukerjee;
Sasha Nemecek; David A. Schneider; Glenn Zorpette
CONTRIBUTING EDITORS: Mark Alpert, Marguerite Holloway,
Steve Mirsky, Paul Wallich

Art
Edward Bell, ART DIRECTOR
Jana Brenning, SENIOR ASSOCIATE ART DIRECTOR
Johnny Johnson, ASSISTANT ART DIRECTOR
Bryan Christie, ASSISTANT ART DIRECTOR
Bridget Gerety, PHOTOGRAPHY EDITOR
Lisa Burnett, PRODUCTION EDITOR

Copy
Maria-Christina Keller, COPY CHIEF
Molly K. Frances; Daniel C. Schlenoff;
Katherine A. Wong; Stephanie J. Arthur

Administration
Rob Gaines, EDITORIAL ADMINISTRATOR
David Wildermuth

Production
Richard Sasso, ASSOCIATE PUBLISHER/
VICE PRESIDENT, PRODUCTION
William Sherman, DIRECTOR, PRODUCTION
Janet Cermak, MANUFACTURING MANAGER
Tanya Goetz, DIGITAL IMAGING MANAGER
Silvia Di Placido, PREPRESS AND QUALITY MANAGER
Madelyn Keyes, CUSTOM PUBLISHING MANAGER
Norma Jones, ASSISTANT PROJECT MANAGER
Carl Cherebin, AD TRAFFIC

Circulation
Lorraine Leib Terlecki, ASSOCIATE PUBLISHER/
CIRCULATION DIRECTOR
Katherine Robold, CIRCULATION MANAGER
Joanne Guralnick, CIRCULATION PROMOTION MANAGER
Rosa Davis, FULFILLMENT MANAGER

Advertising
Kate Dobson, PUBLISHER
OFFICES: NEW YORK:
Thomas Potratz, EASTERN SALES DIRECTOR;
Kevin Gentzel; Randy James;
Stuart M. Keating; Wanda R. Knox.
DETROIT, CHICAGO: Edward A. Bartley, DETROIT MANAGER.
3000 Town Center, Suite 1435, Southfield, MI 48075;
WEST COAST: Lisa K. Carden, WEST COAST MANAGER.
1554 S. Sepulveda Blvd., Suite 212, Los Angeles, CA 90025;
SAN FRANCISCO: Debra Silver.
225 Bush St., Suite 1453, San Francisco, CA 94104
CANADA: Fenn Company, Inc. DALLAS: Griffith Group

Marketing Services
Laura Salant, MARKETING DIRECTOR
Diane Schube, PROMOTION MANAGER
Susan Spirakis, RESEARCH MANAGER
Nancy Mongelli, PROMOTION DESIGN MANAGER

International
EUROPE: Roy Edwards, INTERNATIONAL ADVERTISING DIRECTOR,
London. HONG KONG: Stephen Hutton, Hutton Media Ltd.,
Wanchai. MIDDLE EAST: Peter Smith, Peter Smith Media and
Marketing, Devon, England. BRUSSELS: Reginald Hoe, Europa
S.A. SEOUL: Biscom, Inc. TOKYO: Nikkei International Ltd.

Business Administration
Marie M. Beaumonte, GENERAL MANAGER
Alyson M. Lane, BUSINESS MANAGER
Constance Holmes, MANAGER, ADVERTISING ACCOUNTING
AND COORDINATION

Chairman and Chief Executive Officer
John J. Hanley

Co-Chairman
Rolf Grisebach

Corporate Officers
Joachim P. Rosler, PRESIDENT
Frances Newburg, VICE PRESIDENT

Electronic Publishing
Martin O. K. Paul, DIRECTOR

Ancillary Products
Diane McGarvey, DIRECTOR

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue • New York, NY 10017-1111
(212) 754-0550

PRINTED IN U.S.A.

LETTERS TO THE EDITORS

FIGHTING GERMS

I thank Stuart Levy for firing major artillery in the continuing campaign against antimicrobial-resistant bacteria in his recent article "The Challenge of Antibiotic Resistance" [March]. He points out that our normal bacterial flora fill a niche that is one line of defense against unwanted pathogens. I believe that Levy goes too far, however, when he intimates that the use of household disinfectants and antiseptics can kill "wanted" bacteria and might promote the resistant strains. To the contrary, we should strive to have sterile toys, high chairs, mattress pads and cutting boards. There are no "wanted" microbes in these areas. Sterilizing such items *has* been shown to prevent household spread of infection. But, as Levy points out, the desired agents to sterilize these areas are the alcohol-based products, which evaporate and do not leave residues after they do their job.

WINKLER G. WEINBERG
Infectious Disease Branch
Kaiser Permanente

Levy reports that a major reason for the misuse of antibiotics is that "physicians acquiesce to misguided patients who demand antibiotics." I have found that it is often physicians who administer unnecessary antibiotics. For example, I was recently given a swab test to determine if I had strep throat, a bacterial infection. Instead of recommending that I wait two days for the results, the doctor immediately prescribed antibiotics. He obviously supposed that even if my infection did turn out to be viral, there was no harm in taking the drugs.

It is unfair to blame the patients who request antibiotics—physicians should inform patients that antibiotics may in fact promote resistance and degrade the immune system and should therefore be taken only when required. Education and awareness are crucial to ensure the prudent use of antibiotics.

MITA PATEL
Kanata, Ontario

Levy replies:

In the ideal world, "sterile toys, high chairs" and so on might be desirable, as Weinberg suggests. But complete sterility is impossible in the environment in which we live. Furthermore, this scenario could be risky if, in fact, we need to encounter some microbes to develop the ability to mount an immune response to common pathogens. A clean item, however, is an achievable goal: the numbers of certain harmful bacteria should be reduced so they do not pose a threat to human health.

I fully agree with Patel's final statement. Prescribers and users of antibiotics have both played a role in the problem of antibiotic resistance, and both have a stake in the solution. The physician who acquiesces to a patient's demand for unnecessary antibiotics is not much better than the one who prescribes antibiotics for a viral cold.

THE END OF CHEAP OIL

You must have been wearing blinders when you selected articles for the March special report, "Preventing the Next Oil Crunch." There was absolutely no mention of the increasing harm inflicted on our planet by the extraction, production and consumption of fossil fuels. The irony is that the end of cheap oil may be a good thing. As we use and deplete these fuels, we pollute the air we breathe and the water we drink. Rather than just report on how to wring the last drop of oil from the earth, SCIENTIFIC AMERICAN could have included at least one article that addressed environmental damage and global warming in the context of fossil-fuel use and exhaustion. We will run out of these fuels sooner or later—it would have been good to hear how we might live in a post-fossil-fuel age through conservation and the use of renewable energy sources.

RICHARD REIS
Silver Spring, Md.

LAURIE GRACE

***OIL PRODUCTION** could decline in the next decade.*

Your March coverage of the petroleum scene is rock-solid. I think it is strong testimonial to private enterprise that identification of the impending oil crunch, as well as the various antidotes, all stem from research at the corporate level rather than from government subsidies. Thanks for resisting what must have been a temptation—to present such pork-barrel alternatives as the methanol-from-corn proposals. Gravy for Congressmen, but eventually everybody loses.

DAVID H. RUST
Woodville, Tex.

DYING LANGUAGES

Rodger Doyle's piece on "Languages, Disappearing and Dead" ["By the Numbers," News and Analysis, March] seems to have an omission for the U.S. You left out Appletalk. But I leave it to you to decide whether or not Apple should be classified as "endangered" or "moribund."

DOUG WAUD
Worcester, Mass.

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.

ERRATA

In the map accompanying "Languages, Disappearing and Dead" ["By the Numbers," News and Analysis, March], the data for the British Isles are incorrect. The correct version of the map can be found at http://www.sciam.com/1998/0798issue/0798letters/corrections.html on the World Wide Web.

In "Liquid Fuels from Natural Gas" [March], the company Brown & Root was mistakenly identified as a British company. Brown & Root is a U.S. company headquartered in Houston.

"Japanese Temple Geometry" [May] contains a notational error: in the ellipse problem on page 86, the variables a and b are the semimajor and semiminor axes, respectively.

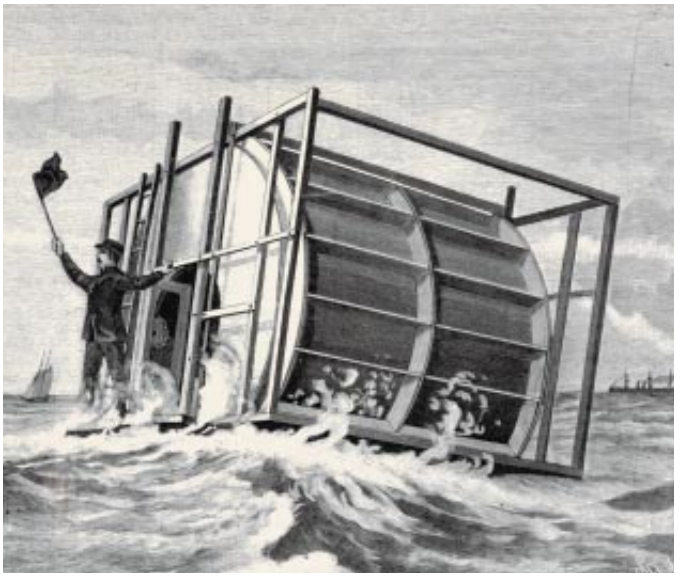
50, 100 AND 150 YEARS AGO

SCIENTIFIC AMERICAN

JULY 1948

ANTIQUITY OF MAN—"Was the beetle-browed Neanderthal man really our ancestor, or an unhappy cousin doomed to extinction? Is *Homo sapiens* a recent arrival in Europe? Last August, in a quiet French village in the Department of Charente, the mystery was solved when a few fragments of an old skull were brushed carefully out of the ancient clays. The most curious fact is that it was a skull very much like your own. There is nothing Neanderthaloid about it. It is within the size range of living females: this woman could have sat across from you on the subway and you would not have screamed. You might even have smiled. The lady of Charente places modern man on the European Continent over 100,000 years ago.—Loren Eiseley"

PRIMORDIAL ATOMS—"Nineteen years after Edwin Hubble's discovery that the galaxies seem to be running away from one another at fabulously high speeds, the picture presented by the expanding universe theory—which assumes that in its original state all matter was squeezed together in one solid mass of extremely high density and temperature—gives us the right conditions for building up all the known elements in the periodic system. According to calculations, the formation of elements must have started five minutes after the maximum compression of the universe. It was fully accomplished, in all essentials, about 10 minutes later.—George Gamow"



Intrepid mariner and his roller-boat

JULY 1898

FEAR—"Of the 298 classes of objects of fear to which 1,707 persons confessed, thunder and lightning lead all the rest. But is there any factual justification for this fear? We believe there is not. As proof we may cite statistics of the United States Weather Bureau. For the years 1890–1893 the deaths from lightning numbered an average of 196 a year. Indeed if one can go by statistics, the risk of meeting death by a horse kick in New York is over 50 per cent greater than that of death by lightning."

SPAIN VERSUS CUBAN GUERRILLAS—"Owing to the peculiar nature of the land in Cuba, a small force is capable of holding a much larger force at bay with the methods of guerrilla warfare that are adopted by the Cuban insurgents against the Spanish soldiers. The armies of Spain have been perpetually harassed by the enemy, and as the Cubans would not meet them in the field, they have devoted their attention to cutting off the various sections of the island to prevent the mobilization of large bodies of insurgent troops; to 'reconcentration,' by which they hoped to starve the Cuban forces by shutting up in the towns the peasants who furnished them with food; and to the protection of large estates and plantations."

ROLL, ROLL, ROLL YOUR BOAT—"The accompanying view is of a roller-boat launched from Bar Harbor, Maine. Our readers will not be surprised to learn that the maiden voyage was disastrous and that after rolling, or rather being blown by the wind, out to sea for fifteen miles, the crew of two were glad to exchange their swinging platform for the solid deck of a seagoing freighter. The vessel consisted of a cylindrical barrel about 10 feet in diameter, built of staves and hooped in the usual barrel fashion. The rolling motion was imparted by hand cranks and gears, and the forward movement of the boat was due to the paddles arranged around the periphery of the barrel."

JULY 1848

SALMON OF OREGON—"Lieut. Howison, of the U.S. Navy, in his report on Oregon, states that the Salmon enter the mouth of the Columbia in May, and make their way up the stream for the distance of twelve hundred miles. The young fry pass out to sea in October, when they are nearly as large as herring. These fish constitute the chief subsistence of many thousand Indians, who reside in the country watered by the Columbia, and its tributaries, and afford an abundant supply to all those and the white settlers of Oregon."

TRAVELS IN BORNEO—"We were escorted through a crowd of wondering Dyaks, to a house in the centre of the village. The structure was round and well ventilated by port-holes in the pointed roof. We ascended to the room above and were taken a-back at finding that we were in the head house, as it is called, and that the beams were lined with human heads, all hanging by a small line passed through the top of the skull. They were painted in the most fantastic and hideous manner. However, the first impression occasioned by this very unusual sight soon wore off, and we succeeded in making an excellent dinner, in company with these gentlemen.—Frank Marryat" [*Excerpted from Marryat's* Borneo and the Indian Archipelago, *published in London in 1848.*]

NEWS AND ANALYSIS

IN FOCUS

STATISTICAL UNCERTAINTY

Researchers warn that continued debate over the 2000 census could doom it to failure

DOUGLAS BURROWS *Gamma Liaison*

ATTEMPTS TO COUNT HOMELESS AMERICANS *in the 1990 census largely failed. The 2000 census will probably do little better.*

Censuses in the U.S. have always seemed straightforward—it's just a head count, right?—and have always proved, in practice, to be just the opposite: logistically complex, politically contentious and statistically inaccurate. Clerks were still tabulating the results of the 1880 census eight years later. The 1920 count revealed such a dramatic shift in population from farms to cities that Congress refused to honor the results. And a mistake in doling out electoral college seats based on the 1870 census handed Rutherford B. Hayes the presidency when Samuel J. Tilden should in fact have been awarded the most votes.

But after 1940 the accuracy of the census at least improved each decade, so that only 1.2 percent of the population slipped past the enumerators in 1980, according to an independent demographic analysis. That trend toward increasing accuracy reversed in 1990, however. The Census Bureau paid 25 percent more per home to count people than it had in 1980, and its hundreds of thousands of workers made repeated attempts to collect information on every person in every house—what is called a full enumeration. Nevertheless, the number of residents left off the rolls for their neighborhood rose to 15 million, while 11 million were counted where they should not have been. The net undercount of four million amounted to 1.8 percent of the populace.

Less than 2 percent might be an acceptable margin of error were it not that some groups of people were missed more than others. A quality-check survey found that blacks, for example, were undercounted by 4.4 percent; rural renters, by 5.9 percent. Because census data are put to so many important uses—from redrawing voting districts and siting schools to distributing congressional seats and divvying up some $150 billion in annual federal spending—all agree that this is a problem.

In response, Congress unanimously passed a bill in 1991 commissioning the National Academy of Sciences (NAS) to study ways to reduce cost and error in the census. The expert

panel arrived at an unequivocal conclusion: the only way to reduce the undercount of all racial groups to acceptable levels at an acceptable cost is to introduce scientific sampling into the April 1, 2000, census and to give up the goal of accounting directly for every individual. Other expert groups, including a special Department of Commerce task force, two other NAS panels, the General Accounting Office and both statisticians' and sociologists' professional societies, have since added their strong endorsement of a census that incorporates random sampling of some kind.

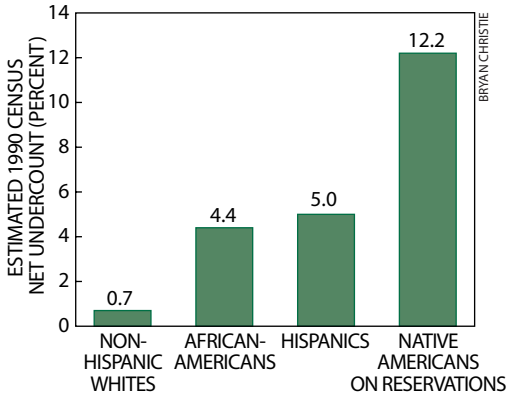
After some waffling, the Census Bureau finally settled last year on a plan to use two kinds of surveys. The first will begin after most people have mailed back the census forms sent to every household. Simulations predict that perhaps one third of the population will neglect to fill out a form—more in some census tracts (clusters of adjacent blocks, housing 2,000 to 7,000 people) than in others, of course. To calculate the remainder of the population, census workers will visit enough randomly selected homes to ensure that at least 90 percent of the households in each tract are accounted for directly.

So if only 600 out of 1,000 homes in a given tract fill out forms, enumerators will knock on the doors of random nonrespondents until they add another 300 to the tally. The number of denizens in the remaining 100 houses can then be determined by extrapolation, explains Howard R. Hogan, who leads the statistical design of the census.

After the initial count is nearly complete, a second wave of census takers will fan out across the country to conduct a much smaller quality-control survey of 750,000 homes. Armed with a more meticulous (and much more expensive) list of addresses than the census used, this so-called integrated coverage measurement (ICM) will be used to gauge how many people in each socioeconomic strata were overcounted or undercounted in the first stage. The results will be used to inflate or deflate the counts for each group in order to arrive at final census figures that are closer to the true population in each region.

"We endorsed the use of sampling [in the first stage] for two reasons," reports James Trussell, director of population research at Princeton University and a member of two NAS panels on the census. "It saves money, and it at least offers the potential for increased accuracy, because you could use a smaller, much better trained force of enumerators." The Census Bureau puts the cost of the recommended, statistics-based plan at about $4 billion. A traditional full enumeration, it estimates, would cost up to $800 million more.

The ICM survey is important, says Alan M. Zaslavsky, a statistician at Harvard Medical School, because it will reduce the lopsided undercounting of certain minorities. "If we did a traditional enumeration," he comments, "then we would in effect be saying one more time that it is okay to undercount blacks by 3 or 4 percent—we've done it in the past, and we'll do it again."



***UNDERCOUNTING** of some racial and ethnic groups that undermined the 1990 census could be reduced by using statistical sampling in the 2000 census.*

Republican leaders in Congress do not like the answers given by such experts. Two representatives and their advocates, including House Speaker Newt Gingrich, filed suits to force the census takers to attempt to enumerate everyone. Oral arguments in one trial were set for June; the cases may not be decided until 1999.

The Republicans' main concern, explains Liz Podhoretz, an aide to the House subcommittee on the census, is "that the ICM is five times bigger than the [quality-check survey performed] in 1990, and they plan to do it in half the time with less qualified people. And it disturbs them that statisticians could delete a person's census data" to adjust for overcounted socioeconomic groups.

Although the great majority of researchers support the new census plan, there are several well-respected dissenters. "I think the 2000 design is going to have more error than the 1990 design," says David A. Freedman of the University of California at Berkeley. The errors to worry about, he argues, are not the well-understood errors introduced by sampling but systematic mistakes made in collecting and processing the data.

As an example, Freedman points out that a computer coding error made in the quality check during the last census would have erased one million people from the country and erroneously moved a congressional seat from Pennsylvania to Arizona had the survey data been used to correct the census. That mistake was not caught until after the results were presented to Congress. "Small mistakes can have large effects on total counts," adds Kenneth W. Wachter, another Berkeley statistician.

"There are ways to improve the accuracy without sampling," Podhoretz asserts. "Simplifying the form and offering it in several languages, as is planned, should help. They should use [presumably more familiar] postal workers as enumerators. They should use administrative records, such as welfare rolls."

"That shows appalling ignorance," Trussell retorts. "Our first report addressed that argument head-on and concluded that you cannot get there by doing it the old way. You're just wasting a lot of money."

Representative Dan Miller of Florida was planning to introduce a bill in June that would make it illegal to delete any nonduplicated census form from the count. Such a restriction would derail the census, Trussell warns. "The idea behind sampling is not to eliminate anybody but to arrive at the best estimate of what the actual population is. Surely the goal is not just to count as many people as possible?"

As the debate drags on, the brinkmanship is making statisticians nervous. Podhoretz predicts that "some kind of a showdown is likely next spring." That may be too late. "You don't want to redesign a census at the last minute," Freedman says.

"I think the two sides should just agree to flip a coin," Trussell says. "To think next year about what we're going to do is madness." Wachter concurs: "We must not let the battle over sampling methods destroy the whole census." Otherwise April 1, 2000, may make all involved look like April fools.

—W. Wayt Gibbs in San Francisco

COSMOLOGY

INFLATION IS DEAD; LONG LIVE INFLATION

How an underdense universe doesn't sink cosmic inflation

Over the past year, observational astronomers have at last convinced theorists that the universe contains less matter than the theory of inflation predicts. The expansion of the universe, as traced by distant supernovae and radio-bright galaxies, is decelerating too slowly. The mass of galaxy clusters, as deduced from their internal motions and their ability to focus the light of more distant objects, is too low. The number of these clusters, which should be growing if there is sufficient raw material, has changed too little. And the abundance of deuterium, which is inversely related to the total amount of matter, is too high. It seems there is only a third of the matter needed for geometric flatness, the expected outcome of inflation.

But far from killing the theory, cosmologists say, the observations make it more necessary than ever—albeit in a new form. No other theory answers a nagging question in big bang cosmology: Why is the universe even vaguely flat? Over time, the cosmos should seem ever more curved as more of it comes into view and its overall shape becomes more apparent. By now, billions of years after the big bang, the universe should be highly curved, which would make it either depressingly desolate or impenetrably dense.

Inflationary theory—developed in the early 1980s by Alan H. Guth, now at the Massachusetts Institute of Technology, and Andrei D. Linde, now at Stanford University—solved the problem by postulating that the universe went through a period of accelerating expansion. Once-adjacent regions separated faster than light (which space can do—Einstein's special theory of relativity applies to speeds within space). As a result, we now see only a fragment of the cosmos. Its overall shape is not visible yet; each fragment looks flat. Inflation also explains the near uniformity of the universe: any lumpiness is too large scale for us to perceive.

But if observers can't find enough matter to flatten space, theorists must draw one of two awkward conclusions. The first is that some new kind of dark matter makes up the difference. The inferred matter goes by the name of "quintessence," first used in this general context by Lawrence M. Krauss of Case Western Reserve University. The usage alludes to Aristotelian ether; besides, anything that accounts for two thirds of physical reality is surely quintessential.

Quintessence joins the two previously postulated kinds of dark matter: dim but otherwise ordinary matter (possibly rogue brown dwarfs) and inherently invisible elementary particles (possibly neutrinos, if these ghostly particles have a slight mass). Both reveal themselves only by tugging at visible stars and galaxies. About quintessence, scientists know even less. Cosmic flatness dictates that it contain energy but

LINDA A. CICERO *Stanford News Service*

***STEPHEN W. HAWKING** and other cosmologists struggle to explain a low value of Ω, the matter density of the universe.*

IN BRIEF

Dust Impact

Good news for the producers of *Deep Impact:* special effects in the sequel could cost much less. Some now credit cold weather caused by cosmic dust with prompting regular mass extinctions. In *Science* in May, Stephen J. Kortenkamp of the Carnegie Institution and Stanley F. Dermott of the University of Florida refined a three-year-old model, which posited that the earth's orbit tilts every 100,000 years, sending the planet through a sun-blocking "dust plane" and into an ice age. They determined that the shape of the earth's orbit—not its tilt—is what matters: when its orbit becomes more circular every 100,000 years, the planet accumulates more dust.

Endangered No More

Never heard of the Missouri bladderpod? Well, there weren't too many of them around until recently. Soon these creatures—along with dozens of other plant and animal species—will be struck from the government's official endangered list. Interior Secretary Bruce Babbitt announced the plan in May. Among the populations most likely to be declared at least partially recovered are the gray wolf, the bald eagle and the peregrine falcon (*photograph*).

WILLIAM H. MULLINS, *Photo Researchers, Inc.*

Wishing on a Star

You might as well have been if you bought the name for one from the International Star Registry (ISR). New York City Consumer Affairs Commissioner Jules Polonetsky has recently issued a violation against the Illinois-based firm for engaging in deceptive trade practices—the first legal action taken against it. ISR charges anywhere from $50 to $100 for the privilege of christening a star. For the money, customers receive a copy of "Your Place in the Cosmos," a listing of stars and their ISR-bought names. The problem is that only the International Astronomical Union has the right to assign star names—and they're not willing to sell it.

More "In Brief" on page 22

does not specify what kind; the universe's expansion and galaxy clustering imply that quintessence exerts a gravitational repulsion and shuns ordinary matter.

A form of quintessence was already thought to have powered inflation and then died out, begetting ordinary matter. Now it may be back, challenging its progeny for control of the universe. If quintessence wins, the universe will expand forever in a new round of inflation. Our fate hinges on what makes up quintessence. The simplest possibility, Einstein's cosmological constant, inexorably gains in relative strength as cosmic expansion dilutes matter. But other forms of quintessence, such as featherweight particles or space-time kinks, might eventually fade away. In May, Christopher T. Hill of Fermi National Accelerator Laboratory speculated that the quintessence mystery is related to another: the neutrino mass.

So far the only proof for quintessence is circumstantial. The latest supernova observations suggest that cosmic expansion is accelerating, and recent cosmic microwave background measurements show that triangles may indeed subtend 180 degrees, as they should in flat space.

But the lack of direct proof—as well as an observed shortage of gravitational lenses, which suggests the universe is smaller than certain forms of quintessence would make it—has led many cosmologists to a different awkward conclusion: maybe inflation stopped before making space exactly flat. In traditional inflation, this would make the universe 100,000 times too lumpy. The new trick is to kill the two birds with two stones: to suppose that the uniformity of the universe does not result from the same process as its shape does. Maybe the cosmos was made uniform by a previous round of inflation, was uniform from birth or has a special shape that let it even itself out quickly.

Two-round inflationary theory was developed in 1995 by two teams: Martin Bucher of Princeton University, Neil G. Turok, now at the University of Cambridge, and Alfred S. Goldhaber of the State University of New York at Stony Brook; and Kazuhiro Yamamoto of Kyoto University and Misao Sasaki and Takahiro Tanaka of Osaka University. In this theory, the first round creates a uniform mega-universe. Within it, bubbles—self-contained universes—spontaneously form. Each undergoes a second round of inflation that ends prematurely, leaving it curved. The amount of curvature varies from bubble to bubble.

The second idea, announced in February by Turok and Stephen W. Hawking of Cambridge, is that the smooth universe gurgled not out of a soda universe but out of utter nothingness. Updating Hawking's decade-old work on creation ex nihilo, they devised an "instanton"—loosely speaking, a mathematical formula for the difference between existence and nonexistence—that implied we should indeed be living in a slightly curved universe.

Finally, maybe the universe has an unusual topology, so that different parts of the cosmos interconnect like pretzel strands. Then the universe merely gives the illusion of immensity, and the multiple pathways allow matter to mix together and become smooth. Such speculation dates to the 1920s but was dusted off two years ago by Neil J. Cornish of Cambridge, David N. Spergel of Princeton and Glenn D. Starkman of Case Western Reserve.

Like all good cosmological theories, these ideas lead to some wacky conclusions. The bubble and ex nihilo universes are infinite, which quantum laws forbid. The solution: let the universe be both infinite and finite. From the outside it is finite, keeping the quantum cops happy; inside, "space" takes on the infinite properties of time. In the pretzel universe, light from a given object has several different ways to reach us, so we should see several copies of it. In principle, we could look out into the heavens and see the earth.

More worrisome is that these models abandon a basic goal of inflationary theory: explaining the universe as the generic outcome of a simple process independent of hard-to-fathom initial conditions. The trade-off is that cosmologists can now subject metaphysical speculation—including interpretations of quantum mechanics and guesses about the "before"—to observational test.

Out of all this brainstorming may emerge an even deeper theory than standard inflation; by throwing a wrench into the works, observers may have fixed them. Upcoming high-resolution observations of the microwave background and galaxy clustering should be decisive. But if not, cosmologists may begin to question the underpinnings of modern physics. "If the experimental data is inconsistent with literally everything, this may be a signal for us to change gravity theory—Einstein theory," Linde says. —*George Musser*

VIRTUAL ANTHROPOLOGY

FACE OFF

Three-dimensional imaging stands in for fossils

GERHARD WEBER Institute of Human Biology/University of Vienna

***VIRTUAL FOSSIL SKULL** of an ancient human from Petralona, Greece, shows hidden features such as sinuses, and an "endocast" enables brain analysis.*

For scientists who study human evolution, fossil remains provide the only direct evidence of our ancient ancestors. Access to these paleoanthropological Rosetta stones, however, is limited by protective curators who are often reluctant to lend the fragile fossils. And in the case of fossil skulls, nature preserves critical information in the largely inaccessible interior. But help is on the way. At the annual meeting of the American Association of Physical Anthropologists in Salt Lake City this past April, researchers discussed how medical imaging, virtual reality and computer-controlled modeling technologies get around these obstacles noninvasively.

Three-dimensional medical imaging based on computed tomography (CT) scans was developed in the early 1980s. On a computer, surgeons could electronically remove the patient's soft tissue and then explore the virtual skull inside and out before operating. It wasn't long before Glenn C. Conroy of Washington University and his colleagues demonstrated that these same techniques could also be applied to fossils, in which sediments take the place of soft tissue.

With advances in computer graphics and computational power, paleoanthropologists can now perform on their computers a wide range of investigations that are impossible to attempt on the original fossil. Missing features on one side of the skull can be re-created by mirroring the preserved features (postmortem deformations can be similarly rectified), and tiny, hidden structures such as the inner ear can be magnified for closer examination. Moreover, as Christoph P. E. Zollikofer and Marcia S. Ponce de León of the University of Zurich and others have shown, anthropologists can reconstruct fragmented fossils on-screen.

The standard repertoire of measurements can also be made virtually, in most cases with the same degree of accuracy afforded by handheld calipers. And with the creation of a virtual "endocast," brain volume can be determined reliably. In fact, Conroy's recent research has revealed a major discrepancy between the estimated and actual brain volume of an early hominid called Stw 505 (or Mr. Ples). Conroy suspects that the estimated cranial ca-

In Brief, *continued from page 20*

Sleep-Hearing

How is it that an infant's whimper will wake a mother in deep sleep, but an alarm clock might not? To find out, Serena J. Gondek, an undergraduate student at Johns Hopkins University, and her supervisor, Gregory L. Krauss, tested five patients who were scheduled for surgery to correct epileptic seizures and so had electrode grids implanted directly onto their brains. They exposed the subjects to various tones and mapped the patterns of activation onto MRI and CT scans. They found that during waking, only areas around the primary auditory cortex were used. During sleep, however, regions of the frontal lobe also responded. Gondek guesses that this site may analyze sounds to determine whether a sleeper must be roused to react appropriately.

Rerouting Electric Cars

The Partnership for a New Generation of Vehicles, a consortium of the government and major U.S. automakers, has hit a roadblock. The midsize car they have come up with—a diesel-electric hybrid that can get up to 80 miles (129 kilometers) to the gallon and meet current emissions standards—is simply too expensive to compete in the U.S. market. All-electric may not be the way to go, either. Alexander Domijan of the University of Florida recently found that even if only one in 10 Floridians switched to electric vehicles (EVs), problems would arise: given the state's current power configurations, whenever an EV owner decided to charge up the buggy, the neighbors' lights might dim.

KIM KULISH *Sygma*

Wonder Spuds

Results from the primary phase of the first human clinical trials of vaccine-containing foods are in: eating potatoes genetically implanted with a vaccine produces immunity to specific diseases. The Boyce Thompson Institute for Plant Research, an affiliate of Cornell University, developed the wonder spuds, which were administered at the University of Maryland. In 10 of 11 test subjects, antibodies against the bacterium *Escherichia coli* rose at least fourfold in their blood.

More "In Brief" on page 28

pacity of some other fossils might also be incorrect—a hunch that, if substantiated, could have important implications for our understanding of brain evolution.

With a virtually manipulated fossil on the computer screen, the next step for the cyberanthropologist is real virtuality. By using a rapid-prototyping technology called stereolithography, the 3-D image on screen can be re-created physically in a transparent, laser-cured resin. Because the resin is cured layer by layer, all the internal features can be replicated, including the braincase, sinuses and nerve canals. "With this technology, we see things that no one has seen before, because we're seeing through the bone itself," observes Dean Falk of the State University of New York at Albany. Falk, Horst Seidler and Gerhard W. Weber of the University of Vienna's Institute of Human Biology and their colleagues have published data suggesting that skulls that resemble one another externally may differ internally. According to the researchers, studying these internal morphological features on stereolithographic models could provide insight into the hotly debated origins of Neanderthals and modern humans. (It might also illuminate the recent finding that Neanderthals may have had a language ability similar to that of modern humans, based on the size of a nerve canal leading to the tongue.)

What many find most attractive about these techniques is the possibility of creating a digital archive containing the CT scan data for important fossils all over the world. With Internet access and the proper software, researchers could download any fossil, perform their own measurements and manipulations, and, if equipped with a stereolithographic apparatus, even create a hard copy. These techniques, however, are not meant to replace calipers and fossil casts (or the originals). Rather they complement the traditional methods, which are still more reliable for certain analyses, such as those requiring detail at the submillimeter level.

Although interest in these techniques is widespread, equipment and software costs are still prohibitive for most institutions. Because sophisticated image-analysis software alone sports a $15,000 price tag and production costs for each stereolithographic model total around $3,000, only a handful of labs now conduct research virtually. Prices, however, are dropping as a greater market is established. "In five to 10 years," Weber predicts, "every important institute working with fossil material will work with these methods." —*Kate Wong*

DEMOGRAPHICS

WHERE HAVE ALL THE BOYS GONE?

The mysterious decline in male births

Despite their macho swagger, males are the more fragile sex of the human species. Male fetuses are less likely than females to come to term: although 125 males are conceived for every 100 females, only about 105 boys are born for every 100 girls. In the first half of this century, improvements in prenatal care reduced the number of miscarriages and stillbirths and hence increased the proportion of baby boys in most industrial countries. But since 1970 the trend has reversed: in the U.S., Canada and several European countries, the percentage of male births has slowly and mysteriously declined.

So far the decrease has not been alarmingly large. In the U.S. in 1970, 51.3 percent of all newborns were boys; by 1990, this figure had slipped to 51.2 percent. But in Canada the decline has been more than twice as great, and similar long-term drops have been reported in the Netherlands and Scandinavia. The U.S. and Canadian data were compiled by Bruce B. Allan, an obstetrician-gynecologist at Foothills Hospital in Calgary, Alberta. Allan claims the widespread nature of the decline suggests that it is more than a statistical fluctuation. "We can't deny that the percentage of boys is falling," Allan says. "But the question is, Why?"

Demographic factors may be playing a role. Different races have slightly different birth ratios: blacks tend to have fewer boys than whites, whereas Asians have fewer girls. (These differences have been observed worldwide.) The parents' ages may also influence the gender of their offspring; studies have shown that older fathers sire fewer sons than young dads. But Allan found that demographic changes in the Canadian population between 1970 and 1990 could not account for the decline in the percentage of baby boys there.

ED LALLO *Liaison International*

NEWBORN BOYS
have become slightly less common in many countries, and no one knows why.

Some researchers believe pollution may be the culprit. A recent article in the *Journal of the American Medical Association* notes that high exposures to certain pesticides may disrupt a father's ability to produce sperm cells with Y chromosomes—the gametes that beget boys. Other toxins may interfere with prenatal development, causing a disproportionate number of miscarriages among the frailer male embryos. (XY embryos require hormonal stimulation to produce masculine genitalia, which may make the unborn males more vulnerable to hazardous chemicals.)

Perhaps the most striking example of a lopsided birth ratio occurred in Seveso, Italy, where a chemical plant explosion in 1976 released a cloud of dioxin into the atmosphere. Of the 74 children born to the most highly exposed adults from 1977 to 1984, only 35 percent were boys. And the nine sets of parents with the highest levels of dioxin in their blood had no boys at all.

Devra Lee Davis, a program director at the World Resources Institute and one of the authors of the *JAMA* article, argues that the declining male birth ratio should be viewed as a "sentinel health event"—a possible indicator of environmental hazards that are difficult to detect by other means. But other researchers say the link between pollution and birth ratios is not so clear. Fiona Williams, an epidemiologist at the University of Dundee in Scotland, found no correlation between birth ratios and levels of air pollution in 24 Scottish localities. Although very high levels of certain pollutants may reduce the percentage of baby boys, she concludes, one cannot assume that lower exposures will have a similar effect.

To solve the mystery of the missing boys, scientists are calling for more detailed regional analyses of birth ratios. In Canada the falloff has been greatest in the Atlantic provinces; in the U.S. it has been most pronounced in the Midwest, the Southeast and the Pacific states. One provocative theory is that the decline in the male birth ratio has been caused by a continentwide dip in the frequency of sex. When couples have sex more often, fertilization is more likely to occur early in the menstrual cycle, which apparently increases the odds of male conception. Some observers believe this conjecture explains why the percentage of baby boys has usually increased after major wars. *—Mark Alpert*

PHYSICS

EINSTEIN'S DRAG

Two satellites reveal how Earth's rotation warps space-time

In 1918 Austrian physicists Joseph Lense and Hans Thirring derived from Einstein's equations of general relativity that an object that spins also twists the fabric of space-time around it. The Lense-Thirring effect is so small, however, that it has been hard to measure. An international team of Italian, Spanish and Greek-American scientists reported in the March 27 issue of *Science* the most precise detection of the Lense-Thirring effect yet. In an elegant approach they measured how the rotation of Earth distorted space-time and thereby altered the paths of two orbiting satellites.

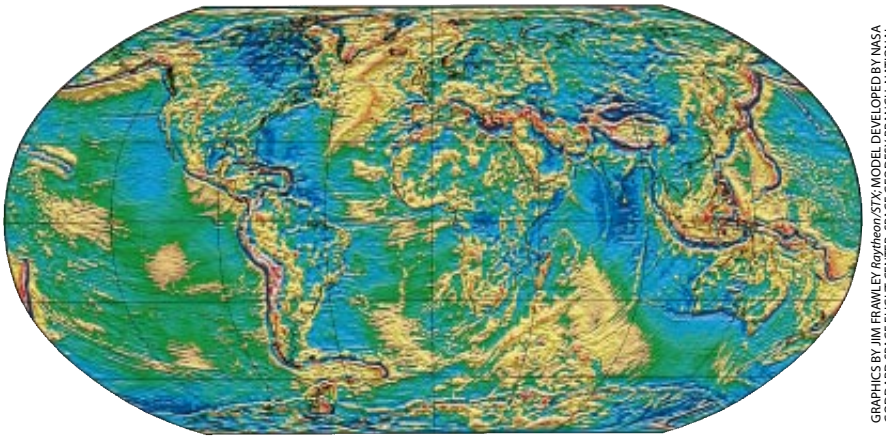
Scientists have known for years where to look for the effect. Calculations show that the rotation of the sun shifts Mercury's orbit six meters a year; unfortunately, that is too small to be detected by today's instruments. Massive objects such as black holes should also demonstrate the effect. Indeed, last year astronomers determined that a disk of gas spiraling into rapidly spinning black holes or neutron stars can precess, in line with the Lense-Thirring effect. Nevertheless, researchers of such observational methods concede there is at least a 400 percent error range.

The recent measurements of the satellites, however, are within about 20 percent of the prediction. The team, led by Ignazio Ciufolini of the National Research Council of Italy and of Rome University, worked with prior laser measurements of the orbits of two satellites originally designed to measure the size and shape of Earth—LAGEOS (Laser Geodynamics Satellite, launched in 1976) and LAGEOS II (sent up in 1992). The laser pulses determined the positions of the satellites with uncertainties of less than one centimeter. The team found that the satellites' trajectories shifted two meters a year because of the spinning Earth.

Crucial to the measurement was canceling ordinary gravity perturbations that masked the relativistic effect. To do that, the workers relied on the Earth gravitational model (EGM-96). Developed by several U.S. institutions, the model estimates the shape and gravitational field of Earth based on orbital data of spacecraft collected over four years. "The experiment would not have been successful without the data provided" by the latest Earth model, says team member Juan Pérez-Mercader of the Laboratory of Astrophysics and Fundamental Physics (LAEFF) in Madrid. The other researchers are Eduardo Fernandes-Vieira of LAEFF, Federico Chieppa of the University of Rome and Erricos Pavlis of the University of Maryland.

One interesting fact about the work is the cost. "Essentially we did not spend one cent, excepting our salaries, travel, computer calculations and the phone calls," Ciufolini says. Pérez-Mercader explains that the work was mainly coordinated through the Internet, a virtual collaboration among members who also share Mediterranean origins—they called themselves *ulivi* ("olives") and *aceitunos* ("olive trees").

Ciufolini, who published previous but less accurate measurements in 1996, is convinced that his team's approach can be improved to get the error down to



GRAPHICS BY JIM FRAWLEY *Raytheon/STX*; MODEL DEVELOPED BY NASA GODDARD SPACE FLIGHT CENTER, SPACE GEODESY BRANCH; NATIONAL IMAGERY AND MAPPING AGENCY AND OHIO STATE UNIVERSITY

EARTH GRAVITATIONAL MODEL *mapping the acceleration caused by gravity over the color spectrum shows that gravity varies from about +0.04 percent* (red) *to –0.03 percent* (blue).

about 2 percent. The key would be a soccer ball–size satellite he calls LARES, or Laser Relativity Satellite, which could be built for less than $10 million. It would contain only mirrors to reflect laser pulses. The orbit of LARES, combined with that of LAGEOS, could cancel almost all the gravity perturbations.

The expense stands in stark contrast to that of the Gravity Probe B, scheduled for launch in 2000. Being built by Stanford University and the National Aeronautics and Space Administration, the probe has taken more than 25 years to develop and will cost about $300 million. The satellite, which will follow a polar orbit, will rely on four gyroscopes the size of haricot beans. In principle, researchers would be able to detect after a year a Lense-Thirring shift in the orientation of the gyroscopes of 42 milliarc seconds, or 0.000012 degree, with an error of 1 percent. That accuracy, Ciufolini calculates, corresponds to measurements in six months of less than 10^{-11} meter—smaller than the radius of a hydrogen atom. (The probe will also measure the curvature of space-time caused by Earth.)

Francis Everitt, one of the main Gravity Probe B researchers, noted that some skepticism about the laser measurement of LAGEOS satellites exists. "I know there is some disagreement among some experts in the data-reduction processing involved in these results," Everitt says. "And I would certainly discourage any different experiment from the way we are doing it. But I think that the observation of satellites may work if the results are confirmed."

Ciufolini and his team do not see Gravity Probe B as the competition, but rather as a complement. Although the design philosophies differ, they have the same goal. "Our results are good news for everybody," Pérez-Mercader remarks. "Gravity Probe B will check the effect in a different way, and that is how science advances."

—Luis Miguel Ariza in Madrid

In Brief, *continued from page 22*

Pretty Big Bang

Twelve billion light-years away, it was the biggest cosmic explosion since the big bang, astronomers say. S. George Djorgovski and Shrinivas R. Kulkarni of the California Institute of Technology reported in *Nature* in May that a gamma-ray burst detected on December 14, 1997, by the Italian-Dutch Beppo-SAX satellite and NASA's Compton Gamma Ray Observatory satellite released several hundred times more energy than an exploding star. The burst, dubbed GRB 971214, lasted less than a minute but for a second was as luminous as the rest of the universe. The scientists do not yet know what causes gamma-ray bursts, but this latest example is making them revise theoretical models.

Ulcers and Heart Disease

As if one condition weren't bad enough. Vincenzo Pasceri and his colleagues at the Catholic University of the Sacred Heart in Rome found evidence of the ulcer-causing bacterium *Helicobacter pylori* (*left*) in 62 percent of the people they tested with heart disease. In comparison, *H. pylori* affected only 40 percent of matched control subjects. Moreover, heart disease was more common among those infected with a particularly virulent strain of the bacterium, one containing a gene called *CagA*. Despite the correlation, Pasceri says patients with *H. pylori*–related ulcers should not necessarily be tested for heart disease.

A. B. DOWSETT SPL/Photo Researchers, Inc.

Virtual Segregation

A disturbing new survey conducted by Donna L. Hoffman and her colleagues at Vanderbilt University's Owen Graduate School of Management found a huge racial divide among Internet users. Whereas 75 percent of the white students they polled had a personal computer at home, fewer than a third of the African-American students they asked did. And even white students who did not own a computer were three times more likely than black students in similar situations to access the World Wide Web from elsewhere. In all, they found that among the roughly 62 million Americans surfing the Web, only five million are black, whereas some 41 million are white. *—Kristin Leutwyler*

ANTI GRAVITY

Gorilla in Our Midst

History was made in April when Koko, the signing gorilla, took part in the first live Internet "chat" between humans and another species, on America Online. Koko responded to questions posed by AOL subscribers, sometimes in a fashion that required elaboration by her mentor, Francine "Penny" Patterson. Recently SCIENTIFIC AMERICAN uncovered sections of the transcript that were mysteriously excised from the official, published version. We print them in the interests of better interspecies communication and to fill a gaping two-column hole left in our news section as we go to press.

To appreciate the value of the "lost Koko transcripts," here first are sections of the actual AOL chat of April 27.

AOL Member: Koko, are you going to have a baby in the future?

Koko: Pink.

AOL Member: I'd like to know what you'd like for your birthday.

Koko: Birthday. Food and smokes.

Dr. Patterson: You have to understand … Smoke is also the name of her kitten.

AOL Member: Do you feel love from the humans who have raised you and cared for you?

Dr. Patterson: She's reading a birthday card.

Koko: Lips, apple give me.

The recovered transcript remnants provide more detail and insight into Koko's thinking.

Recovered section 2:

Koko: Yes, Smoke is a kitten. But when I said smokes, I meant smokes. Cigars. Cubans, in fact. I know a guy who knows a guy who brings them in from Toronto.

Gourmandy: What do you eat?

Koko: I'm a vegetarian.

Pittyting: What about having a baby?

Koko: Pink. Like I said. I'm being ironic of course, poking fun at human gender stereotypes. I mean, I'd like a girl.

Washoerules: What bothers you?

Koko: Grad students. I am not an animal. Well, you know what I mean.

Recovered section 7:

AnnSully: Is signing hard to learn?

Koko: I continue to confuse "heuristic" with "hermeneutic."

MCrawford: Can you read?

Koko: I find Woody Allen's early writings piquant. Hemingway used little words to say big things. I've dabbled in Chomsky but find him pedantic, and I disagree with fundamental aspects of his theses. Goodall raises some interesting issues.

HennyYman: Where does a big gorilla like you sleep, anyway?

Koko: Wait for it … anywhere I want. Of course.

Recovered section 11:

Bigstick99: Do you do any sports?

Koko: I get some exercise. I enjoy jumping up and down on luggage. I also enjoy throwing luggage.

NobelLore: Do you follow the current scientific scene?

Koko: Unless a finding is published in the major journals, one is unlikely to find mention of it in popular reportage. I therefore attempt to browse the primary literature when possible. Thank God for the Internet, eh? LOL.

Recovered section 14:

Host: What did you think of your chat experience?

Koko: Frankly, I found it a bit jejune. I avoid chat rooms. I usually log on only to retrieve e-mail and check my stocks.

Host: Thank you. By the way, what is your e-mail address?

Koko: I don't give that out.

Host. Anything else you'd like to say?

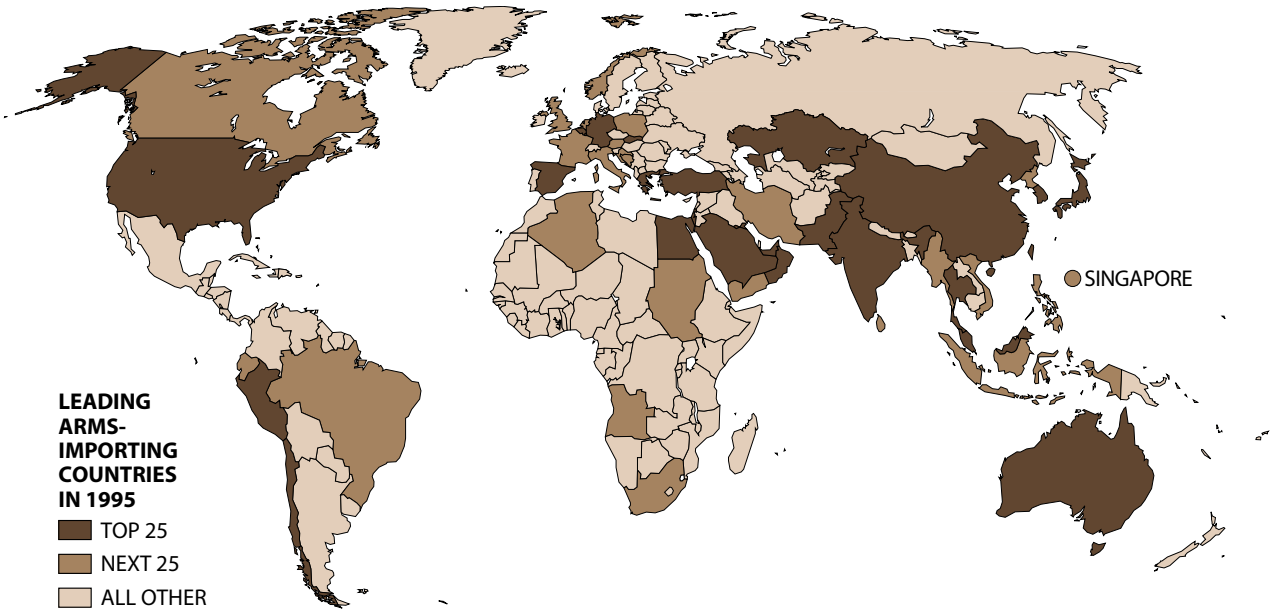
Koko: Lips loose ships sink.

Host: What's that?

Koko: Good night. *—Steve Mirsky*

MICHAEL CRAWFORD

The Arms Trade



RODGER DOYLE

The proliferation of arms to developing countries continues to be a matter of concern, despite a substantial drop in arms shipments since the mid-1980s. Arms sales can lead to major tactical problems, as France discovered in the Gulf War when it could not use its top-of-the-line Mirage fighters because they were indistinguishable from Mirages sold to Iraq. Even worse is the effect on fragile traditional societies, for example, that of Saudi Arabia, which by some accounts is in danger of being destabilized by its substantial expenditures on arms.

Of more immediate concern is the possibility that some rogue state or terrorist group will use weapons of mass destruction. At least 10 countries—China, Taiwan, North Korea, Iran, Iraq, Syria, Israel, Egypt, Libya and Russia—are known to have or are suspected of having a biological warfare capability. All of these, plus Vietnam, are thought to have a chemical warfare capability. The proliferation of such weapons illustrates one of the difficulties of arms limitation: the knowledge that underpins higher living standards—knowledge, for instance, of pharmaceutical and chemical manufacturing—is the same as that needed to produce anthrax bacteria or the nerve gas sarin.

Governments, of course, have a strategic interest in transferring arms to allies, as in the case of American weapons sent to Egypt, Israel and Saudi Arabia. They also tend to promote foreign arms sales to keep their domestic arms suppliers viable. This is certainly true of the major Western suppliers, including the U.S., and of Russia, whose most technologically competitive industry is arms manufacturing. Because of increased excess capacity since the end of the cold war, a buyer's market exists not only for arms but also for the sale of techniques and expertise needed to build up domestic arms production.

A number of international agreements, including the Nuclear Nonproliferation Treaty, the Biological and Toxin Weapons Convention and the Chemical Weapons Convention, are designed to prevent weapons proliferation. Several international organizations are now working on problems of proliferation as well. In addition to the European Union, NATO and the United Nations, they include the Nuclear Suppliers Group; the Missile Technology Control Regime; the Australia Group, which deals with biological and chemical weapons; and the Wassenaar Arrangement, which handles conventional weapons and technologies that can be used for both peaceful and military purposes.

The question that naturally arises is how successful these groups can be when they are undercut by the actions of their own governments. Reports earlier this year disclosed that, despite pleas from Vice President Al Gore and other high U.S. officials, Russia is helping India build submarine-launched ballistic missiles capable of carrying nuclear warheads deep into Pakistan. At about the same time, there was a report alleging that the Clinton administration had been quietly circumventing a new law intended to keep supercomputers (which can be used to design ballistic missiles) out of the hands of Russia, China and other countries that do not control their exports properly or have failed to ratify the Nuclear Nonproliferation Treaty.

The data shown on the map are estimates from the U.S. Arms Control and Disarmament Agency (ACDA), which is now in the process of revising certain figures upward. Although the data understate the true level in some cases, they nonetheless provide a roughly accurate picture of arms transfers in 1995. Imports by countries that support clandestine or black-market purchases may be understated, according to the ACDA. The black market in arms has become more important since the late 1980s, when Egypt, Iran, Iraq, Libya, Pakistan and Syria, among others, purchased Western technologies for developing nuclear weapons, ballistic missiles and chemical weapons. In the 1990s considerable black-market conventional arms flowed to the contending parties in the former Yugoslavia, Somalia, Rwanda and Congo (formerly Zaire).

—Rodger Doyle (rdoyle2@aol.com)

PROFILE

Big Tobacco's Worst Nightmare

Industry secrets exposed by Stanton A. Glantz helped to put tobacco companies on the run. Show them no mercy, he urges

"They are attacking us for things we didn't say!" Two minutes into a phone conversation, and Stan Glantz is already yelling into the receiver, even though the door to his office is wide open and a reporter is sitting nearby. As he excitedly plots with a colleague the best rejoinder to critics of his latest research paper, Glantz leans back in his chair until his short legs leave the floor. One foot starts swinging as an incongruous smile gradually betrays his pleasure at the controversy.

Centered above Glantz's cluttered desk, at the spot where a professor of medicine and cardiology at the University of California at San Francisco would normally hang his Ph.D., sits a large, red steel cabinet labeled "Fire Alarm Control." How fitting. Not just because this office, hardly bigger than a janitor's closet and packed almost to the ceiling with tens of thousands of documents that American cigarette companies would just as soon see burn, is itself arguably a fire hazard. But what better symbol for a tobacco-control researcher to stare at all day as he tallies the damage that tobacco use inflicts on society and as he plots how best to stamp out cigarette smoking forever?

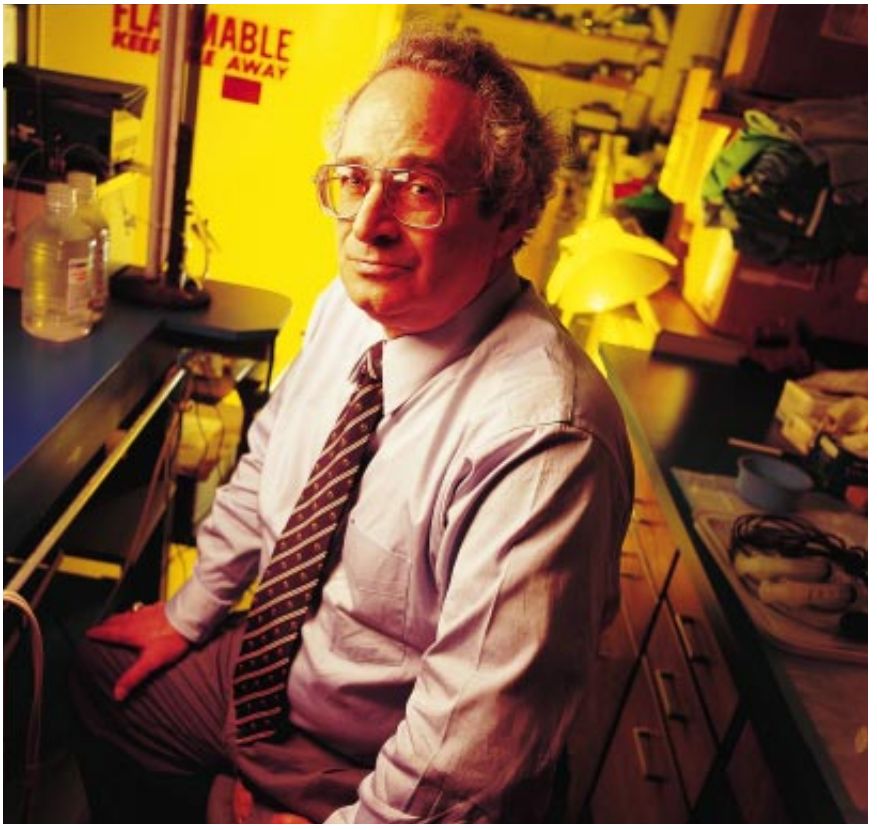
That public health advocates can even dream today of such a possibility—and tobacco companies lose sleep over it—is due in no small part to a bulky Federal Express delivery Glantz received on May 12, 1994, and to what he did with it. Inside the boxes were some 4,000 confidential memos and reports from Brown & Williamson Tobacco and its parent company, BAT Industries, two of the largest cigarette makers in the U.S. and the U.K. The return address read simply "Mr. Butts." A paralegal named Merrell Williams, Jr., it was later reported, had spirited the copies out of a law firm contracted by B&W to review 8.6 million pages of internal communications.

Williams also sent some documents to the *New York Times*, ABC Television, two congressmen and others, but Glantz received the largest set. "My first instinct," he recalls, "was to give them to Richard Daynard," a law professor at Northeastern University who specializes in tobacco litigation. "But then I started looking at them and got sucked in." The *Times* and other media began reporting the most surprising excerpts from the memos—such as the now famous statement made in 1963 by Addison Yeaman, then vice president and general counsel of B&W, that "we are, then, in the business of selling nicotine, an addictive drug effective in the release of stress mechanisms." Meanwhile Glantz and several colleagues at U.C.S.F. began a more thorough and systematic analysis of what the documents revealed about the industry's knowledge of the harmful effects of smoking.

It was an enormous task, and Glantz, Daynard and their colleagues felt the hot breath of B&W's lawyers on their necks. "It was like the Manhattan Project," Daynard reflects. "Stan wanted to finish the bomb before the industry could enjoin the documents. His co-authors felt considerable stress—if they stopped work at 11:30 rather than at midnight, Stan would berate them the next day—but he himself never doubted that what he was doing was critical."

Fortunately, neither did the university, which boldly placed the documents on public display in the library and fought all the way to the California Supreme Court to prevent B&W from getting them back. Publishing the analysis was another matter. Some two dozen publishing houses turned down *The Cigarette Papers*, many out of fear that legal battles with B&W would cost more than the book would earn, Glantz says. In the end, the University of California Press put out the book—and faced no legal resistance whatsoever.

In July 1995 the *Journal of the American Medical Association* ran five long articles by Glantz and his co-workers on the pirated B&W memos. "We knew that we could have been sued for all we were worth," says George D. Lundberg,



TIMOTHY ARCHIBALD

BATTLE-HARDENED TOBACCO FIGHTER
Stanton A. Glantz hopes cigarettes will prove harmful to their manufacturers' health.

the journal's editor. "Yet for the first and only time in the journal's history, the board of trustees and officers of the AMA [American Medical Association] all co-signed as an editorial in that issue."

Glantz's group, the doctors noted, had demonstrated that over the past 30 years the cigarette industry had often known more about the harmful and addictive effects of smoking than the medical community did, and yet it not only concealed its discoveries behind legal technicalities but also lied about them to the public. "As a parent and grandparent, as a physician and as a scientist, this use of the law to harm the public health created in me and many others a sense of outrage," Lundberg says. The AMA's incensed leaders called for the elimination of all tobacco advertising, the regulation of cigarettes as drug-delivery devices, the prohibition of tobacco exports and aggressive legal action to recover medical costs from the tobacco industry. As it turned out, *JAMA* was not sued, and that issue became probably its most famous, as television coverage broadcast the news to 120 million viewers.

"That was a major turning point in the tobacco wars," comments John R. Garrison, director of the American Lung Association. "Stan really started the whole ball moving, and the momentum has continued to build." Indeed, Glantz's five *JAMA* articles had been cited 145 times in major scientific journals by this past February, about three times the typical impact of *JAMA* articles. Another paper Glantz published earlier in 1995 reviewing the link between passive smoking and heart disease has since garnered an extraordinary 129 citations.

Legally, the articles "had a transformative effect on tobacco cases," Daynard says. The *JAMA* issue was introduced as evidence in Minnesota's successful suit against the major tobacco companies, as well as several others, Daynard and Lundberg report.

Then there were the political repercussions. Lundberg confides that "David Kessler [former head of the Food and Drug Administration] knew about this special issue in advance" and timed his assertion of the FDA's authority to regulate tobacco accordingly. Kessler gained President Bill Clinton's support for the move in part because of Glantz's articles, Daynard claims. "Clinton read the whole issue, then called Donna Shalala [secretary of health and human services] and told her to sock it to the bastards," Daynard says.

Such influence comes with consequences. One week after Glantz's articles appeared in *JAMA,* Republican legislators tried to revoke Glantz's grant from the National Cancer Institute (NCI)—the first time Congress had ever singled out an NCI grant for defunding. But with support from Shalala and a group of 29 prominent public health officials, Glantz retained his research money. Then, last year, an industry-sponsored nonprofit group took U.C.S.F. to court, charging Glantz with improper use of tax funds. After six months of legal wrangling, the suit was dismissed.

> "Do we really want to tout cigarette smoke as a drug? It is, of course, but there are dangerous FDA implications to having such conceptualizations go beyond these walls."
>
> —Confidential memo from W. L. Dunn to Dr. H. Wakeham of Philip Morris, 1969, presented at the Minnesota trial

Glantz says that such public wrestling matches with the industry, though exhausting, serve an important purpose. In 1978 "I worked on a campaign to pass a state initiative banning smoking in some public places. The industry beat us," he explains. "But after it was over, we realized that we had inadvertently tricked the industry into running a huge public awareness campaign on the health effects of smoking."

Years later, when a similar initiative passed, it was actually enforced. "Smoking is a social problem," Glantz observes. "Ordinances only work to the extent that they sanctify a change in public attitudes." That is one reason he opposes any congressional bill preempting the 41 states that have sued cigarette makers. "These trials educate the public," he points out.

Indeed, the Minnesota trial uncovered 39,000 more documents showing that the companies knew even more about the dangers of smoking—and knew it even earlier—than the B&W papers had suggested. Faced with such incriminating evidence, the industry was forced to agree in its settlement on May 8 to disband the Council for Tobacco Research (CTR), the research unit that sponsored biased smoking studies for decades, and to turn over all the CTR's research data to the FDA. The companies also ceded 33 million pages of internal documents to public access. And they agreed never again to assert, falsely, that nicotine is not addictive or that smoking has not been proved to cause lethal diseases.

Add to those provisions the roughly $36 billion that the industry has agreed to pay Texas, Florida, Mississippi and Minnesota over the next 25 years, and it is clear, Glantz says, that "Congress is the absolute worst place to deal with this issue." Glantz calculated that the (now defunct) national settlement proposed in June 1997 would have reimbursed society just 10 cents for each dollar in damage caused by tobacco addiction. "So far the states have won basically full reimbursement in their settlements," he says.

Glantz advocates either no new federal law or a very simple one that stiffly increases the tax on cigarettes and slaps steep fines on their manufacturers if smoking rates among children fail to fall. "That strategy is far better than anything that has surfaced so far in Washington," Garrison concurs.

A top tobacco official "once complained that with so many state and local battles, 'it is like being pecked to death by ducks,' " Glantz says. And the flock is growing. A private health insurer joined the Minnesota suit and won a separate settlement of $469 million. "It is a qualitative leap forward," Daynard observes. "The obvious next step is for major employers to sue for damages. The tidal wave has begun."

And if the cigarette companies are hounded into bankruptcy? "The notion that they would go bankrupt voluntarily is silly," Daynard responds. "And besides, bankruptcy doesn't mean that cigarette production and tobacco farming would stop." Nevertheless, Glantz suggests, "If we could be rid of the cigarette companies, that would be good. With a big enough [advertising] campaign, I believe we could eliminate smoking in 10 years." It sounds like a pipe dream. But then again, five years ago so did the suggestion that anyone could sue cigarette manufacturers for 15 billion bucks and win.

—W. Wayt Gibbs in San Francisco

ENVIRONMENTAL POLICY

HOT COOLANTS

An international clampdown is planned on the black market in CFCs and other banned chemicals

Growing evidence of large-scale smuggling in chlorofluorocarbons (CFCs), coolants that deplete the earth's protective ozone layer, has forced the world's rich countries to agree on coordinated action to enforce the Montreal Protocol. That 1987 treaty was intended to reduce and ultimately phase out the chemicals. But although the treaty has driven a 90 percent decrease in CFC production over the past decade, the fall has been slowed by a thriving global black market in the chemicals, fed by factories in Russia, India and China, among other places. "The illegal CFC trade is one of the greatest threats to ozone-layer recovery," says John Passacantando of Ozone Action, an advocacy group in Washington, D.C. Legal loopholes mean that controlling the traffic is turning out to be a challenge.

According to Duncan Brack of the Royal Institute of International Affairs in London, about 15 percent of CFCs in use around the world—tens of thousands of tons—have been smuggled at some point, many shipped via Europe. Russian CFCs, which originally dominated the trade, have given way to Chinese and, most recently, Indian material. As developing nations, China and India qualify to manufacture CFCs in bulk until 2010. Industrial countries can make them only for special purposes—or for export to developing ones.

The group of eight major industrial nations, meeting in the U.K., agreed in May to beef up antismuggling efforts and to coordinate them with developing countries. Customs officers face a challenge because they have to distinguish illegal CFCs from those that are aboveboard. And although alternative coolants are available for most applications, many U.S. automobile air-conditioning systems still need CFCs, which creates a huge market for bootleg supplies. "The next two years will be critical" in the fight against illegal imports, predicts W. Bruce Passfield of the Justice Department, who heads the U.S. enforcement effort.

JAMES VALLETTE *International Trade Information Service*

CONTRABAND CHLOROFLUOROCARBONS made in Mexico by AlliedSignal were seized in Texas by U.S. customs agents.

The U.S., not usually singled out for praise by environmental organizations, is winning plaudits for its anti-CFC-smuggling endeavors. "Operation Cool Breeze," based in Miami, has led to 17 convictions, some resulting in heavy fines and jail terms. But the U.S. price for the most common CFC, freon, is now 10 times higher than the going price in South America. "The incentives to smuggle have increased," Passacantando says.

Thomas A. Watts-Fitzgerald of the Justice Department, who leads Operation Cool Breeze, says convictions have deterred large-scale operators in Miami. Much of the traffic, he points out, is now carried overland from Mexico by "mules," people paid to deliver a package but who may be ignorant of its contents. "This summer we'll see the biggest wave of smuggling yet," predicts James Vallette, an analyst who has worked for Greenpeace and Ozone Action. A favorite trick of some shippers is to create a false paper trail that makes it appear as though a CFC cargo is legitimately just passing through the U.S. Other entrepreneurs taint the chemicals with oil, so that newly manufactured CFCs test like recycled material.

Smuggled environmental hazards may be growing. Petitions to import Chinese halons, fire retardants that are 10 times more harmful to the ozone layer than CFCs, have jumped in the past two years. U.S. authorities have concluded that many of these compounds, which are valuable in military aircraft and oil pipelines for suppressing fires, are falsely labeled as recycled. Investigations are in progress, and the Environmental Protection Agency currently denies China import licenses for the chemicals, Passfield says.

In Europe, efforts to limit the clandestine CFC trade have been languid by comparison—perhaps because governments there do not impose excise taxes on the chemicals. In the U.S., that incentive brings in more than $100 million a year. But the growing political prominence of environmental crime suggests the lackadaisical attitude toward smuggling could be coming to an end. The head of a Frankfurt company was arrested a year ago with 1,000 metric tons of CFCs and halons from China, notes Julian Newman of the Environmental Investigation Agency in London. And Newman's organization, which is a private agency, disclosed last fall that it had created a dummy company to identify Chinese traders willing to supply CFCs wrongly labeled as recycled.

For now, loopholes in both Europe and the U.S. continue to provide opportunities for shady profiteers. But the European Commission has recently decided to enact a sales ban on CFCs, which would simplify policing there—and perhaps lead to some European efforts modeled on Operation Cool Breeze. Nations that were happy a decade ago to limit production of an environmental contaminant are now learning that eliminating its use will take a more serious effort.

—Tim Beardsley in Washington, D.C.

FIELD NOTES

TEMPEST IN A TEACUP

At a navy test facility, a U.S. team prepares to regain sailing's America's Cup

All eyes are on the bright-orange boat and the ripples fanning out from its bow as it pushes through the flat water. This is no toy sailboat bobbing in the pond at Central Park on a clear spring day. It is a $50,000, 7.6-meter-long (25-foot-long) replica of an America's Cup hull—built to within one millimeter of accuracy—inside a cold, dark testing tank at the David Taylor Model Basin in Bethesda, Md., where the U.S. Navy usually tests its destroyers and submarines. Scientists and journalists huddle in silence around a computer that records the water resistance, or drag, on the model as it goes through its motions.

John K. Marshall—president and chief executive officer of the New York Yacht Club/Young America campaign to win back the America's Cup from New Zealand in 2000—looks on with a particular intensity. For this former Olympic sailor, engineer and nine-time America's Cup participant, the race has already started. "The America's Cup is, of course, an athletic and sporting event of the first order," Marshall says, "but it is also an international competition in technology." And that battle—at the drafting board—began several years ago.

Indeed, Marshall and the Young America design team got to work on the year 2000 technology challenge "while we were still weeping over our loss in 1995," Marshall says. To date, they appear to be in the lead. Young America was the first U.S. team to start testing physical models, which Marshall feels is a vital part of the design process. Computer simulations, though vastly improved in recent years, can predict only so much about a hull's performance, he says: "Continual validation and benchmarking are essential." And of the 17 competitors in all, only one other team—the defenders—is as far ahead.

The Maryland setup is impressive. The tank itself, which looks like a station for floating subway cars, is 15.5 meters wide and 6.7 meters deep, making its cross section twice as large as that of the Royal Navy's tank, which New Zealand is currently using. Size offers two clear advantages. Waves created by the moving model inevitably interact with the tank walls, artificially adding drag. But a larger tank minimizes such interference and hence reduces experimental uncertainty. Also, the tank is big enough so that Young America can test models a third the size of an actual yacht—the maximum allowed for model testing under the regatta's rules. (Bigger models mean smaller errors.)

Crossing the tank like a bridge is a 60-ton movable carriage that resembles a giant Erector set. Above it sits a makeshift shack filled with computers; below, a dynamometer leased from the Canadian National Research Council's Institute of Marine Dynamics (IMD). The 330-kilogram instrument, considered the most accurate in the world, hovers over the model hull, measuring the precise forces acting on it as it heels and yaws. In all, Young America plans to test some 20 different models under the supervision of IMD's lead scientist, Robert Pallard. Each model will go through many different runs, each of which is about 200 meters and takes about 15 minutes.

It is slow work, but the outcome will be a fast shape. They hope. Few design breakthroughs have occurred during the past century. "The New York Yacht Club keeps all boats built for members on record, not on paper but as scale models," Marshall observes. "And through the years, very little has changed." He notes one exception: the winglet, an innovation borrowed from aviation by the yacht designers behind the first successful challenge to the America's Cup by Australia in 1983. Placing the tiny fins on a keel increases the apparent draft of this structure and reduces lift-induced drag in much the same way as it does for airplane wings.

The day they invited the press to observe, Young America had nothing so radical on display. They were retesting an old hull design from 1995 for comparison's sake. For the next two years, a great deal of work remains—including wind-tunnel tests, further computer analyses and full-scale two-boat trials. But Marshall says the design team is hungry: "They're just as competitive as the sailors are. This is unlike normal research. There is a final deadline, and then you win or lose." —*Kristin Leutwyler*

DAN NERNEY

AMERICA'S CUP MODEL, measuring 7.6 meters long, sails through still waters in the testing tank at the David Taylor Model Basin in Bethesda, Md.

COMPUTER SCIENCE

GETTING REAL?

Synthetic emotions could make computers nicer

When today's users respond emotionally to a computer, they typically call it unprintable names, perhaps hold down all the keys and maybe contemplate throwing it out a window. But such unpleas-

antness could be a thing of the past if projects at Stanford University and at the Massachusetts Institute of Technology Media Laboratory bear fruit. Researchers are studying how to make people feel happy about the relationship between man and machine—and how to make computers more soothing when they detect frustration. The approach has started to attract serious attention from computer and software designers—as well as criticism that it is misconceived and ethically questionable.

The new interest in how people feel about computers, as opposed to simply how they use them, has been driven in large part by Byron Reeves and Clifford I. Nass of Stanford, who have long studied how people respond to what Nass is happy to call a computer's personality. Reeves and Nass have shown that even computer-literate people respond emotionally to machine-generated messages they see on a screen, as well as to apparently irrelevant details, such as the quality of a synthesized voice. Their responses are much like those that would be elicited by a real person.

An unhelpful error message, for example, elicits the same signs of irritation as an impolite comment from an unlikable person. Such involuntary and largely unconscious responses have potentially important consequences. Users engage in gender stereotyping of machines, for example, being more likely to rate a "macho" voice as authoritative than a female one. Users also enjoyed interacting better with a screen character of their own ethnicity than with a character portrayed differently. Because so many people today spend more time interacting with a computer than with other people, hardware and software designers have a keen interest in such issues—as the imposing list of corporate sponsors supporting Reeves and Nass's work testifies.

At M.I.T., Rosalind W. Picard and her students are trying to take the next step—giving computers the power to sense their users' emotional state. Picard is convinced that computers will need the ability to recognize and express emotions in order to be "genuinely intelligent." Psychologists, she points out, have established that emotions greatly affect how people make decisions in the real world. So a computer that recognized and responded to emotions might be a better collaborator than today's insensitive, pigheaded machines.

Detecting emotions is difficult for a machine, especially when someone is trying to conceal them. But Picard says she has at least one system "that definitely looks useful." The apparatus detects frowning in volunteers who are asked to perform a simple computer-based task and are then frustrated by a simulated glitch. The setup monitors the frown muscles by means of a sensor attached to special eyeglasses. Other studies she has conducted with Raul Fernandez have achieved "better than random" detection of frustration responses in 21 out of 24 subjects by monitoring skin conductance and blood flow in a fingertip. Picard's work, too, has attracted industry interest.

Jonathan T. Klein, also at M.I.T., is building on Picard's results to try to make friendlier digital helpmates. Klein is testing strategies for calming down frustrated users. Klein's system may, for example, solicit a dialogue or comment on the user's annoyance sympathetically without judgment. (These strategies were inferred from observations of skilled human listeners, according to Klein.) Nass suggests that computers might one day detect when a user is feeling down—and try to adapt by livening things up.

But the notion that computers might respond emotionally—or what psychologists call "affectively"—itself causes frustration in Ben Shneiderman, a computer-interface guru at the University of Maryland. Shneiderman says people want computers to be "predictable, controllable and comprehensible"—not adaptive, autonomous and intelligent. Shneiderman likens an effective computer interface to a good tool, which should do what it is instructed to do and nothing else. He cites the failed "Postal Buddy" stamp-selling robot, the extinct talking automobile and Microsoft's defunct "Bob" computer character as evidence of the futility of making machines like people. And there are significant ethical questions about allowing people to be manipulated by machines in ways they are not aware of, Shneiderman contends.

Picard, though, says her studies address only emotions that people do not try to hide. And Nass, who acknowledges Shneiderman's ethical concerns, notes that Microsoft Bob's digital progeny are alive and well—as the humanoid assistants, such as "Einstein" and "Clip-It," that dispense advice in Office 97's built-in help system. Machines are already becoming more polite, Nass states, and more friendliness is on the way. So if you are reading this on-line, thank you for visiting the *Scientific American* Web site. We hope you'll come back another day.

—Tim Beardsley in Washington, D.C.

EARTH SCIENCE

TECTONICS IN A SANDBOX

Researchers model the earth's motions at small scale

How do rivers cut their banks? For decades, hydrologists have used tanks filled with water and fine sand to find out. In many cases, experimenting with such small-scale physical models has proved superior to using computer codes, which often do not mimic nature nearly as well. That is why a group of researchers at the University of Minnesota's St. Anthony Falls Laboratory decided to build what may be the most elaborate modeling tank of this kind yet, one that allows the scientists to control more than the usual parameters of sediment type, water flow and level. The room-size device now under construction will also be able to imitate the way the earth's crust gradually shifts up and down. Tests of the concept have already produced scale models of continental shelves that, when opened up, look so much like cross sections of the real thing that petroleum geologists may learn something new about these vast reservoirs of oil and gas.

Because the apparatus can simulate the evolution of subsurface geology, its builders have dubbed it "Jurassic tank." It was conceived by two of the university's faculty: Christopher Paola, a geologist, and Gary Parker, a civil engineer. When they first thought of building a tank that could model tectonic subsidence and uplift, they designed one with a hinged but otherwise rigid floor that could be jacked up and down in various ways. But after consulting with an engineering firm that fabricates earthquake simulators and similar equipment, they realized that their initial concept was essentially unworkable, even with their $500,000 budget.

LAYNE KENNEDY

***HONEYCOMB OF FUNNELS** will hold enough gravel to support a rubber floor for the giant modeling tank being built by University of Minnesota researchers.*

Paola and Parker began to despair, but their hopes rebounded when James P. Mullin, an engineer in the lab, showed them something resembling an ant farm. Mullin demonstrated that the best way to make a flexible floor for a large tank was to use granular material that could be withdrawn from the bottom. "Like the agricultural feeders you see around here," Mullin notes.

After some experimenting, the team built 10 steep-sided funnels with hexagonal rims to test the idea. The researchers clustered the hexagons together in beehive fashion within a small tank, filled the cones with pea-size gravel and laid more gravel and a rubber mat on top. Pushing gravel out a little at a time from the bottom of the funnels (with a water jet) allowed them to move the rubber floor of the tank downward with a precisely controlled motion. The prototype worked so well that they then used it to simulate coastal deposition, introducing at one end fine sand and crushed coal so that the resulting light and dark bands would delineate discrete sedimentary beds.

As the flow of water transported these sediments across the model terrain and into the diminutive ocean pooled at the far side, the scientists adjusted the water level and withdrew gravel from the underside. In this way, they could simulate changing sea level and tectonic subsidence. The experiment succeeded, but it created another technical problem: how to dissect the slab of sediment resting on the rubber mat to reveal the "geologic" structure hidden within it.

Again Mullin came up with the solution. "He made essentially a big microtome," quips Christopher R. Ellis, a researcher on the Jurassic tank team. Like the microtomes used to slice biological specimens for microscopic analysis, the apparatus Mullin built shaves thin layers from the side of the model deposit. After each slice, photographs are taken, and the results are recorded on a computer, which can later render cross sections at any orientation.

The six- by 11-meter Jurassic tank, with its full set of 432 hexagonal funnels, should be complete by the fall. Although the tank will eventually be used to model other geologic settings, initial experiments will simulate sedimentation along continental margins. Once the slices are translated to a form that resembles the seismic cross sections used by oil-exploration geologists, Parker boasts, "we can produce images that will fool them." *—David Schneider*

THE INTERNET

LOST IN CYBERSPACE

Scientists look for a better way to search the Web

As any World Wide Web surfer knows, finding information over the Internet can be painfully time-consuming. Search engines such as Yahoo!, AltaVista and Infoseek help, but an improperly honed query can easily result in digital diarrhea—tens of thousands of Web pages that are irrelevant. A new technique that analyzes how documents posted on the Internet are linked to one another could provide relief. Developed by researchers from IBM, Cornell University and the University of California at Berkeley, the method finds two types of Web sites for a particular desired subject: "authorities" (pages that are cited by many other documents on that topic) and "hubs" (sites that link to many of those authorities).

The system, dubbed automatic resource compiler (ARC), first performs an ordinary Boolean text-based search (for example, locating documents that contain the words "diamond" and "mineral" but not "baseball") using an engine such as AltaVista. After generating a quick list of about 200 pages, ARC then expands that set to include documents linked to and from those 200 pages. The step is repeated to obtain a collection of up to 3,000 locations. ARC then analyzes the interconnections between those documents, essentially giving higher authority scores to those pages that are frequently cited, with the assumption that such documents are more useful (just as scientific papers that are referenced by many other articles are deemed most important). Also, hubs are given high marks for having linked to those authorities.

One feature of ARC is that it leads to the natural separation of Web sites—both authorities and hubs—into communities. A search for information on abortion, for instance, will result in two sets of sites, pro-life and pro-choice, because documents from one group are more likely to link to one another than to pages from the other community.

Though clever, ARC is not perfect. "It is possible for a query to go awry," admits Jon M. Kleinberg, an assistant professor at Cornell's computer science department and developer of the algorithm at the heart of ARC. For one, searches on a specific topic such as "Steffi Graf" can result in Web pages on the general subject of tennis without any mention of the German star athlete.

Consequently, some researchers feel that future tools will need to offer a variety of techniques, depending on the type of information desired. "There's a danger in the one-size-fits-all approach," warns Louis Monier, technical director for AltaVista. Other methods being investigated include morphological and linguistic analyses that might, for example, aid in finding a person's home page (as opposed to articles written about that person) by exploiting certain distinct characteristics. Specifically, home pages usually contain photographs of the person, and the language used tends to be less sophisticated. Recently Infoseek implemented a new proprietary search technology that takes into account about a dozen factors, including the number of times a page is cited as well as the date when the document was last modified.

Whatever the approach, one thing is for sure: the need for the next generation of search tools is becoming critical, asserts Prabhakar Raghavan, one of the IBM researchers who helped to develop ARC. (When and how ARC and others will be introduced commercially, however, is unclear.) "The amount of stuff on the Web is growing exponentially," he says, "but the amount we can digest is not. So the information you do retrieve must be exemplary."

—Alden M. Hayashi

NUCLEAR ENERGY

A New Twist in Fusion

For the past 30 years, fusion energy researchers have been forecasting that commercially viable reactors are just a decade away. One reason that great day keeps receding into the distant future is that holding a gas of charged deuterium or tritium (isotopes of hydrogen) steady while its atoms fuse into helium is harder than almost anyone expected. The most popular reactor designs, called tokamaks, try to confine the hydrogen plasma inside shifting magnetic fields generated both by currents inside the plasma itself and by giant external magnets. If there are leaks in this magnetic bottle, the plasma hits the reactor's inner walls and loses its energy.

NATIONAL INSTITUTE OF FUSION SCIENCE

Physicists have known for many years that another kind of magnetic fusion device, called a stellarator, might get around this problem. In a stellarator, intertwined spiral magnets (*photograph*) and several ring magnets do all the work of confining the plasma inside its doughnut-shaped chamber. Because, unlike tokamaks, there is no need to pass electric current through the plasma, the arrangement is inherently stable. But this design has never been tested at large scales.

That is all changing. On March 31, Japan's National Institute of Fusion Science injected the first high-energy plasma into a new stellarator that is more than 10 times the size of any built before. Eight years in construction, the reactor sports eight huge superconducting magnetic coils. It took a full month of refrigeration to cool them down to the near absolute zero temperature needed to eliminate their electrical resistance. When it ramps up to full power in a few months, the Large Helical Device, as it is called, should be able to perform about as well as the renowned Tokamak Fusion Test Reactor (TFTR) at Princeton University. This stellarator will never create more energy than it burns, but bigger, better ones might—in about 10 years, give or take 30.

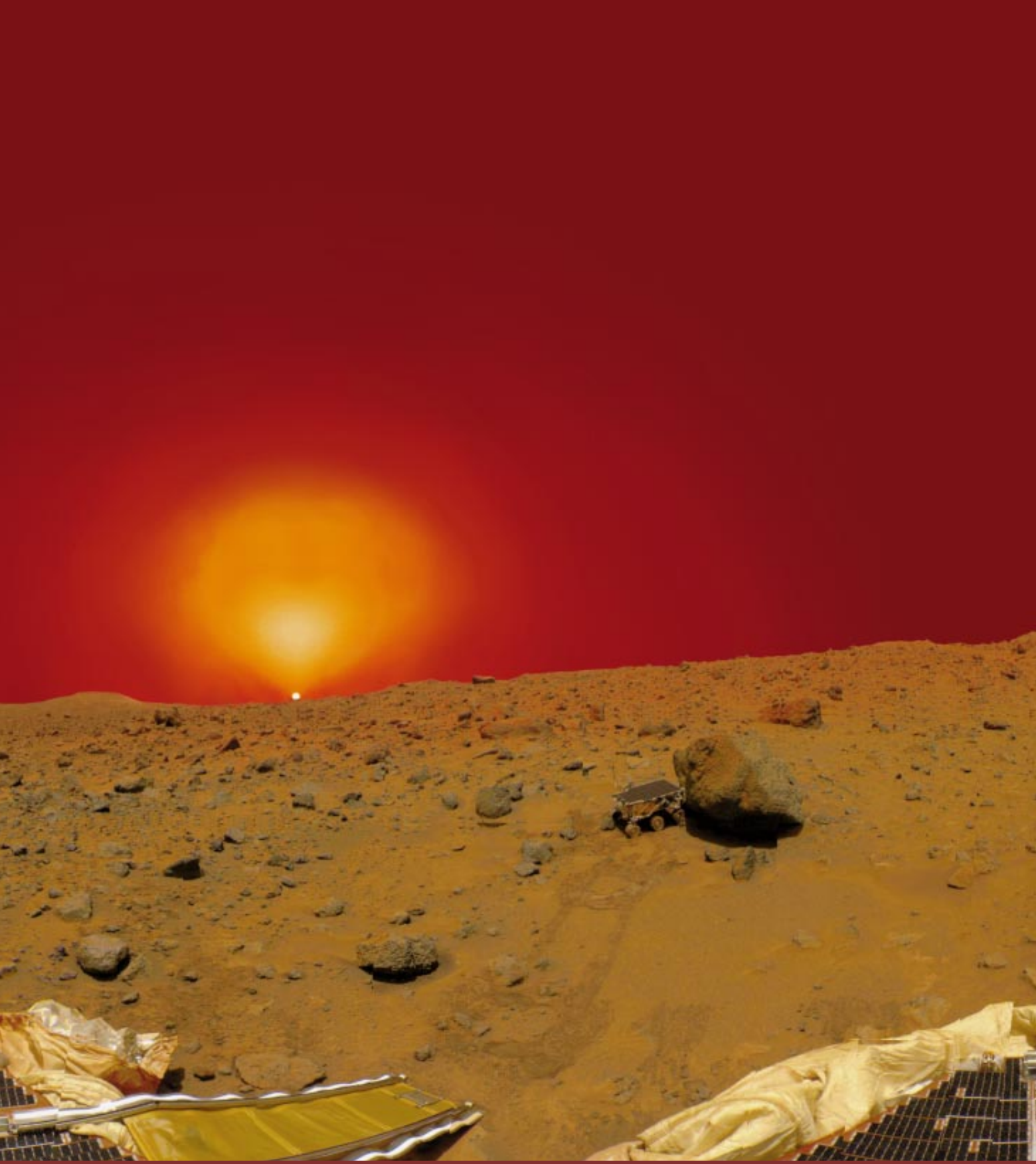
—W. Wayt Gibbs in San Francisco

The Mars Pathfinder Mission

Last summer the first ever Mars rover found in situ evidence that the Red Planet may once have been hospitable to life

by Matthew P. Golombek

Rocks, rocks, look at those rocks," I exclaimed to everyone in the Mars Pathfinder control room at about 4:30 P.M. on July 4, 1997. The Pathfinder lander was sending back its first images of the surface of Mars, and everyone was focused on the television screens. We had gone to Mars to look at rocks, but no one knew for sure whether we would find any, because the landing site had been selected using orbital images with a resolution of roughly a kilometer. Pathfinder could have landed on a flat, rock-free plain. The first radio downlink indicated that the lander was nearly horizontal, which was worrisome for those of us interested in rocks, as most expected that a rocky surface would result in a tilted lander. The very first images were of the lander so that we could ascertain its condition, and it was not



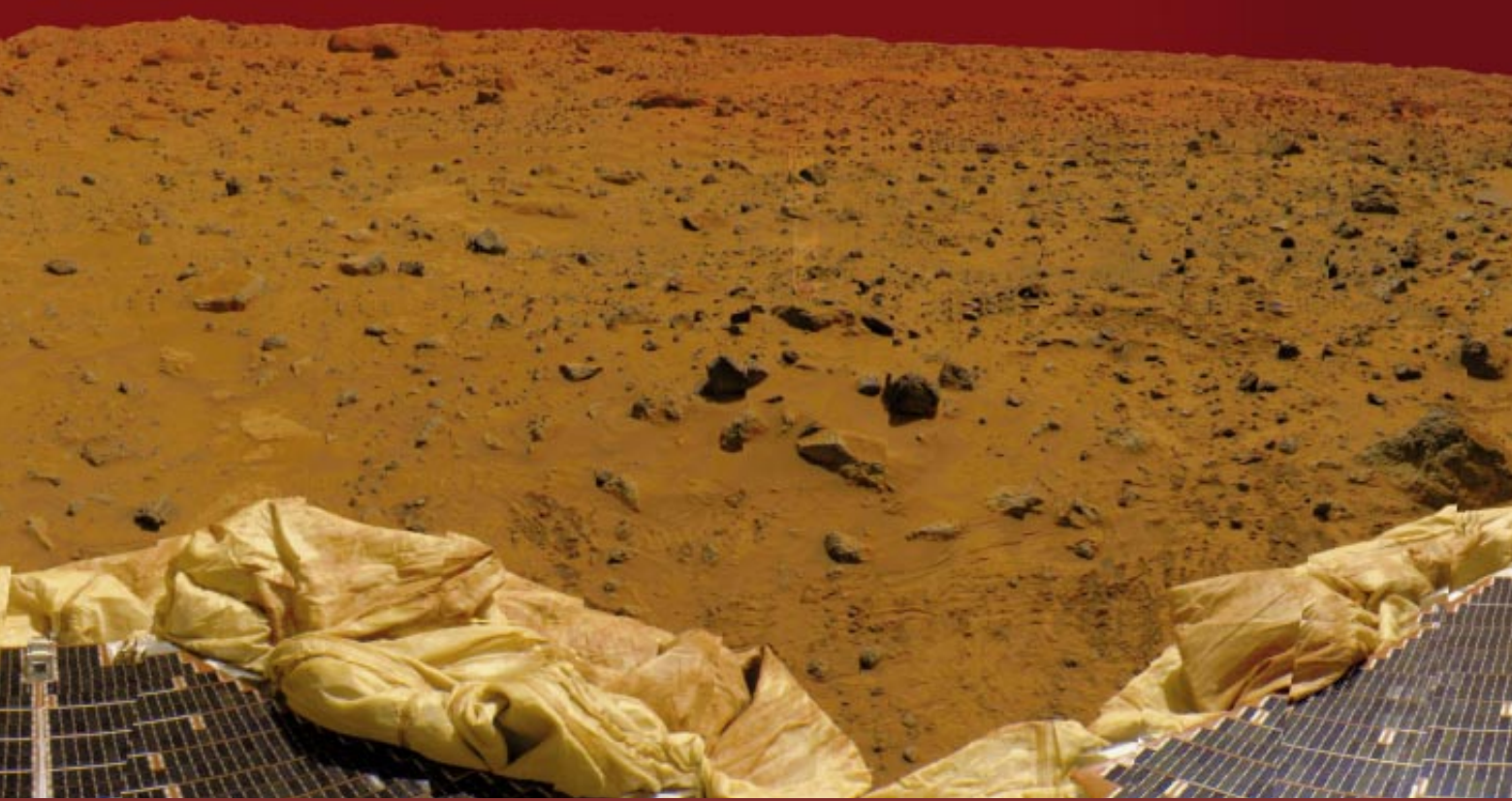
TWILIGHT AT ARES VALLIS,

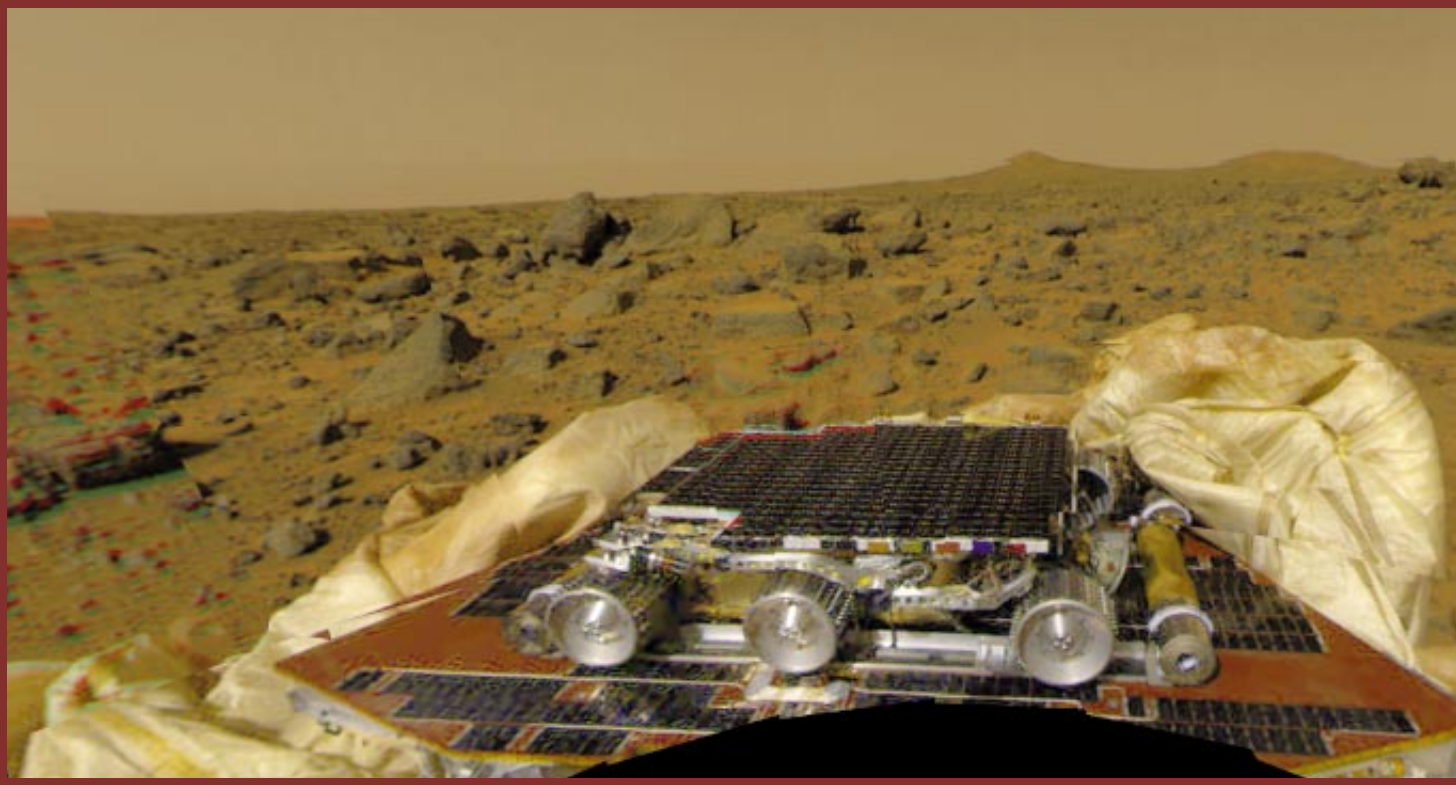
Pathfinder's landing site, is evoked in this 360-degree panorama, a composite of a true sunset (*inset at right*) and other images. The rover is analyzing the rock Yogi to the right of the lander's rear ramp. Farther right are whitish-pink patches on the ground known as Scooby Doo (*closer to lander*) and Baker's Bench. The rover tried to scratch the surface of Scooby Doo but could not, indicating that the soil in these patches is cemented together. The much studied Rock Garden appears left of center. Flat Top, the flat rock in front of the garden, is covered with dust, but steep faces on other large rocks are clean; the rover analyzed all of them. (In this simulation, parts of the sky and terrain were computer-adjusted to complete the scene. During a real sunset, shadows would of course be longer and the ground would appear darker.) —*The Editors*

until a few tense minutes later that the first pictures of the surface showed a rocky plain—exactly as we had hoped and planned for [*see top illustration on page 43*].

Why did we want rocks? Every rock carries the history of its formation locked in its minerals, so we hoped the rocks would tell us about the early Martian environment. The two-part Pathfinder payload, consisting of a main lander with a multispectral camera and a mobile rover with a chemical analyzer, was suited to looking at rocks. Although it could not identify the minerals directly—its analyzer could measure only their constituent chemical elements—our plan was to identify them indirectly based on the elemental composition and the shapes, textures and colors of the rocks. By landing Pathfinder at the mouth of a giant channel where a huge volume of water once flowed briefly, we sought rocks that had washed down from the ancient, heavily cratered highlands. Such rocks could offer clues to the early climate of Mars and to whether conditions were once conducive to the development of life [*see top illustration on page 46*].

U.S. GEOLOGICAL SURVEY AND NASA/JET PROPULSION LABORATORY; SUNSET SIMULATION BY LAURIE GRACE





FIRST IMAGES
from Mars Pathfinder were assembled into this panorama of dark rocks, bright-red dust and a butterscotch sky. Many rocks, particularly in the Rock Garden (*center*), are inclined and stacked—a sign that they were deposited by fast-moving water. About a kilometer behind the garden on the west-southwest horizon are the Twin Peaks, whose prominence identified the landing site on Viking orbiter images. The day after touching down, the lander pulled back the air bag and unfurled two ramps; the rover trundled down the rear ramp onto the surface. (The small green and red streaks are artifacts of data compression.)

NASA/JET PROPULSION LABORATORY

SANDBLASTED ROCK
named Moe resembles terrestrial rocks known as ventifacts. Their fluted texture develops when sand-size particles hop along the surface in the wind and erode rocks in their path. On Earth, such particles are typically produced when water breaks down rocks. Moe's grooves all point to the southwest, which is roughly the same orientation as wind streaks seen elsewhere at the site.

NASA/JET PROPULSION LABORATORY

The most important requirement for life on Earth (the only kind we know) is liquid water. Under present conditions on Mars, liquid water is unstable: because the temperature and pressure are so low, water is stable only as ice or vapor; liquid would survive for just a brief time before freezing or evaporating. Yet Viking images taken two decades ago show drainage channels and evidence for lakes in the highlands. These features hint at a warmer and wetter past on Mars in which water could persist on the surface [see "Global Climatic Change on Mars," by Jeffrey S. Kargel and Robert G. Strom; SCIENTIFIC AMERICAN, November 1996]. To be sure, other explanations have also been suggested, such as sapping processes driven by geothermal heating in an otherwise frigid and dry environment. One of Pathfinder's scientific goals was to look for evidence of a formerly warm, wet Mars.

The possible lakebeds are found in terrain that, judging from its density of impact craters, is roughly the same age as the oldest rocks on Earth,which show clear evidence for life 3.9 billion to 3.6 billion years ago. If life was able to develop on Earth at this time, why not on Mars, too, if the conditions were similar? This is what makes studying Mars so compelling. By exploring our neighboring planet, we can seek answers to some of the most important questions in science: Are we alone in the universe? Will life arise anywhere that liquid water is stable, or does the formation of life require something else as well? And if life did develop on Mars, what happened to it? If life did not develop, why not?

Pathfinding

Pathfinder was a Discovery-class mission—one of the National Aeronautics and Space Administration's "faster, cheaper, better" spacecraft—to demonstrate a low-cost means of landing a small payload and mobile vehicle on Mars. It was developed, launched and operated under a fixed budget comparable to that of a major motion picture (between $200 million and $300 million), which is a mere fraction of the budget typically allocated for space missions. Built and launched in a short time (three and a half years), Pathfinder included three science instruments: the Imager for Mars Pathfinder, the Alpha Proton X-Ray Spectrometer and the Atmospheric Structure Instrument/Meteorology Package. The rover itself also acted as an instrument; it was used to conduct 10 technology experiments, which studied the abrasion of metal films on a rover wheel and the adherence of dust to a solar cell as well as other ways the equipment reacted to its surroundings.

In comparison, the Viking mission, which included two orbiter-lander pairs, was carried out more than 20 years ago at roughly 20 times the cost. Viking was very successful, returning more than 57,000 images that scientists have been studying ever since. The landers carried sophisticated experiments that tested for organisms at two locations; they found none.

The hardest part of Pathfinder's mission was the five minutes during which the spacecraft went from the relative security of interplanetary cruising to the stress of atmospheric entry, descent and landing [*see illustration on page 47*]. In that short time, more than 50 critical events had to be triggered at exactly the right times for the spacecraft to land safely. About 30 minutes before entry, the backpack-style cruise stage separated from the rest of the lander. At 130 kilometers above the surface, the spacecraft entered the atmosphere behind a protective aeroshell. A parachute unfurled 134 seconds before landing, and then the aeroshell was jettisoned. During descent, the lander was lowered beneath its back cover on a 20-meter-long bridle, or tether.

As Pathfinder approached the surface, its radar altimeter triggered the firing of three small rockets to slow it down further. Giant air bags inflated around each face of the tetrahedral lander, the bridle was cut and the lander bounced onto the Martian surface at 50 kilometers per hour. Accelerometer measurements indicate that the air-bag-enshrouded lander bounced at least 15 times without losing air-bag pressure. After rolling at last to a stop, the lander deflated the air bags and opened to begin surface operations.

Although demonstrating this novel landing sequence was actually Pathfinder's primary goal, the rest of the mission also met or exceeded expectations. The lander lasted three times longer than its minimum design criteria, the rover 12 times longer. The mission returned 2.3 billion bits of new data from Mars, including more than

SAND DUNES provide circumstantial evidence for a watery past. These dunes, which lay in the trough behind the Rock Garden, are thought to have formed when windblown sand hopped up the gentle slope to the dune crest and cascaded down the steep side (which faces away from the rover in this image). Larger dunes have been observed from orbit, but none in the Pathfinder site. The discovery of these smaller dunes suggests that sand is more common on Mars than scientists had thought. The formation of sand on Earth is principally accomplished by moving water.

16,500 lander and 550 rover images and roughly 8.5 million individual temperature, pressure and wind measurements. The rover traversed a total of 100 meters in 230 commanded movements, thereby exploring more than 200 square meters of the surface. It obtained 16 measurements of rock and soil chemistry, performed soil-mechanics experiments and successfully completed the numerous technology experiments. The mission also captured the imagination of the public, garnering front-page headlines for a week, and became the largest Internet event in history, with a total of about 566 million hits for the first month of the mission—47 million on July 8 alone.

Flood Stage

The mosaic of the landscape constructed from the first images revealed a rocky plain (about 20 percent of which was covered by rocks) that appears to have been deposited and shaped by catastrophic floods [*see top illustration on page 43*]. This was what we had predicted based on remote-sensing data and the location of the landing site (19.13 degrees north, 33.22 degrees west), which is downstream from the mouth of Ares Vallis in the low area known as Chryse Planitia. In Viking orbiter images, the area appears analogous to the Channeled Scabland in eastern and central Washington State. This analogy suggests that Ares Vallis formed when roughly the same volume of water as in the Great Lakes (hundreds of cubic kilometers) was catastrophically released, carving the observed channel in a few weeks. The density of impact craters in the region indicates it formed at an intermediate time in Mars's history, somewhere between 1.8 billion and 3.5 billion years ago.

The Pathfinder images support this interpretation. They show semirounded pebbles, cobbles and boulders similar to those deposited by terrestrial catastrophic floods. Rocks in what we dubbed the Rock Garden, a collection of rocks to the southwest of the lander, with the names Shark, Half Dome and Moe, are inclined and stacked, as if deposited by rapidly flowing water. Large rocks in the images (0.5 meter or larger) are flat-topped and often perched, also consistent with deposition by a flood. Twin Peaks, a pair of hills on the southwest horizon, are streamlined. Viking images suggest that the lander is on the flank of a broad, gentle ridge trending northeast from Twin Peaks; this ridge may be a debris tail deposited in the wake of the peaks. Small channels throughout the scene resemble those in the Channeled Scabland, where drainage in the last stage of the flood preferentially removed fine-grained materials.

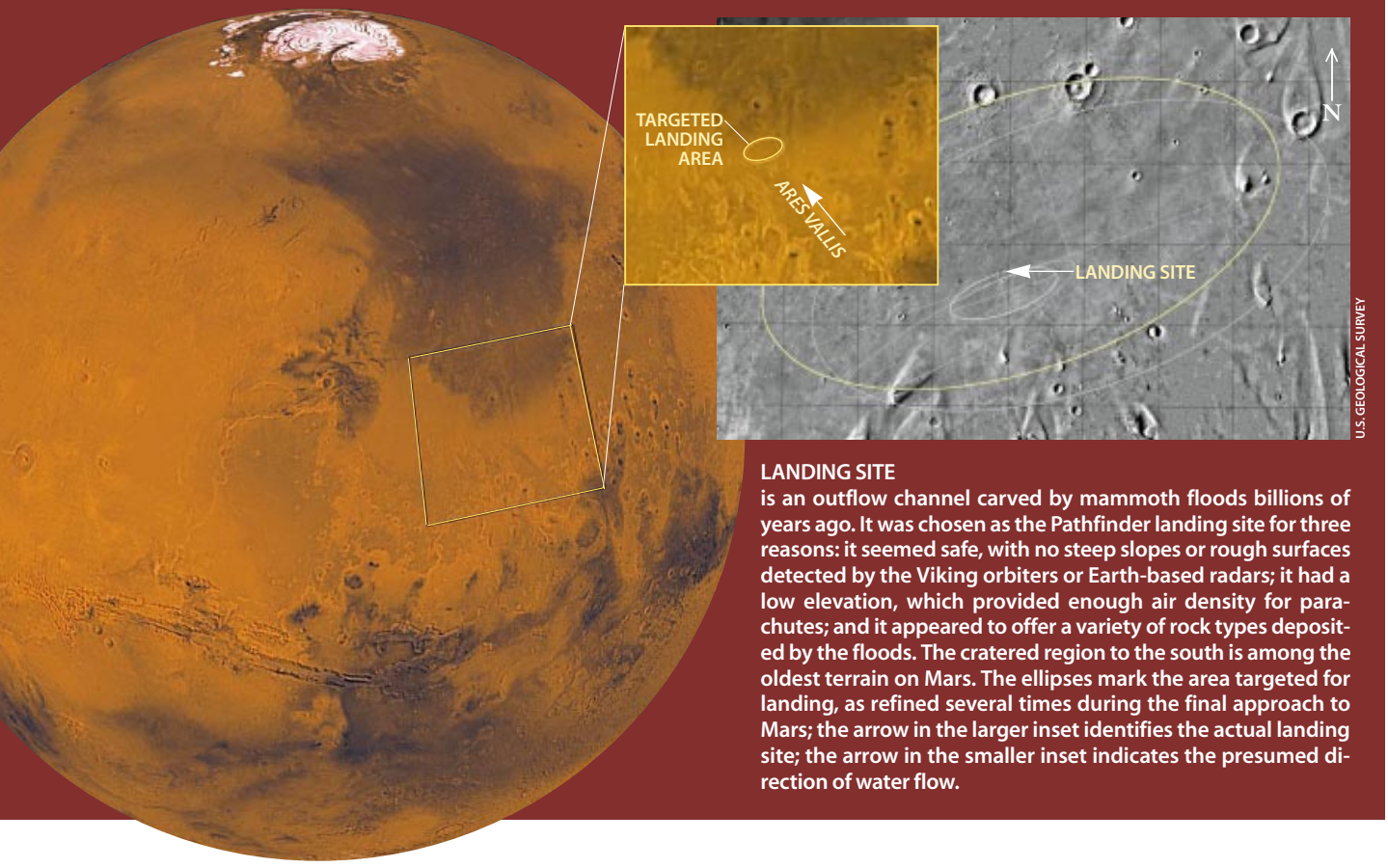
The rocks in the scene are dark gray and covered with various amounts of yellowish-brown dust. This dust appears to be the same as that seen in the atmosphere, which, as imaging in different filters and locations in the sky suggests, is very fine grained (a micron in size). The dust also collected in wind streaks behind rocks.

Some of the rocks have been fluted and grooved, presumably by sand-size particles (less than one millimeter) that hopped along the surface in the wind [*see bottom illustration on page 43*]. The rover's camera also saw sand dunes in the trough behind the Rock Garden [*see illustration below*]. Dirt covers the lower few centimeters of some rocks, suggesting that they have been exhumed by wind. Despite these signs of slow erosion by the wind, the rocks and surface appear to have changed little since they were deposited by the flood.

Sedimentary Rocks on Mars?

The Alpha Proton X-Ray Spectrometer on the rover measured the compositions of eight rocks. The silicon content of some of the rocks is much higher than that of the Martian meteorites, our only other samples of Mars. The Martian meteorites are all mafic igneous rocks, volcanic rocks that are relatively low in silicon and high in iron and magnesium. Such rocks form when the upper mantle of a planet melts. The melt rises up through the crust and solidifies at or near the surface. These types of rocks, referred to as basalts, are the most common rock on Earth and have also been found on the moon. Based on the composition of the Martian meteorites and the presence of plains and mountains that look like features produced by basaltic volcanism on Earth, geologists expected to find basalts on Mars.

NASA/JET PROPULSION LABORATORY



LANDING SITE
is an outflow channel carved by mammoth floods billions of years ago. It was chosen as the Pathfinder landing site for three reasons: it seemed safe, with no steep slopes or rough surfaces detected by the Viking orbiters or Earth-based radars; it had a low elevation, which provided enough air density for parachutes; and it appeared to offer a variety of rock types deposited by the floods. The cratered region to the south is among the oldest terrain on Mars. The ellipses mark the area targeted for landing, as refined several times during the final approach to Mars; the arrow in the larger inset identifies the actual landing site; the arrow in the smaller inset indicates the presumed direction of water flow.

The rocks analyzed by Pathfinder, however, are not basalts. If they are volcanic, as suggested by their vesicular surface texture, presumably formed when gases trapped during cooling left small holes in the rock, their silicon content classifies them as andesites. Andesites form when the basaltic melt from the mantle intrudes deep within the crust. Crystals rich in iron and magnesium form and sink back down, leaving a more silicon-rich melt that erupts onto the surface. The andesites were a great surprise, but because we do not know where these rocks came from on the Martian surface, we do not know the full implications of this discovery. If the andesites are representative of the highlands, they suggest that ancient crust on Mars is similar in composition to continental crust on Earth. This similarity would be difficult to reconcile with the very different geologic histories of the two planets. Alternatively, the rocks could represent a minor proportion of high-silicon rocks from a predominately basaltic plain.

Intriguingly, not all the rocks appear to be volcanic. Some have layers like those in terrestrial sedimentary rocks, which form by deposition of smaller fragments of rocks in water. Indeed, rover images show many rounded pebbles and cobbles on the ground. In addition, some larger rocks have what look like embedded pebbles [*see illustration at left*] and shiny indentations, where it looks as though rounded pebbles that were pressed into the rock during its formation have fallen out, leaving holes. These rocks may be conglomerates formed by flowing liquid water. The water would have rounded the pebbles and deposited them in a sand, silt and clay matrix; the matrix was subsequently compressed, forming a rock, and carried to its present location by the flood. Because conglomerates require a long time to form, if these Martian rocks are conglomerates they strongly suggest that liquid water was once stable and that the climate was therefore warmer and wetter than at present.

POSSIBLE CONGLOMERATE ROCK
may be the best proof yet that Mars was once warm and wet. The rock, known as Ender, has pits and pebbles. It and other rocks could be conglomerates, which require flowing water to form. The lander is visible in the background; the lattice mast holds the lander camera, and the mast on the right holds the meteorological sensors.

Soils at the landing site vary from the bright-red dust to darker-red and darker-gray material. Overall, the soils are lower in silicon than the rocks and richer in sulfur, iron and magnesium. Soil compositions are generally the same as those measured at the Viking sites, which are on opposite hemispheres (Viking 1 is 800 kilometers west of Pathfinder; Viking 2 is thousands of kilometers away on the opposite, eastern side of the northern hemisphere). Thus, this soil may be a globally deposited unit. The similarity in compositions among the soils implies that their differences in color may be the result of slight variations in iron mineralogy or in particle size and shape [*see top right illustration on next page*].

A bright-red or pink material also covered part of the site. Similar to the soils in composition, it seems to be indurated or cemented because it was not damaged by scraping with the rover wheels.

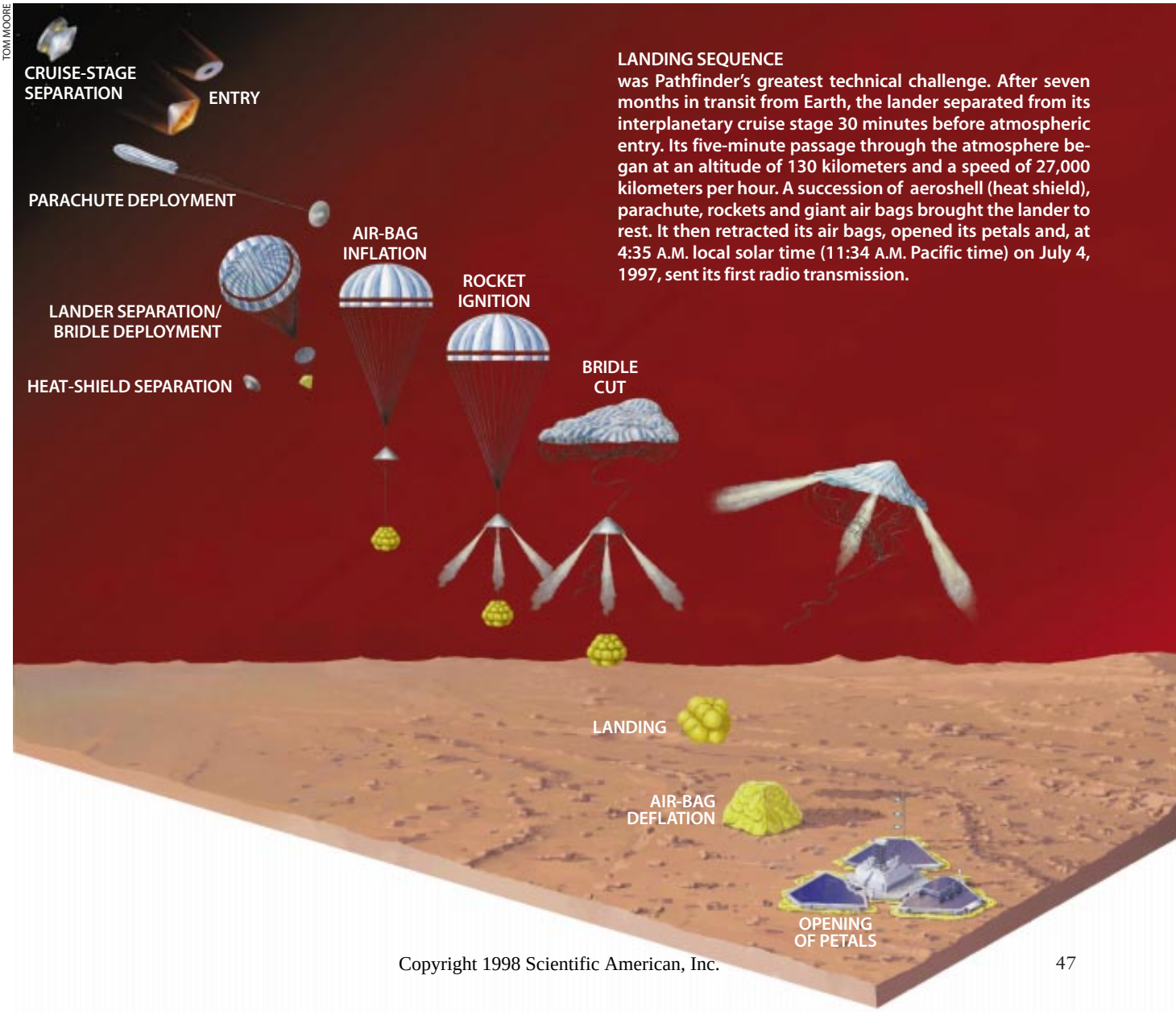
Pathfinder also investigated the dust in the atmosphere of Mars by observing its deposition on a series of magnetic targets on the spacecraft. The dust, it turned out, is highly magnetic. It may consist of small silicate (perhaps clay) particles, with some stain or cement of a highly magnetic mineral known as maghemite. This finding, too, is consistent with a watery past. The iron may have dissolved out of crustal materials in water, and the maghemite may be a freeze-dried precipitate.

The sky on Mars had the same butterscotch color as it did when imaged by the Viking landers. Fine-grained dust in the atmosphere would explain this color. Hubble Space Telescope images had suggested a very clear atmosphere; scientists thought it might even appear blue from the surface. But Pathfinder found otherwise, suggesting either that the atmosphere always has some dust in it from local dust storms or that the atmospheric opacity varies appreciably over a short time. The inferred dust-particle size (roughly a micron) and shape and the amount of water vapor (equivalent to a pitiful hundredth of a millimeter of rainfall) in the atmosphere are also consistent with measurements made by Viking. Even if Mars was once lush, it is now drier and dustier than any desert on Earth.

Freezing Air

The meteorological sensors gave further information about the atmosphere. They found patterns of diurnal and longer-term pressure and temperature fluctuations. The temperature reached its maximum of 263 kelvins

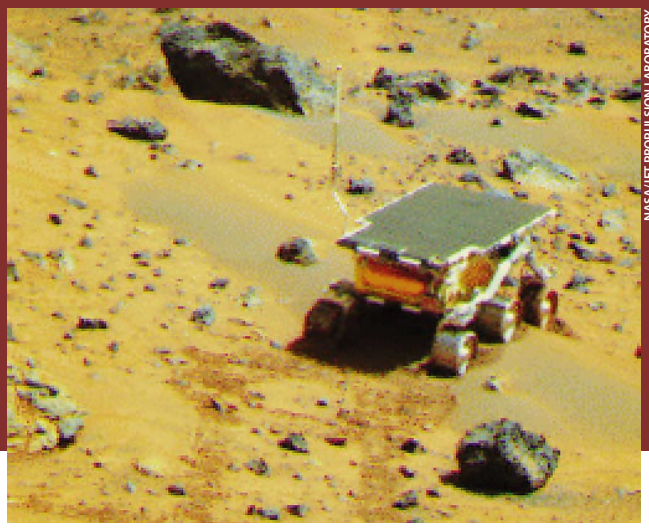
TOM MOORE



LANDING SEQUENCE
was Pathfinder's greatest technical challenge. After seven months in transit from Earth, the lander separated from its interplanetary cruise stage 30 minutes before atmospheric entry. Its five-minute passage through the atmosphere began at an altitude of 130 kilometers and a speed of 27,000 kilometers per hour. A succession of aeroshell (heat shield), parachute, rockets and giant air bags brought the lander to rest. It then retracted its air bags, opened its petals and, at 4:35 A.M. local solar time (11:34 A.M. Pacific time) on July 4, 1997, sent its first radio transmission.

WISPY, BLUE CLOUDS
in the dawn sky, shown in this color-enhanced image taken on sol 39 (the 39th Martian day after landing), probably consist of water ice. During the night, water vapor froze around fine-grained dust particles; after sunrise, the ice evaporated. The total amount of water vapor in the present-day Martian atmosphere is paltry; if it all rained out, it would cover the surface to a depth of a hundredth of a millimeter. The basic appearance of the atmosphere is similar to what the Viking landers saw more than 20 years ago.

MULTICOLORED SOILS
were exposed by the rover's wheels. The rover straddles Mermaid Dune, a pile of material covered by dark, sand-size granules. Its wheel tracks also reveal bright-red dust and darker-red soil (*bottom left*). Scientists were able to deduce the properties of surface materials by studying the effect that the wheels had on them.



NASA/JET PROPULSION LABORATORY

(–10 degrees Celsius) every day at 2:00 P.M. local solar time and its minimum of 197 kelvins (–76 degrees C) just before sunrise. The pressure minimum of just under 6.7 millibars (roughly 0.67 percent of pressure at sea level on Earth) was reached on sol 20, the 20th Martian day after landing. On Mars the air pressure varies with the seasons. During winter, it is so cold that 20 to 30 percent of the entire atmosphere freezes out at the pole, forming a huge pile of solid carbon dioxide. The pressure minimum seen by Pathfinder indicates that the atmosphere was at its thinnest, and the south polar cap its largest, on sol 20.

Morning temperatures fluctuated abruptly with time and height; the sensors positioned 0.25, 0.5 and one meter above the spacecraft took different readings. If you were standing on Mars, your nose would be at least 20 degrees C colder than your feet. This suggests that cold morning air is warmed by the surface and rises in small eddies, or whirlpools, which is very different from what happens on Earth, where such large temperature disparities do not occur. Afternoon temperatures, after the air has warmed, do not show these variations.

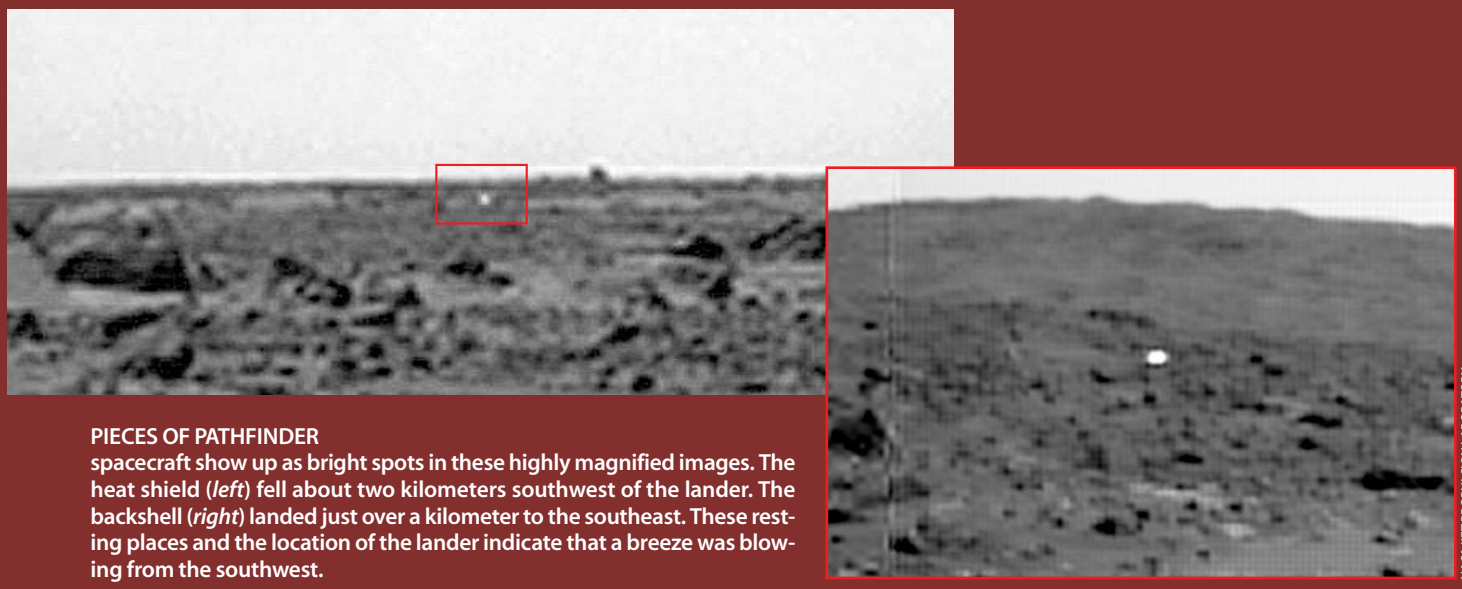
In the early afternoon, dust devils repeatedly swept across the lander. They showed up as sharp, short-lived pressure changes and were probably similar to events detected by the Viking landers and orbiters; they may be an important mechanism for raising dust into the Martian atmosphere. Otherwise, the prevailing winds were light (clocked at less than

Summary of Evidence for a Warmer, Wetter Mars

Over the past three decades, scientists have built the case that Mars once looked much like Earth, with rainfall, rivers, lakes, maybe even an ocean. Pathfinder has added evidence that strengthens this case (*red*).

GEOLOGIC FEATURE	PROBABLE ORIGIN	IMPLICATION
Riverlike valley networks	Water flow out of ground or from rain	Either atmosphere was thicker (allowing rain) or geothermal heating was stronger (causing groundwater sapping)
Central channel ("thalweg") in broader valleys	Fluid flow down valley center	Valleys were formed by water flow, not by landslides or sapping
Lakelike depressions with drainage networks; layered deposits in canyons	Flow through channels into lake	Water existed at the surface, but for unknown time
Possible strand lines and erosional beaches and terraces	Possible shoreline	Northern hemisphere might have had an ocean
Rimless craters and highly eroded ancient terrain	High erosion rates	Water, including rain, eroded surface
Rounded pebbles and possible conglomerate rock	Rock formation in flowing water	Liquid water was stable, so atmosphere was thicker and warmer
Abundant sand	Action of water on rocks	Water was widespread
Highly magnetic dust	Maghemite stain or cement on small (micron-size) silicate grains	Active hydrologic cycle leached iron from crustal materials to form maghemite

LISA BURNETT



PIECES OF PATHFINDER
spacecraft show up as bright spots in these highly magnified images. The heat shield (*left*) fell about two kilometers southwest of the lander. The backshell (*right*) landed just over a kilometer to the southeast. These resting places and the location of the lander indicate that a breeze was blowing from the southwest.

NASA/JET PROPULSION LABORATORY

36 kilometers per hour) and variable.

Pathfinder measured atmospheric conditions at higher altitudes during its descent. The upper atmosphere (altitude above 60 kilometers) was colder than Viking had measured. This finding may simply reflect seasonal variations and the time of entry: Pathfinder came in at 3:00 A.M. local solar time, whereas Viking arrived at 4:00 P.M., when the atmosphere is naturally warmer. The lower atmosphere was similar to that measured by Viking, and its conditions can be attributed to dust mixed uniformly in comparatively warm air.

As a bonus, mission scientists were able to use radio communications signals from Pathfinder to measure the rotation of Mars. Daily Doppler tracking and less frequent two-way ranging during communication sessions determined the position of the lander with a precision of 100 meters. The last such positional measurement was done by Viking more than 20 years ago. In the interim, the pole of rotation has precessed—that is, the direction of the tilt of the planet has changed, just as a spinning top slowly wobbles. The difference between the two positional measurements yields the precession rate. The rate is governed by the moment of inertia of the planet, a function of the distribution of mass within the planet. The moment of inertia had been the single most important number about Mars that we did not yet know.

From Pathfinder's determination of the moment of inertia we now know that Mars must have a central metallic core that is between 1,300 and 2,400 kilometers in radius. With assumptions about the mantle composition, derived from the compositions of the Martian meteorites and the rocks measured by the rover, scientists can now start to put constraints on interior temperatures. Before Pathfinder, the composition of the Martian meteorites argued for a core, but the size of this core was completely unknown. The new information about the interior will help geophysicists understand how Mars has evolved over time. In addition to the long-term precession, Pathfinder detected an annual variation in the planet's rotation rate, which is just what would be expected from the seasonal exchange of carbon dioxide between the atmosphere and the ice caps.

Taking all the results together suggests that Mars was once more Earth-like than previously appreciated. Some crustal materials on Mars resemble, in silicon content, continental crust on Earth. Moreover, the rounded pebbles and the possible conglomerate, as well as the abundant sand- and dust-size particles, argue for a formerly water-rich planet. The earlier environment may have been warmer and wetter, perhaps similar to that of the early Earth. In contrast, since floods produced the landing site 1.8 billion to 3.5 billion years ago, Mars has been a very un-Earth-like place. The site appears almost unaltered since it was deposited, indicating very low erosion rates and thus no water in relatively recent times.

Although we are not certain that early Mars was more like Earth, the data returned from Pathfinder are very suggestive. Information from the Mars Global Surveyor, now orbiting the Red Planet, should help answer this crucial question about our neighboring world. SA

The Author

MATTHEW P. GOLOMBEK is project scientist of Mars Pathfinder, with responsibility for the overall scientific content of the mission. He conducts his work at the Jet Propulsion Laboratory in Pasadena, Calif. He is chair of the Pathfinder Project Science Group, deputy of the Experiment Operations Team and a member of the project management group. He has written numerous papers on the spacecraft and its results and has organized press conferences and scientific meetings. Golombek's research focuses on the structural geology and tectonics of Earth and the other planets, particularly Mars. He became interested in geology because he wanted to know why Earth had mountains and valleys.

Further Reading

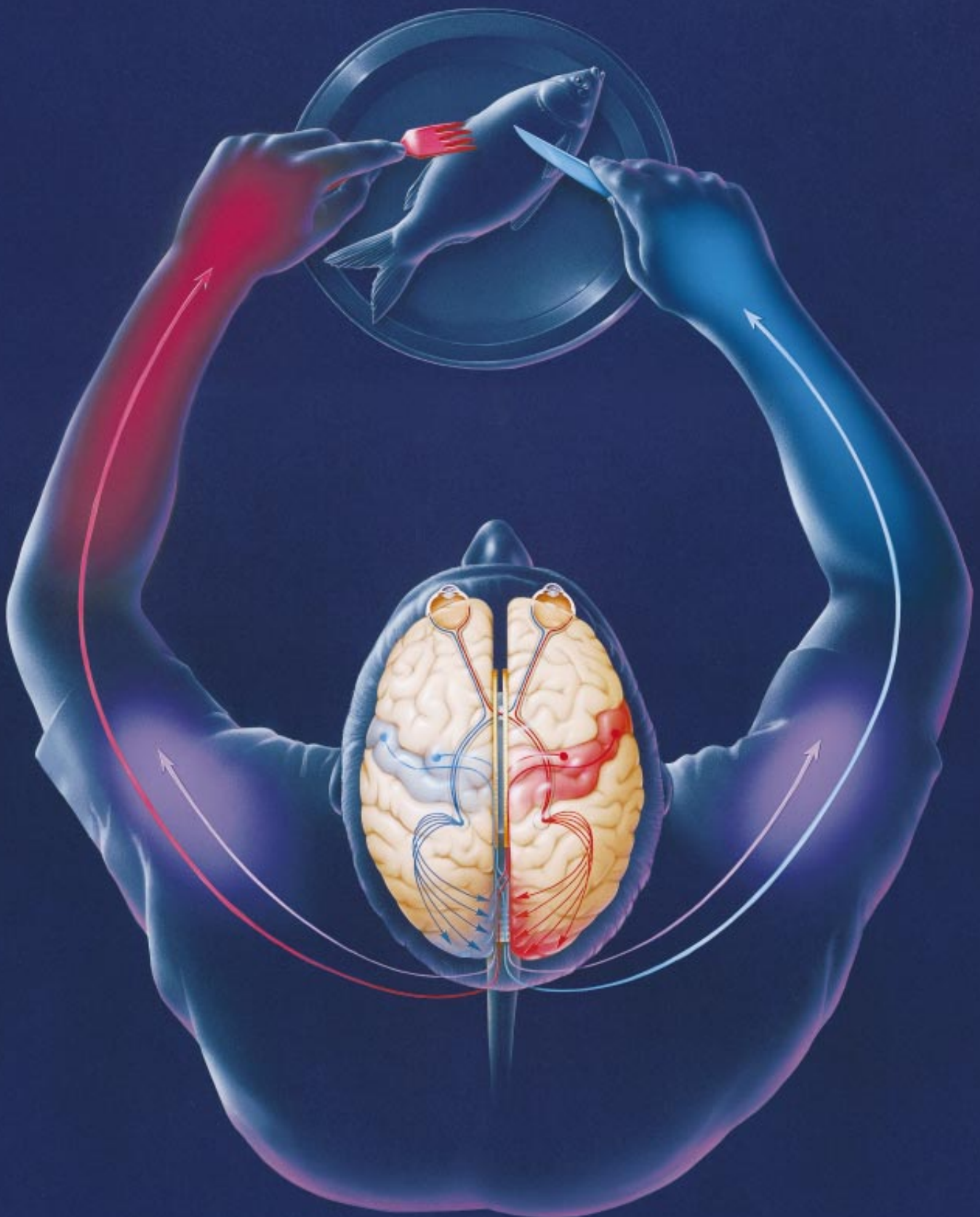
MARS. Edited by Hugh H. Kieffer, Bruce M. Jakosky, Conway W. Snyder and Mildred S. Matthews. University of Arizona Press, 1992.

WATER ON MARS. Michael H. Carr. Oxford University Press, 1996.

MARS PATHFINDER MISSION AND ARES VALLIS LANDING SITE. Matthew P. Golombek et al. in *Journal of Geophysical Research,* Vol. 102, No. E2, pages 3951–4229; February 25, 1997.

MARS PATHFINDER. Matthew P. Golombek et al. in *Science,* Vol. 278, pages 1734–1774; December 5, 1997.

The Split Brain Revisited



JOHN W. KARAPELOU

Groundbreaking work that began more than a quarter of a century ago has led to ongoing insights about brain organization and consciousness

by Michael S. Gazzaniga

About 30 years ago in these very pages, I wrote about dramatic new studies of the brain. Three patients who were seeking relief from epilepsy had undergone surgery that severed the corpus callosum—the superhighway of neurons connecting the halves of the brain. By working with these patients, my colleagues Roger W. Sperry, Joseph E. Bogen, P. J. Vogel and I witnessed what happened when the left and the right hemispheres were unable to communicate with each other.

It became clear that visual information no longer moved between the two sides. If we projected an image to the right visual field—that is, to the left hemisphere, which is where information from the right field is processed—the patients could describe what they saw. But when the same image was displayed to the left visual field, the patients drew a blank: they said they didn't see anything. Yet if we asked them to point to an object similar to the one being projected, they could do so with ease. The right brain saw the image and could mobilize a nonverbal response. It simply couldn't talk about what it saw.

The same kind of finding proved true for touch, smell and sound. Additionally, each half of the brain could control the upper muscles of both arms, but the muscles manipulating hand and finger movement could be orchestrated only by the contralateral hemisphere. In other words, the right hemisphere could control only the left hand and the left hemisphere only the right hand.

Ultimately, we discovered that the two hemispheres control vastly different aspects of thought and action. Each half has its own specialization and thus its own limitations and advantages. The left brain is dominant for language and speech. The right excels at visual-motor tasks. The language of these findings has become part of our culture: writers refer to themselves as left-brained, visual artists as right-brained.

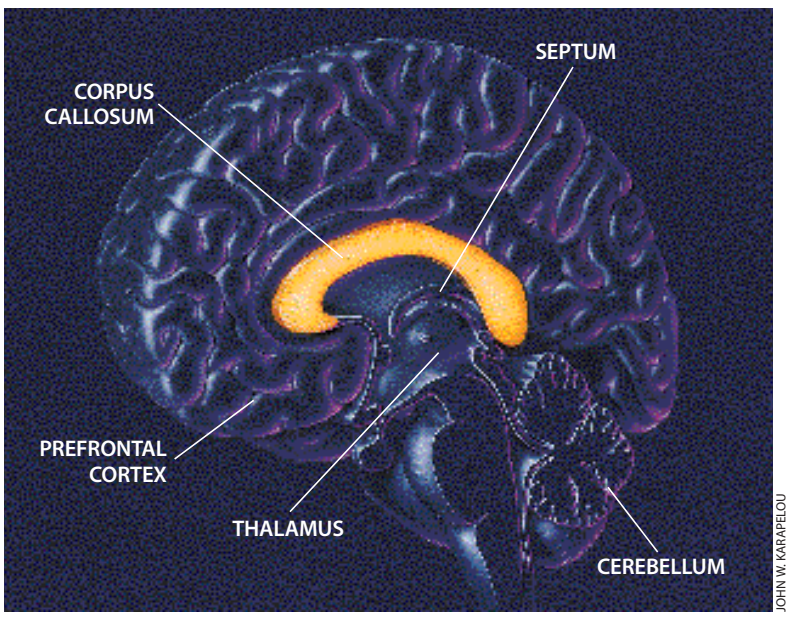
In the intervening decades, split-brain research has continued to illuminate many areas of neuroscience. Not only have we and others learned even more about how the hemispheres differ, but we also have been able to understand how they communicate once they have been separated. Split-brain studies have shed light on language, on mechanisms of perception and attention, and on brain organization as well as the potential seat of false memories. Perhaps most intriguing has been the contribution of these studies to our understanding of consciousness and evolution.

The original split-brain studies raised many interesting questions, including ones about whether the distinct halves could still "talk" to each other and what role any such communication played in thought and action. There are several bridges of neurons, called commissures, that connect the hemispheres. The corpus callosum is the most massive of these and typically the only one severed during surgery for epilepsy. But what of the many other, smaller commissures?

Remaining Bridges

By studying the attentional system, researchers have been able to address this question. Attention involves many structures in the cortex and the subcortex—the older, more primitive part of our brains. In the 1980s Jeffrey D. Holtzman of Cornell University Medical College found that each hemisphere is able to direct spatial attention not only to its own sensory sphere but also to certain points in the sensory sphere of the opposite, disconnected hemisphere. This discovery suggests that the attentional system is common to both hemispheres—at least with regard to spatial information—and can still operate via some remaining interhemispheric connections.

Holtzman's work was especially intriguing because it raised the possibility that there were finite attentional "resources." He posited that working on one kind of task uses certain



JOHN W. KARAPELOU

BRAIN WIRING is, in many cases, contralateral (*opposite page*). The right hemisphere processes information from the left visual field, whereas the left hemisphere processes data from the right visual field. For hand movement as well, the right hemisphere controls the hand and fingers of the left arm; the left hemisphere controls the right. Both hemispheres, however, dictate the movement of the upper arms. The two hemispheres are connected by neuronal bridges called commissures. The largest of these, and the one severed during split-brain operations, is the corpus callosum (*right*).

brain resources; the harder the task, the more of these resources are needed—and the more one half of the brain must call on the subcortex or the other hemisphere for help. In 1982 Holtzman led the way again, discovering that, indeed, the harder one half of a split brain worked, the harder it was for the other half to carry out another task simultaneously.

Recent investigations by Steve J. Luck of the University of Iowa, Steven A. Hillyard and his colleagues at the University of California at San Diego and Ronald Mangun of the University of California at Davis show that another aspect of attention is also preserved in the split brain. They looked at what happens when a person searches a visual field for a pattern or an object. The researchers found that split-brain patients perform better than normal people do in some of these visual-searching tasks. The intact brain appears to inhibit the search mechanisms that each hemisphere naturally possesses.

The left hemisphere, in particular, can exert powerful control over such tasks. Alan Kingstone of the University of Alberta found that the left hemisphere is "smart" about its search strategies, whereas the right is not. In tests where a person can deduce how to search efficiently an array of similar items for an odd exception, the left does better than the right. Thus, it seems that the more competent left hemisphere can hijack the intact attentional system.

Although these and other studies indicated that some communication between the split hemispheres remains, other apparent interhemispheric links proved illusory. I conducted an experiment with Kingstone, for instance, that nearly misled us on this front. We flashed two words to a patient and then asked him to draw what he saw. "Bow" was flashed to one hemisphere and "arrow" to the other. To our surprise, our patient drew a bow and arrow! It appeared as though he had internally integrated the information in one hemisphere; that hemisphere had, in turn, directed the drawn response.

We were wrong. We finally determined that integration had actually taken place on the paper, not in the brain. One hemisphere had drawn its item—the bow—and then the other hemisphere had gained control of the writing hand, drawing its stimulus—the arrow—on top of the bow. The image merely looked coordinated. We discovered this chimera by giving less easily integrated word pairs like "sky" and "scraper." The subject did not draw a tall building; instead he drew the sky over a picture of a scraper.

The Limits of Extrapolation

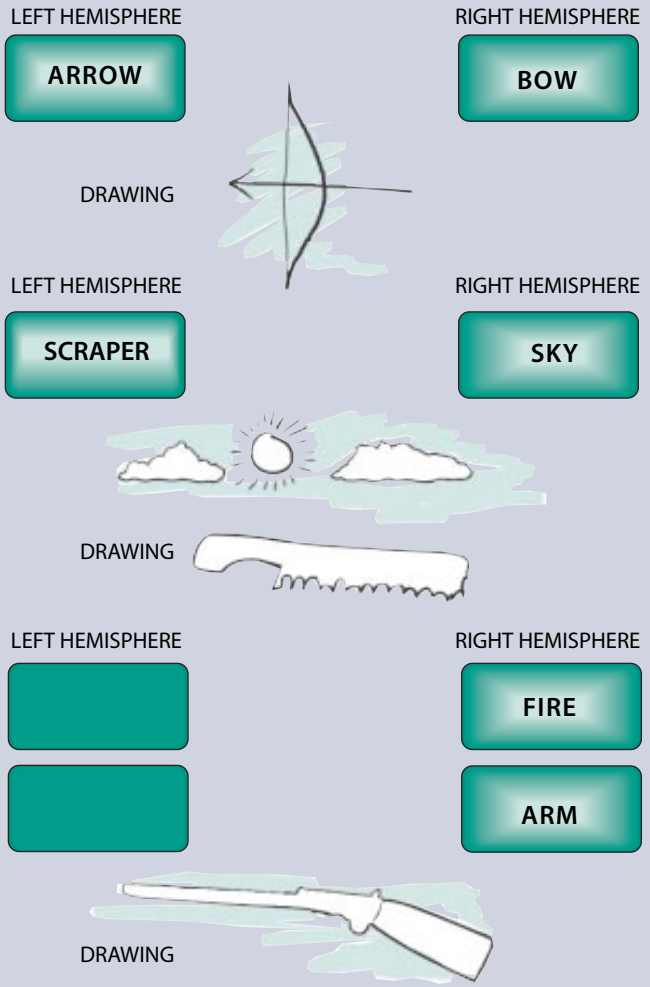
In addition to helping neuroscientists determine which systems still work and which are severed along with the corpus callosum, studies of communication between the hemispheres led to an important finding about the limits of nonhuman studies. Humans often turn to the study of animals to understand themselves. For many years, neuroscientists have examined the brains of monkeys and other creatures to explore the ways in which the human brain operates. Indeed, it has been a common belief—emphatically disseminated by Charles Darwin—that the brains of our closest relatives have an organization and function largely similar, if not identical, to our own.

Split-brain research has shown that this assumption can be spurious. Although some structures and functions are remarkably alike, differences abound. The anterior commissure provides one dramatic example. This small structure lies somewhat below the corpus callosum. When this commissure is left intact in otherwise split-brain monkeys, the animals retain the ability to

Testing for Synthesis

Ability to synthesize information is lost after split-brain surgery, as this experiment shows. One hemisphere of a patient was flashed a card with the word "bow"; the other hemisphere saw "arrow." Because the patient drew a bow and arrow, my colleagues and I assumed the two hemispheres were still able to communicate with each other—despite the severing of the corpus callosum—and had integrated the words into a meaningful composite.

The next test proved us wrong. We flashed "sky" to one hemisphere, "scraper" to the other. The resulting image revealed that the patient was not synthesizing information: sky atop a comblike scraper was drawn, rather than a tall building. One hemisphere drew what it had seen, then the other drew its word. In the case of bow and arrow, the superposition of the two images misled us because the picture appeared integrated. Finally, we tested to see whether each hemisphere could, on its own, integrate words. We flashed "fire" and then "arm" to the right hemisphere. The left hand drew a rifle rather than an arm on fire, so it was clear that each hemisphere was capable of synthesis. —*M.S.G.*



LAURIE GRACE

transfer visual information from one hemisphere to the other. People, however, do not transfer visual information in any way. Hence, the same structure carries out different functions in different species—an illustration of the limits of extrapolating from one species to another.

Even extrapolating between people can be dangerous. One of our first striking findings was that the left brain could freely process language and speak about its experience. Although the right was not so free, we also found that it could process some language. Among other skills, the right hemisphere could match words to pictures, do spelling and rhyming, and categorize objects. Although we never found any sophisticated capacity for syntax in that half of the brain, we believed the extent of its lexical knowledge to be quite impressive.

Over the years it has become clear that our first three cases were unusual. Most people's right hemispheres cannot handle even the most rudimentary language, contrary to what we initially observed. This finding is in keeping with other neurological data, particularly those from stroke victims. Damage to the left hemisphere is far more detrimental to language function than is damage to the right.

Nevertheless, there exists a great deal of plasticity and individual variation. One patient, dubbed J.W., developed the capacity to speak out of the right hemisphere—13 years after surgery. J.W. can now speak about information presented to the left or to the right brain.

Kathleen B. Baynes of the University of California at Davis reports another unique case. A left-handed patient spoke out of her left brain after split-brain surgery—not a surprising finding in itself. But the patient could *write* only out of her right, nonspeaking hemisphere. This dissociation confirms the idea that the capacity to write need not be associated with the capacity for phonological representation. Put differently, writing appears to be an independent system, an invention of the human species. It can stand alone and does not need to be part of our inherited spoken language system.

Brain Modules

Despite myriad exceptions, the bulk of split-brain research has revealed an enormous degree of lateralization—that is, specialization in each of the hemispheres. As investigators have struggled to understand how the brain achieves its goals and how it is organized, the lateralization revealed by split-brain studies has figured into what is called the modular model. Research in cognitive science, artificial intelligence, evolutionary psychology and neuroscience has directed attention to the idea that brain and mind are built from discrete units—or modules—that carry out specific functions. According to this theory, the brain is not a general problem-solving device whose every part is capable of any function. Rather it is a collection of devices that assists the mind's information-processing demands.

Within that modular system, the left hemisphere has proved quite dominant for major cognitive activities, such as problem solving. Split-brain surgery does not seem to affect these functions. It is as if the left hemisphere has no need for the vast computational power of the other half of the brain to carry out high-level activities. The right hemisphere, meanwhile, is severely deficient in difficult problem solving.

Joseph E. LeDoux of New York University and I discovered this quality of the left brain almost 20 years ago. We had asked a simple question: How does the left hemisphere respond to behaviors produced by the silent right brain? Each hemisphere was presented a picture that related to one of four pictures placed in front of the split-brain subject. The left and the right hemispheres easily picked the correct card. The left hand pointed to the right hemisphere's choice and the right hand to the left hemisphere's choice [*see illustration above*].

We then asked the left hemisphere—the only one that can talk—why the left hand was pointing to the object. It really did not know, because the decision to point to the card was made in the right hemisphere. Yet, quick as a flash, it made

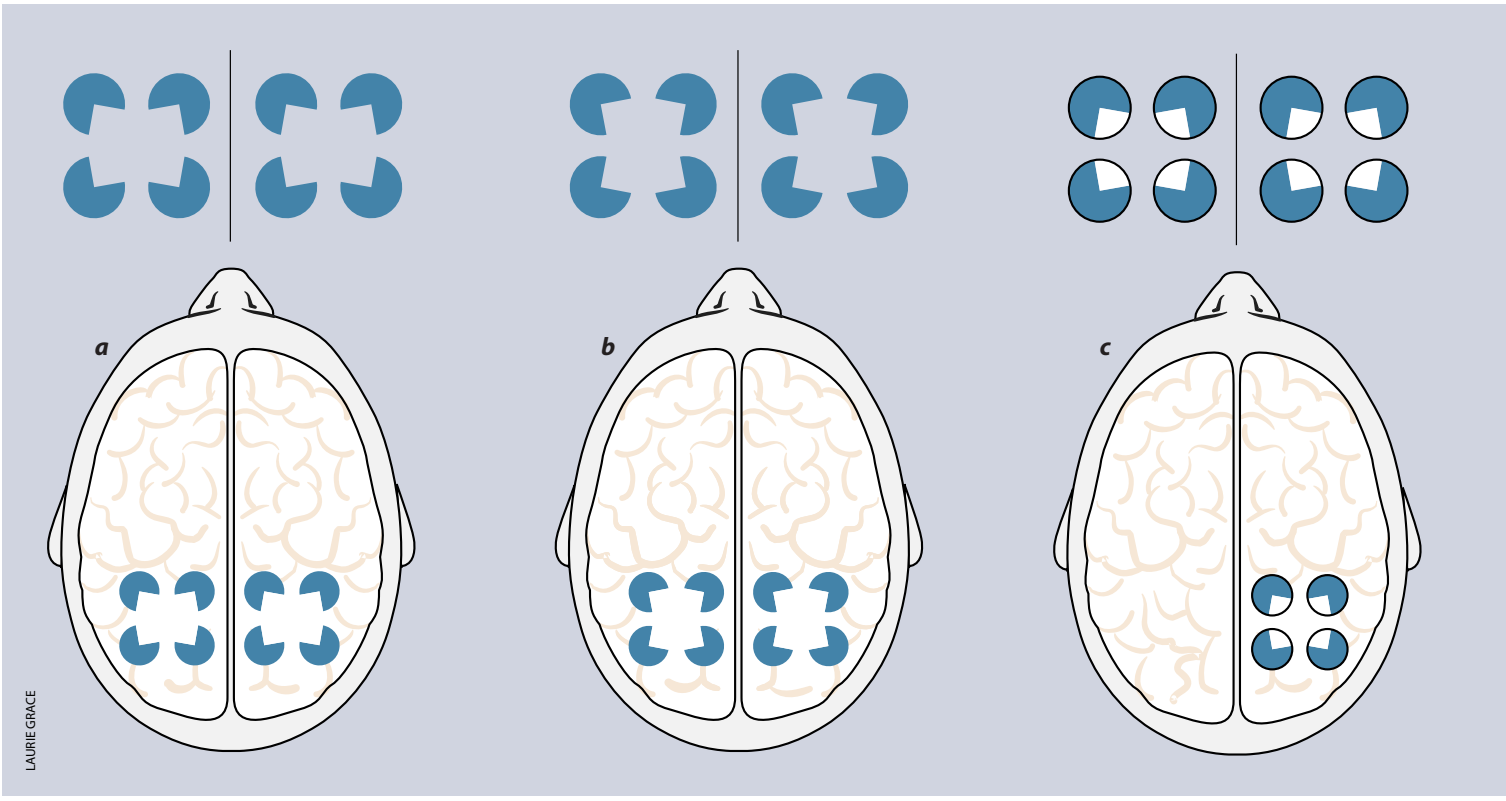
Finding False Memory

False memories originate in the left hemisphere. As this MRI image indicates, a region in both the right and left hemispheres is active when a false memory is recalled (*yellow*); only the right is active during a true memory (*red*). My colleagues and I studied this phenomenon by testing the narrative ability of the left hemisphere. Each hemisphere was shown four small pictures, one of which related to a larger picture also presented to that hemisphere. The patient had to choose the most appropriate small picture.

As seen below, the right hemisphere—that is, the left hand—correctly picked the shovel for the snowstorm; the right hand, controlled by the left hemisphere, correctly picked the chicken to go with the bird's foot. Then we asked the patient why the left hand—or right hemisphere—was pointing to the shovel. Because only the left hemisphere retains the ability to talk, it answered. But because it could not know why the right hemisphere was doing what it was doing, it made up a story about what it could see—namely, the chicken. It said the right hemisphere chose the shovel to clean out a chicken shed. —*M.S.G.*

MICHAEL B. MILLER

JOHN W. KARAPELOU



LAURIE GRACE

up an explanation. We dubbed this creative, narrative talent the interpreter mechanism.

This fascinating ability has been studied recently to determine how the left hemisphere interpreter affects memory. Elizabeth A. Phelps of Yale University, Janet Metcalfe of Columbia University and Margaret Funnell, a postdoctoral fellow at Dartmouth College, have found that the two hemispheres differ in their ability to process new data. When presented with new information, people usually remember much of what they experience. When questioned, they also usually claim to remember things that were not truly part of the experience. If split-brain patients are given such tests, the left hemisphere generates many false reports. But the right brain does not; it provides a much more veridical account.

This finding may help researchers determine where and how false memories develop. There are several views about when in the cycle of information processing such memories are laid down. Some researchers suggest they develop early in the cycle, that erroneous accounts are actually encoded at the time of the event. Others believe false memories reflect an error in reconstructing past experience: in other words, that people develop a schema about what happened and retrospectively fit untrue events—that are nonetheless consistent with the schema—into their recollection of the original experience.

The left hemisphere has exhibited certain characteristics that support the latter view. First, developing such schemata is exactly what the left hemisphere interpreter excels at. Second, Funnell has discovered that the left hemisphere has an ability to determine the source of a memory, based on the context or the surrounding events. Her work indicates that the left hemisphere actively places its experiences in a larger context, whereas the right simply attends to the perceptual aspects of the stimulus. Finally, Michael B. Miller, a graduate student at Dartmouth, has demonstrated that the left prefrontal regions of normal subjects are activated when they recall false memories.

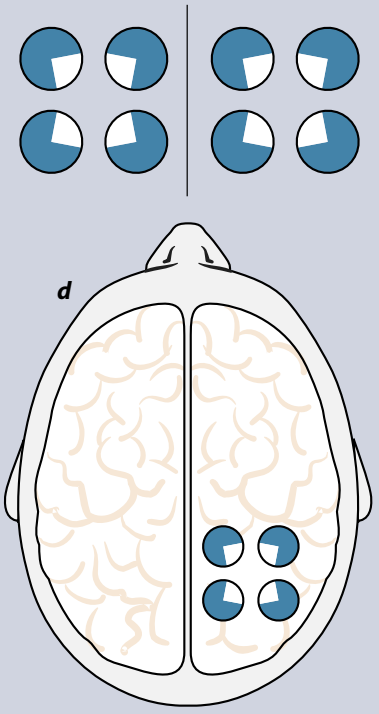
These findings all suggest that the interpretive mechanism of the left hemisphere is always hard at work, seeking the meaning of events. It is constantly looking for order and reason, even when there is none—which leads it continually to make mistakes. It tends to overgeneralize, frequently constructing a potential past as opposed to a true one.

The Evolutionary Perspective

George L. Wolford of Dartmouth has lent even more support to this view of the left hemisphere. In a simple test that requires a person to guess whether a light is going to appear on the top or bottom of a computer screen, humans perform inventively. The experimenter manipulates the stimulus so that the light appears on the top 80 percent of the time but in a random sequence. While it quickly becomes evident that the top button is being illuminated more often, people invariably try to figure out the entire pattern or sequence—and they deeply believe they can. Yet by adopting this strategy, they are correct only 68 percent of the time. If they always pressed the top button, they would be correct 80 percent of the time.

Rats and other animals, on the other hand, are more likely to "learn to maximize" and to press only the top button. It turns out the right hemisphere behaves in the same way: it does not try to interpret its experience and find deeper meaning. It continues to live only in the thin moment of the present—and to be correct 80 percent of the time. But the left, when asked to explain why it is attempting to figure the whole sequence, always comes up with a theory, no matter how outlandish.

This narrative phenomenon is best explained by evolutionary theory. The human brain, like any brain, is a collection of neurological adaptations established through natural selection. These adaptations each have their own representation—



Looking for Illusions

Illusory contours reveal that the human right brain can process some things the left cannot. Both hemispheres can "see" whether the illusory rectangles of this experiment are fat (*a*) or thin (*b*). But when outlines are added, only the right brain can still tell the difference (*c* and *d*). In mice, however, both hemispheres can consistently perceive these differences. For a rodent to perform better than we do suggests that some capabilities were lost from one hemisphere or the other as the human brain evolved. New capabilities may have squeezed out old ones in a race for space. —*M.S.G.*

that is, they can be lateralized to specific regions or networks in the brain. Throughout the animal kingdom, however, capacities are generally not lateralized. Instead they tend to be found in both hemispheres to roughly equal degrees. And although monkeys show some signs of lateral specialization, these are rare and inconsistent.

For this reason, it has always appeared that the lateralization seen in the human brain was an evolutionary add-on—mechanisms or abilities that were laid down in one hemisphere only. We recently stumbled across an amazing hemispheric dissociation that challenges this view. It forced us to speculate that some lateralized phenomena may arise from a hemisphere's losing an ability—not gaining it.

In what must have been fierce competition for cortical space, the evolving primate brain would have been hard-pressed to gain new faculties without losing old ones. Lateralization could have been its salvation. Because the two hemispheres are connected, mutational tinkering with a homologous cortical region could give rise to a new function—yet not cost the animal, because the other side would remain unaffected.

Paul M. Corballis, a postdoctoral fellow at Dartmouth, and Robert Fendrich of Dartmouth, Robert M. Shapley of New York University and I studied in many split-brain patients the perception of what are called illusory contours. Earlier work had suggested that seeing the well-known illusory contours of Gaetano Kanizsa of the University of Trieste was the right hemisphere's specialty. Our experiments revealed a different situation.

We discovered that both hemispheres could perceive illusory contours—but that the right hemisphere was able to grasp certain perceptual groupings that the left could not. Thus, while both hemispheres in a split-brain person can judge whether the illusory rectangles are fat or thin when no line is drawn around the openings of the "Pacman" figures, only the right can continue to make the judgment after the line has been drawn [*see illustration at left*]. This setup is called the amodal version of the test.

What is so interesting is that Kanizsa himself has demonstrated that mice can do the amodal version. That a lowly mouse can perceive perceptual groupings, whereas a human's left hemisphere cannot, suggests that a capacity has been lost. Could it be that the emergence of a human capacity like language—or an interpretive mechanism—chased this perceptual skill out of the left brain? We think so, and this opinion gives rise to a fresh way of thinking about the origins of lateral specialization.

Our uniquely human skills may well be produced by minute and circumscribed neuronal networks. And yet our highly modularized brain generates the feeling in all of us that we are integrated and unified. How so, given that we are a collection of specialized modules?

The answer may be that the left hemisphere seeks explanations for why events occur. The advantage of such a system is obvious. By going beyond the simple observation of events and asking why they happened, a brain can cope with these same events better, should they happen again.

Realizing the strengths and weaknesses of each hemisphere prompted us to think about the basis of mind, about this overarching organization. After many years of fascinating research on the split brain, it appears that the inventive and interpreting left hemisphere has a conscious experience very different from that of the truthful, literal right brain. Although both hemispheres can be viewed as conscious, the left brain's consciousness far surpasses that of the right. Which raises another set of questions that should keep us busy for the next 30 years or so. SA

The Author

MICHAEL S. GAZZANIGA is professor of cognitive neuroscience and director of the Center for Cognitive Neuroscience at Dartmouth College. He received his Ph.D. at the California Institute of Technology, where he, Roger W. Sperry and Joseph E. Bogen initiated split-brain studies. Since then, he has published in many areas and is credited with launching the field of cognitive neuroscience in the early 1980s. Gazzaniga likes to ski and to arrange small, intense intellectual meetings in exotic places.

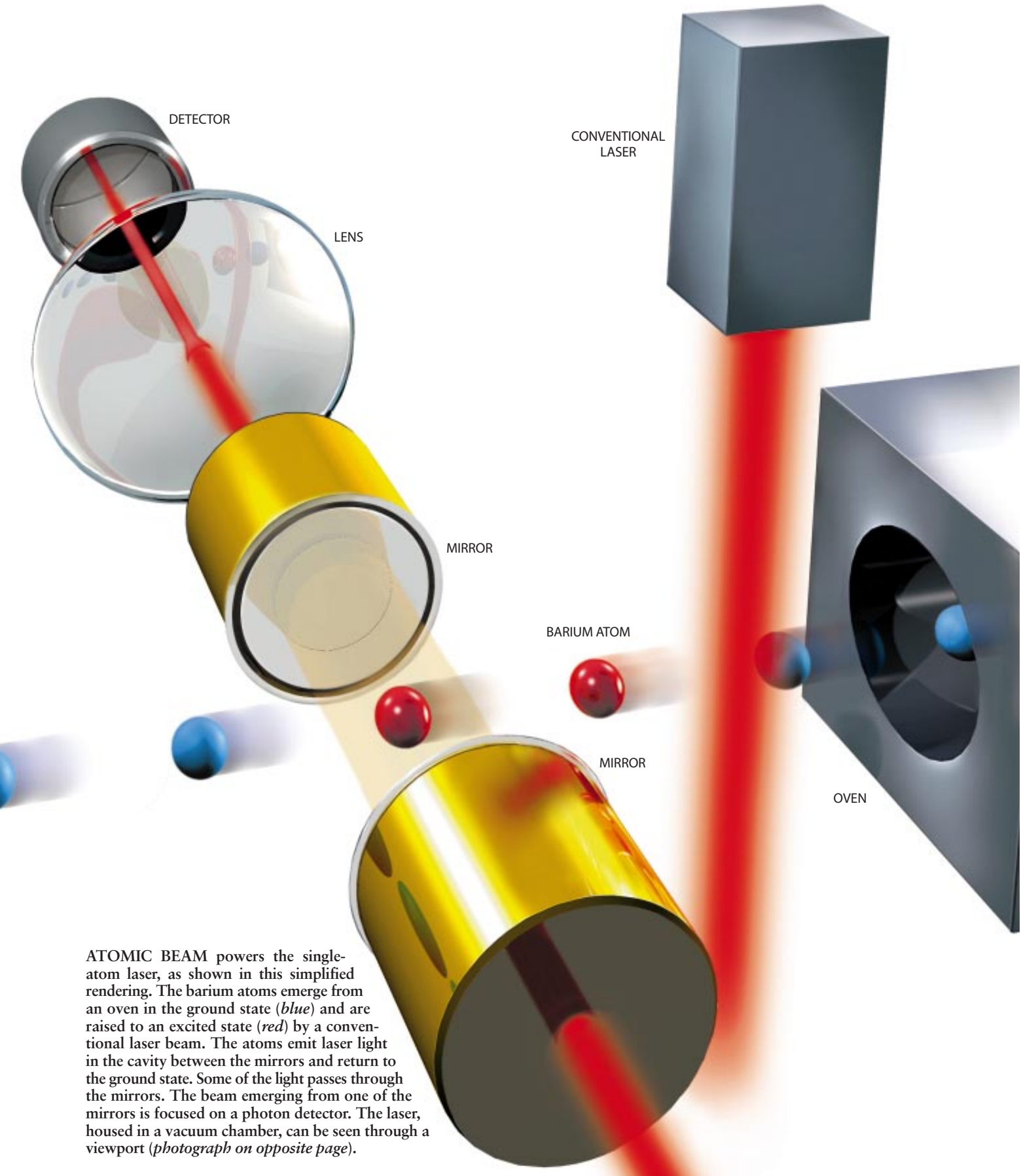
Further Reading

HEMISPHERIC SPECIALIZATION AND INTERHEMISPHERIC INTEGRATION. M. J. Tramo, K. Baynes, R. Fendrich, G. R. Mangun, E. A. Phelps, P. A. Reuter-Lorenz and M. S. Gazzaniga in *Epilepsy and the Corpus Callosum*. Second edition. Plenum Press, 1995.

HOW THE MIND WORKS. Steven Pinker. W. W. Norton, 1997.

THE MIND'S PAST. Michael S. Gazzaniga. University of California Press, 1998.

THE TWO SIDES OF PERCEPTION. Richard B. Ivry and Lynn C. Robertson. MIT Press, 1998.



ATOMIC BEAM powers the single-atom laser, as shown in this simplified rendering. The barium atoms emerge from an oven in the ground state (*blue*) and are raised to an excited state (*red*) by a conventional laser beam. The atoms emit laser light in the cavity between the mirrors and return to the ground state. Some of the light passes through the mirrors. The beam emerging from one of the mirrors is focused on a photon detector. The laser, housed in a vacuum chamber, can be seen through a viewport (*photograph on opposite page*).

MARK GERBER

The Single-Atom Laser

A new type of laser that harnesses the energy of individual atoms reveals how light interacts with matter

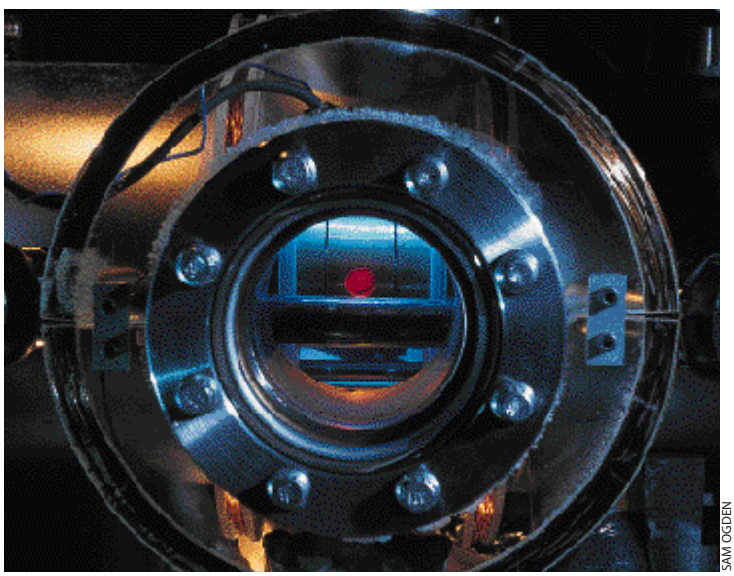
by Michael S. Feld and Kyungwon An

Since its invention 40 years ago, the laser has become a nearly ubiquitous device. Teachers and tour guides use its slender beam as a high-tech pointer. Supermarket clerks use lasers to read the bar codes on groceries, and music lovers enjoy high-quality recordings thanks to the miniature lasers in their compact-disc players. All these lasers rely on the same principle: a huge number of atoms or molecules—from the millions to the trillions—working together to produce an intense, monochromatic beam of light. Recently, however, we have developed a laser that works with the smallest number of atoms possible: just one.

The single-atom laser cannot read bar codes or play music; its output power is about a trillionth of a watt (lasers in CD players are many millions of times more powerful). But the device has already proved to be an important experimental tool. The light generated by the single-atom laser exhibits properties that can be explained only by quantum mechanics, the theory that governs interactions on the atomic and subatomic scale. By analyzing the laser's performance under varying conditions, scientists can test the predictions of quantum theory and gain new insights into the nature of laser light.

The Conventional Laser

To appreciate the single-atom laser, one must first understand how conventional lasers operate. Every laser contains two essential components: an optical resonator, usually formed by two aligned mirrors that reflect light back and forth; and a laser medium positioned between the mirrors, which generates and amplifies the light. In the helium-neon laser, for example, this medium is a gas; in the neodymium-YAG laser, the medium consists of neodymium ions embedded in an yttrium aluminum garnet crystal. The atoms or molecules in the medium are not homogeneous—they are distributed among many distinct quantum states, or energy levels. Only one of the elements or compounds in the medium participates in the laser process, and only a tiny fraction of these atoms actually emit laser light; these are known as the active atoms, whereas the remainder are called background atoms. The active atoms alternate between two of the energy states. Light amplification can occur only when the number of active atoms in the higher-energy, or excited, state exceeds the number in the lower-energy state. This condition is called population inversion.



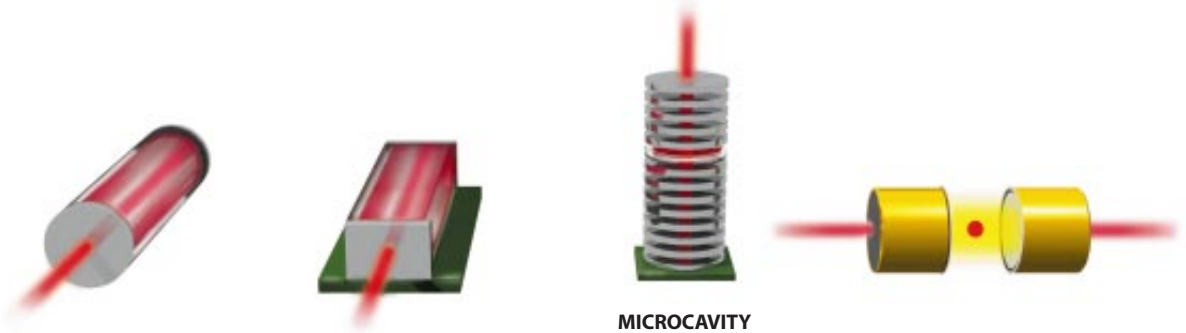
SAM OGDEN

The laser uses an energy source—an electrical discharge, for example—to raise the active atoms from the lower-energy state to the excited state. But excited atoms do not remain excited indefinitely: their natural tendency is to drop back to the lower-energy state. When they do, they emit light in random directions, a process called spontaneous emission. A small proportion of this light, typically a few millionths, heads toward one of the mirrors and is reflected back to the laser medium. The reflected light induces the excited atoms to emit additional photons with the same wavelength, direction and phase. This production of light with the same characteristics as the original light, called stimulated emission, is responsible for the many extraordinary properties of laser radiation. As the light bounces back and forth between the mirrors, it is continually amplified by the laser medium. A small portion of this light is transmitted through one of the mirrors, which is designed to be slightly less reflective than the other, allowing the familiar laser beam to emerge.

For laser oscillation to occur, the laser gain—the increase in the intensity of the light as it passes through the laser medium—must exceed the loss of beam intensity caused by imperfections in the mirrors and other factors. This requirement is known as the laser threshold condition. Conventional lasers must have a medium containing huge numbers of atoms or molecules to provide enough gain to exceed the loss. For example, a helium-neon laser emitting one milliwatt of power contains several thousand trillion neon atoms and about 10 times as many helium atoms. The light amplification reaches equilibrium when about one billion photons are bouncing between the resonator mirrors. In other words, several million neon atoms and tens of millions of helium atoms are needed to maintain each photon in the resonator.

There are several ways to reduce the number of atoms required. One can limit the loss of beam intensity by improving

LASER TYPES



	HELIUM-NEON	NEODYMIUM-YAG	MICROCAVITY SEMICONDUCTOR	SINGLE-ATOM
DISTANCE BETWEEN MIRRORS	20 TO 100 CENTIMETERS	5 TO 15 CENTIMETERS	2 TO 5 MICRONS	1 MILLIMETER
NUMBER OF ATOMS	10^{16} NEON ATOMS	10^{19} NEODYMIUM IONS	10^{6} ELECTRON-HOLE PAIRS	1 BARIUM ATOM
POWER (WATTS)	0.001	1.0	0.0001	0.000000000001

MARK GERBER

FOUR LASERS exhibit very different characteristics, although they all produce monochromatic beams of light. The helium-neon laser can be found in any supermarket bar-code scanner, whereas the more powerful neodymium-YAG laser is used for surgery. The microcavity semiconductor laser, still in the developmental stage, may someday be used in optical computers. The single-atom laser is primarily an experimental tool, but it may also find applications in low-noise information processing and precision spectroscopy.

the reflectivity of the mirrors, so that the photons are stored longer in the resonator and can build up more easily. And in some cases, one can increase the laser gain by reducing the number of background atoms, which interfere with light amplification by colliding with the active atoms. But in practice, even the most efficient conventional lasers require at least 100,000 atoms for every photon stored in the resonator. Clearly, a conventional laser cannot generate a beam with only one atom.

Atoms in Cavities

Our single-atom laser uses an alternative method of light amplification based on a process called quantized Rabi oscillation. This phenomenon has been studied by scientists working in the field of cavity quantum electrodynamics, which governs the behavior of atoms in extremely small resonators [see "Cavity Quantum Electrodynamics," by Serge Haroche and Jean-Michel Raimond; Scientific American, April 1993]. It is perhaps the most elementary form of interaction between light and matter.

Rabi oscillation is the periodic exchange of energy between atoms and an electromagnetic field. Physicist I. I. Rabi first studied this process in the 1930s. When Rabi exposed a sample of atoms to specially tuned radio waves, he found that the ground-state atoms rose to an excited state as they absorbed energy from the field. This absorption occurred because the energy of the radio-frequency photons matched the difference in energy between the ground state and the excited state. Once all the atoms had reached the excited state, however, they could not absorb more energy from the waves. So the process reversed itself: with continued irradiation the atoms began to emit their energy back into the applied field and return to the ground state. Then they began to absorb energy from the field again, repeating the cycle.

In Rabi's experiments the actions of individual atoms and photons were not observable. Because radio-frequency photons have very little energy, even a low-power radio wave contains an enormous number of them. So many energy exchanges take place between the atoms and the photons that their effects average out, leaving no opportunity to study the quantum-mechanical nature of the exchange. But in the early 1960s Edwin T. Jaynes of Washington University and Frederick W. Cummings of the research laboratories of Ford Motor Company developed a theory to explain how a single atom with two energy levels would interact with a light wave that contains only a small number of photons. With just one atom, the frequencies of the Rabi oscillations cannot be arbitrary—they must take quantized values, much like the distinct energy levels of atoms. In other words, a two-level atom will emit and absorb photons at distinct rates that are determined by the strength of the surrounding electromagnetic field.

One remarkable consequence of this theory is that an excited atom can be induced to emit a photon simply by placing it in a very small cavity. If the cavity is resonant—that is, if its walls are reflective and its dimensions are adjusted so that the photons emitted by the atom can build up inside—a quantum-mechanical "coupling" occurs, causing the atom to emit a photon much more quickly than it would in free space. If the atom remains in the cavity, it will absorb the photon it emitted and then repeat the cycle. This process is called vacuum Rabi oscillation because there is initially no electromagnetic field in the cavity. If one or more photons are in the cavity before the excited atom enters, it will undergo quantized Rabi oscillation and emit and absorb photons at an even faster rate.

This phenomenon was demonstrated in the laboratory in 1984, when Herbert Walther of the Max Planck Institute in Garching, Germany, developed the micromaser, a microwave device based on quantum theory. In Walther's experiment, a beam of Rydberg atoms—atoms in which the outer electrons are excited into large, circular orbits—flowed one by one through a small metal cavity with highly reflective walls. The cavity was analogous to the resonator in a laser—its dimensions conformed to the wavelength of the microwave photons emitted by the Rydberg atoms when they dropped back to a lower-energy state. As the atoms passed through the cavity, they emitted photons at an increased rate, as predicted by Jaynes and Cummings. The photons were able to accumulate in the resonator because its walls were superconducting and could be cooled to just above absolute zero to maximize their reflectivity.

The single-atom laser is an optical version of the micromaser. Excited two-level atoms stream one by one into a tiny resonator and emit infrared photons that

are just beyond the visible range. The first photon is emitted into the empty cavity by vacuum Rabi oscillation, and further amplification of the light occurs through the quantized Rabi oscillation process. As the number of photons in the cavity grows, the probability that an atom passing through the resonator will emit another photon increases. This enhancement is the fundamental process underlying the phenomenon of stimulated emission in a conventional laser.

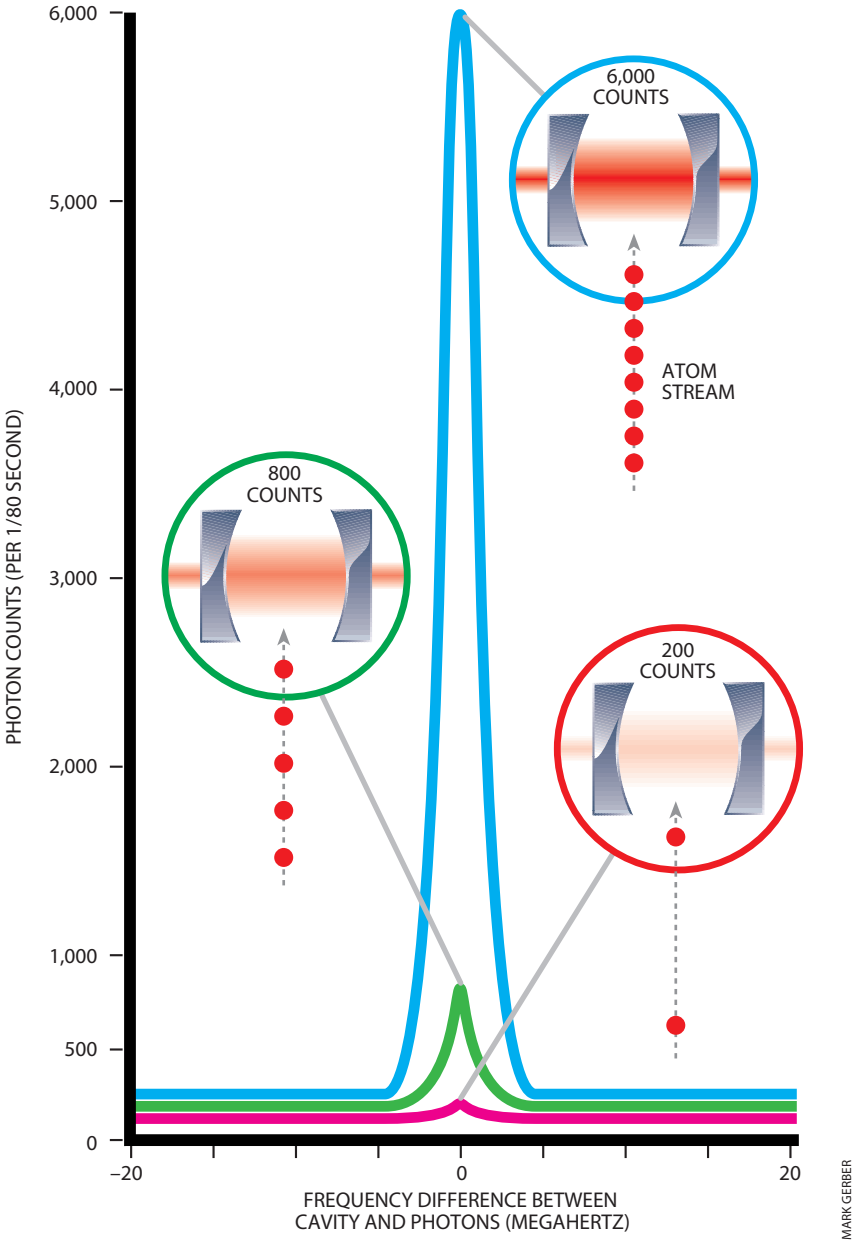
The construction of a working single-atom laser depended on designing an optical resonator that could store a photon for a relatively long time before it is absorbed by one of the mirrors or transmitted out of the cavity. We used a new type of resonator called a supercavity, which consists of two precisely aligned mirrors with ultrahigh reflectivity. In the 1960s, while engineers at the National Aeronautics and Space Administration were trying to develop a propulsion system based on high-speed ions, they discovered that the ion beams coated the walls of a vacuum chamber with a highly reflective film of ions. In the 1970s and 1980s engineers used the ion beams to coat mirrors for laser gyroscopes. Although the technology provided unsurpassed reflectivity, tiny imperfections in the shape of the mirror limited its performance in a resonator.

If, however, experimenters used only a small spot on the mirror—typically on the scale of one millimeter—the imperfections would become unimportant. The mirrors would then work well in a resonator, with reflectivity as high as 99.9999 percent. (The mirrors in a conventional laser are generally 99 percent reflective, whereas a typical wall mirror is only 90 percent reflective.) Photons can be stored in such a resonator 10,000 times more efficiently than in a conventional laser resonator. The mirrors in our single-atom laser were 99.9997 percent reflective and spaced just one millimeter apart. Photons could be reflected back and forth about a quarter of a million times before being absorbed by the mirrors or transmitted out of the resonator.

Unfortunately, it is difficult to keep the supercavity "in tune" with the frequency of the photons emitted by the atoms when they fall to a lower-energy state. If the mirrors move even slightly and the resonant match is lost, Rabi oscillations cannot occur and the atoms in the cavity will not emit photons. To ensure atom-cavity resonance, we used a piezoelectric transducer—a dielectric crystal that converts an applied voltage to mechanical stress—to adjust the spacing between the mirrors. A servo loop monitored the spacing and corrected any deviation, even as small as a ten-thousandth of a nanometer, from the proper distance between the mirrors.

An equally crucial issue was selecting an appropriate atom to put inside the resonator. The atom had to have a suitable pair of energy levels and a low rate of spontaneous emission, because this process would disrupt the interaction between the atom and the cavity. We chose to use barium atoms, which emit photons with a wavelength of 791 nanometers when they drop from the excited state to the ground state. These atoms were prepared by evaporating barium metal in an oven and then directing the vapor toward the gap between the mirrors. The oven produced a narrow beam of barium atoms traveling at an average speed of 360 meters per second. Because the resonator was small and the density of the stream of atoms was low, no more than one barium atom was inside the resonator at any moment.

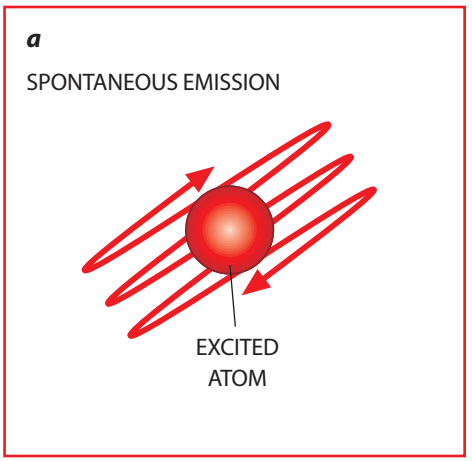
Just before the barium atoms entered the cavity, they passed through a beam of light from a conventional titanium-



BEAM INTENSITY of the single-atom laser rises sharply when a dense stream of barium atoms enters the resonator. When the beam of atoms is sparse, laser output is much lower. Note that the photon count falls to a background level when the resonant frequency of the cavity does not match the frequency of the emitted photons.

UPI/CORBIS-BETTMANN

RABI OSCILLATIONS—first studied by physicist I. I. Rabi (*left*) in the 1930s—cause the emission of photons in the single-atom laser. In free space, an excited atom will spontaneously emit a photon in a random direction (*a*). But in a resonant cavity, the atom will emit a photon more quickly toward the precisely spaced mirrors (*b*). This process is called vacuum Rabi oscillation. If another excited atom then enters the cavity, the presence of the photon will cause the atom to undergo quantized Rabi oscillation. The atom will emit an identical photon in the same direction but at an even faster rate (*c*).



sapphire laser, which was precisely tuned to excite the ground-state atoms to the higher-energy state. If left alone, the atoms would have spontaneously dropped back to the ground state in an average of three millionths of a second, emitting photons with a wavelength of 791 nanometers. But because the cavity had the same resonant wavelength, it induced some of the barium atoms to emit their photons during the 200 billionths of a second when they were streaming between the mirrors.

When the first atom entered the empty resonator, there was a 23 percent chance that it would undergo vacuum Rabi oscillation and emit a photon. But once the first photon was emitted, the electromagnetic field created inside the resonator exerted a stronger influence on the next barium atom to enter the cavity. The chance that this atom would emit a photon jumped to 42 percent. As the number of photons in the resonator increased, the probability of photon emission rose still higher. What is more, all the emitted photons shared the same direction and phase, which were defined by the geometry of the resonator. The result was a weak beam of laser photons, transmitted out of the cavity in a direction perpendicular to the beam of atoms.

Because of the low resonator loss, the photons could be stored in the cavity for almost one millionth of a second, a considerable amount of time by atomic standards. We estimated the number of photons stored in our resonator by measuring the amount of laser light transmitted through one of the cavity's mirrors. We used a highly efficient detector that could count 40 percent of all the photons that reached it.

In our experiment the photon buildup continued until the rate of loss from absorption and transmission through the mirrors equaled the rate of photon emission from the barium atoms. We were able to vary the density of the stream of atoms by adjusting the temperature of the barium oven. When the average number of atoms in the cavity was 0.1—in other words, when the cavity was occupied by a barium atom 10 percent of the time—photon buildup was minimal. Most of the emitted photons left the resonator before another barium atom could enter. But when we raised the average number of atoms in the cavity to 0.4, the atoms emitted about a million photons each second, enough to maintain one photon in the resonator at all times. The presence of a photon in the cavity increased the probability of additional photon emission: when we boosted the average number of atoms to 0.7, the power of the laser rose sevenfold.

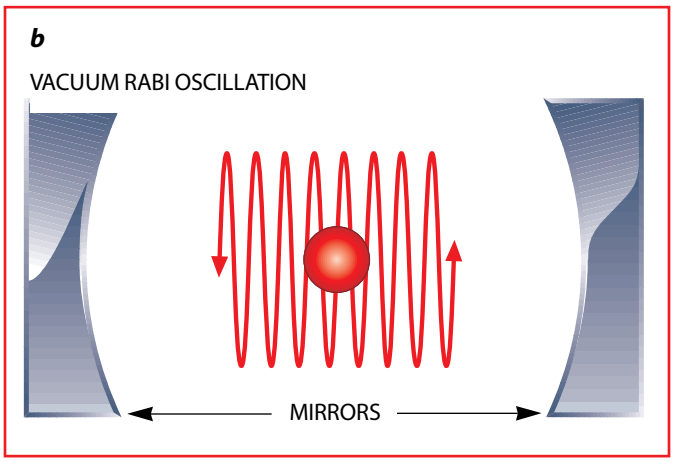
Lasers of the Future

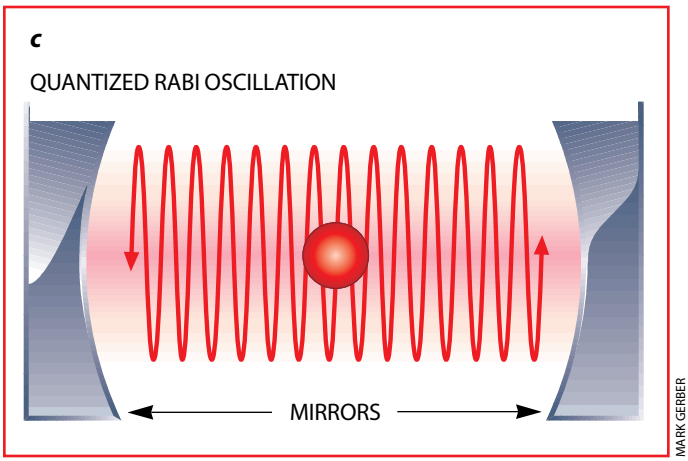
The single-atom laser is remarkably efficient. As the photons build up in the resonator, the probability that the barium atoms will emit another photon can theoretically approach 100 percent. In our prototype the probability of emission peaked at 50 percent, which meant that half the energy absorbed by the barium atoms was converted to laser light. Most conventional lasers, in contrast, have efficiencies ranging from 1 to about 30 percent.

But the true value of the single-atom laser lies in its usefulness as an experimental tool. Because the laser's light is generated by quantum-mechanical phenomena, scientists can test the predictions of quantum theory by setting up atomic interactions in the resonator and then observing the output from the device. Our original laser was not ideal for such experiments, however, because the atomic interactions in the resonator were not uniform. The electromagnetic field in the supercavity had the form of a standing wave between the mirrors, with the amplitude rising and falling in a sinusoidal curve, like the vibrations in a taut piano wire. Because of this setup, the interactions between the atoms and the cavity varied according to the paths of the atoms: those passing through the areas of high amplitude emitted photons, whereas those passing through the low-amplitude areas were unaffected.

We have solved this problem by tilting the atomic beam slightly away from a 90-degree angle with the orientation of the mirrors. Because of the Doppler shift, the atom no longer "sees" the field as a standing wave but rather as a pair of traveling waves moving in opposite directions. We can adjust the distance between the mirrors such that only one of the traveling waves is resonant with the atom. This adjustment makes the atomic interactions more uniform: because all the atoms stream through high- and low-amplitude regions of the traveling wave, they share the same chance of emitting a photon if they pass through the same field.

Another problem with our original laser was the wide velocity distribution of the barium atoms emerging from the





MARK GERBER

oven. The faster atoms traversing the cavity had less time to interact with the field and thus were less likely to emit photons. To make the velocities more uniform, we have modified our system for exciting the barium atoms, irradiating them with two conventional lasers to ensure that only the atoms with the desired velocity are raised to the higher-energy state.

With these improvements, we are now preparing to analyze the emission spectrum of the single-atom laser. Previous experiments have shown that an atom inside a resonant cavity will interact with a beam of light from a conventional laser. When the cavity is empty, the intensity of the beam passing through it will rise steeply as the frequency of the laser approaches the resonant frequency of the cavity. In other words, the cavity's "tuning curve" is single-peaked. But when an atom is inside the cavity, the intensity of the transmitted laser beam will rise sharply at two frequencies above and below the resonant frequency. The cavity's tuning curve becomes double-peaked.

The emission spectrum of the single-atom laser should also become a double-peak curve when the stream of atoms is sparse—that is, when the time interval between atoms is much longer than the cavity decay time, and all the photons leave the resonator before another atom enters. When the stream of atoms is denser and photons build up in the resonator, the emission spectrum should evolve to the single-peak curve characteristic of laser light. But there are no theoretical predictions of how this transition occurs. By providing spectra for the intermediate stages, the single-atom laser may improve our understanding of how laser light evolves from chaotic photon emission.

We also plan to use the device to investigate the "trapped state," a quantum phenomenon that occurs when an excited atom enters the cavity and undergoes one or more complete Rabi oscillations. In a complete oscillation the atom emits a photon into the cavity and then absorbs a photon from the field and leaves the resonator in an excited state. Because the atom enters and exits the cavity in the same state, the number of photons in the field does not increase. We did not observe complete Rabi oscillations in our original experiment, because the barium atoms were moving too rapidly; their transit time through the cavity was just long enough for them to undergo one sixth of an oscillation. We can study trapped states only if we slow the stream of atoms. This in turn requires us to increase the amount of time that photons can be stored in the cavity. We are currently trying to triple the storage time of our resonator.

These studies may yield a practical benefit as well: because trapped states have very low uncertainty in field amplitude, a single-atom laser operating in that mode could be used for low-noise information processing or precision spectroscopy. The knowledge gained from the single-atom laser may also hasten the development of microcavity semiconductor lasers—tiny devices that may someday be used to create optical computers. Many scientists believe they could greatly enhance the efficiency of semiconductor lasers if they could manipulate the basic process of photon emission. The single-atom laser may help scientists master this process by shining its light on the quantum-mechanical world. SA

The Authors

MICHAEL S. FELD and KYUNGWON AN are co-inventors of the single-atom laser. Feld is professor of physics at the Massachusetts Institute of Technology and director of the George R. Harrison Spectroscopy Laboratory. He received his Ph.D. from M.I.T. in 1967. His works include the first demonstration of superradiance in the optical region. Feld has also been active in biomedical applications of lasers and spectroscopy. An received his Ph.D. in physics from M.I.T. in 1994; the single-atom laser was his dissertation project. He continued at M.I.T. as a research scientist in quantum optics and atomic physics until January 1998. An is now an assistant professor at the Korea Advanced Institute of Science and Technology in Taejon, South Korea.

The authors thank Chris Fang-Yen, Abdul Aljalal and Ramachandra R. Dasari for useful discussions. This work was supported by the National Science Foundation under grants PHY-9512056 and CHE-9304251.

Further Reading

CAVITY QUANTUM ELECTRODYNAMICS. Serge Haroche and Daniel Kleppner in *Physics Today,* Vol. 42, No. 1, pages 24–30; January 1989.

THE MICROMASER: A PROVING GROUND FOR QUANTUM PHYSICS. G. Raithel et al. in *Cavity Quantum Electrodynamics*. Edited by Paul R. Berman. Academic Press, 1994.

THE MICROLASER: A FUNDAMENTAL QUANTUM GENERATOR OF LIGHT. Kyungwon An, Ramachandra R. Dasari and Michael S. Feld in *Atomic and Quantum Optics: High-Precision Measurements,* SPIE Proceedings Series, Vol. 2799, pages 14–21; 1996.

Mating Strategies in Butterflies

Butterflies meet, woo and win their mates using seductive signals and clever strategies honed by evolution

by Ronald L. Rutowski

As any postpubescent human knows, interest in potential mating partners is heavily influenced by sensory cues. A glimpse of lustrous hair or of piercing eyes can suddenly cause a man to be smitten with a woman, or she with him. The detection of a provocative scent or a sensuous touch may also kindle desire.

Grace Kelly's or Errol Flynn's obvious charms notwithstanding, an unbiased observer might find butterflies far more sensually appealing than humans. Perhaps unsurprisingly then, visual and other sensory cues also appear to govern these tiny creatures' decisions about mates. At stake is nothing less than the opportunity to produce offspring carrying an individual's genes through time.

Although Charles Darwin knew nothing of genes, he knew a great deal about sex (Gregor Mendel's work was not rediscovered until the early 1900s). Darwin first argued in 1871 that species tend to evolve attributes and behaviors that enhance courtship—and thus reproductive success. Some traits might render an individual more attractive to the opposite sex, whereas others might enable triumph over competing suitors. He specifically pondered butterflies when proposing this theory of sexual selection, largely because of the insects' vivid markings, which he felt might be influential in mate choice. "Their colours and elegant patterns are arranged and exhibited as if for display," he wrote in *The Descent of Man, and Selection in Relation to Sex*. "Hence I am led to suppose that the females generally prefer, or are most excited by the more brilliant males."

Recent experimental work with butterflies has borne out Darwin's suspicions of more than a century ago. Color is now known to spark sexual interest for some species in the butterfly world, as do other sensory signals that were beyond Darwin's human perception. But the creatures are more discerning than this observation might suggest. Ostentatious coloration or scent may do more than attract attention. Appearance and aroma may be shorthand notations of their bearer's health and heartiness.

Color Cues

The clearest evidence for the role of color in sexual attraction among butterflies comes from studies of species in which males and females have distinctly different appearances. Obviously, to mate successfully, individuals must be able to determine whether other conspecific butterflies are of their own or of the opposite sex. The rest, it can be argued, is fine-tuning.

A gorgeous butterfly species whose males and females differ in color is the Little Yellow, *Eurema lisa*. Both sexes appear an identical yellow to the human eye, the shade being produced by pigments in the tiny scales that cover the butterflies' translucent wings. Males and females look quite different to butterflies, however, which perceive light at wavelengths beyond the human visible range and into the ultraviolet. Yellow wing scales on the upper surface of the males' wings reflect ultraviolet light, and those of females do not.

On encountering a female, a Little Yellow male flutters about her briefly before landing and attempting to copulate. On confronting another male, he speeds away and continues his search. These

BILL BEATTY *Animals Animals*

MATING is the culmination of a courtship in which visual and chemical cues play crucial roles. Shown are Orange Sulphurs (*Colias eurytheme*), a butterfly common to many regions of North America. The butterfly holding onto the leaf is the male.

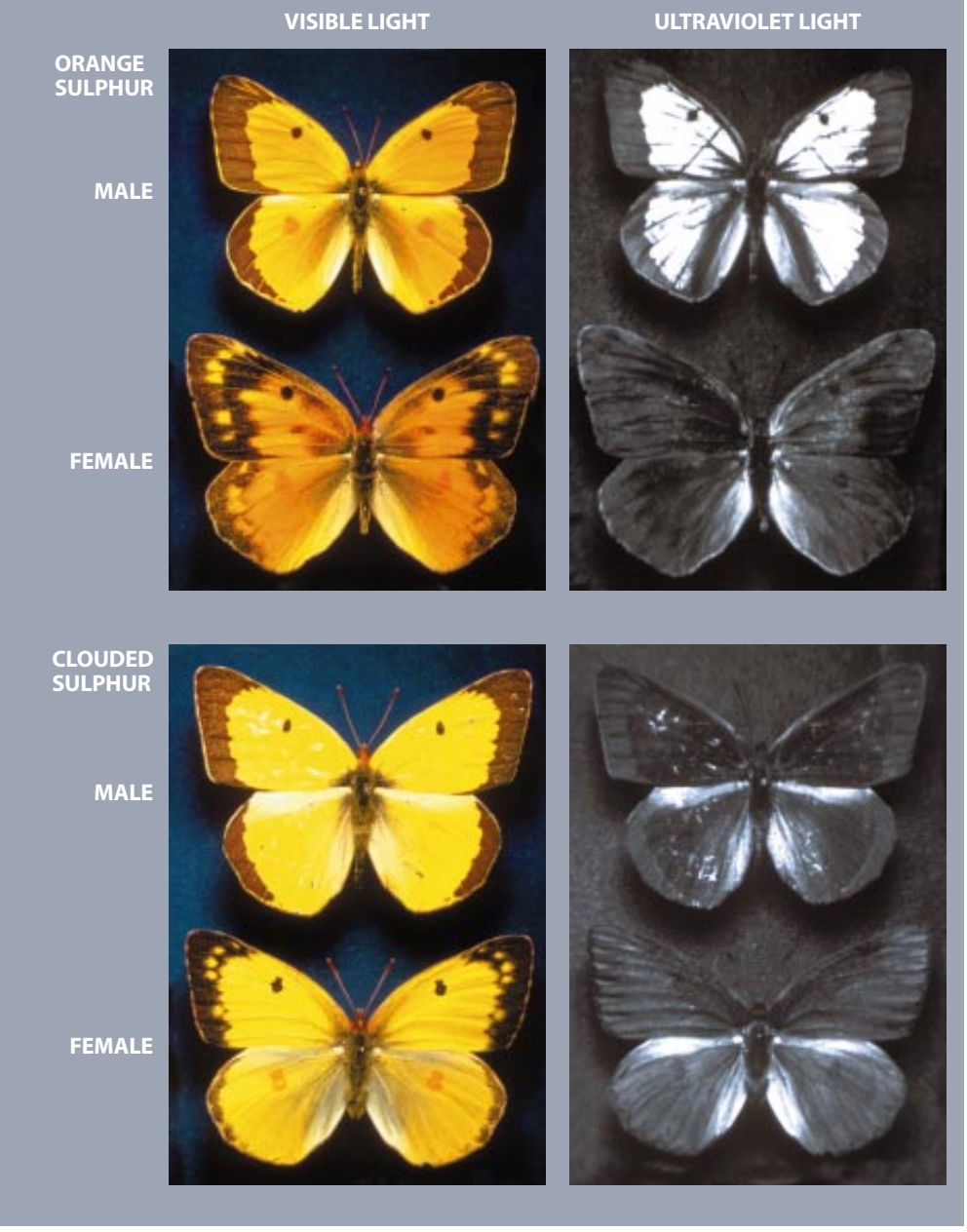
simple behaviors allowed me to develop a test for the cues males use to recognize females. I first glued Little Yellow wings to cards and presented them to males. Males landed on, and even attempted to copulate with, female wings. But male study subjects paid scant attention to male wings similarly mounted.

The next phase of the experiment showed that color was responsible for this choice. I prepared a card with two sets of male wings. A quartz slide that transmits both visible and ultraviolet light covered one set of wings, and a filter that blocks ultraviolet wavelengths overlaid the other. Males now attempted to mate with the male wings under the filter—wings that appeared to be female. The late Robert E. Silberglied and Orley R. "Chip" Taylor, Jr., now at the University of Kansas, got similar results in their study of the Orange Sulphur (*Colias eurytheme*). This species displays a sexual difference in ultraviolet reflectance similar to that in the Little Yellow, and after a male's ultraviolet reflectance is obliterated other males treat him like a female.

Color also can influence mate recognition by females. My research group at Arizona State University took advantage of a dense population of a species known as the Checkered White, *Pontia protodice*, in a rural area near Phoenix to study this phenomenon. We focused on a well-known tendency among virgin females (as well as those who have not mated recently) to approach and chase males occasionally.

We captured Checkered Whites of both sexes and tethered them by tying one end of a thread around the waist between the thorax and abdomen and the other end to a length of stiff wire. We then used this wire like a fishing pole to display the captive butterflies in sight of females in the field. These free females often took off after the tethered butterflies. Their chases after tethered females halted quickly, whereas they showed far greater perseverance toward the males.

As with Little Yellows, male and female members of this species appear quite different in the ultraviolet wavelength, but in the opposite direction. Females reflect ultraviolet light, but the wings of male Checkered Whites con-



PHOTOGRAPHS BY ORLEY R. "CHIP" TAYLOR, JR. *University of Kansas*

ULTRAVIOLET REFLECTANCE of the male Orange Sulphur (*top right*) deters other males and is attractive to females; scales on the wings of the females do not reflect ultraviolet. This difference enables the insects, which can detect ultraviolet wavelengths, to distinguish members of the opposite sex. In the related Clouded Sulphur (*C. philodice*), neither sex reflects ultraviolet light, and the creatures use different cues to identify potential mates, probably chemical compounds emitted by special scales on the males' wings.

tain an ultraviolet-absorbing pigment. This pigment is easily extracted, however, by dipping the wings in a dilute ammonia solution. Such treatment made male wings reflective of ultraviolet, like female wings, without altering any other markings.

I built lifelike models from ammonia-treated wings and then, using stiff angling wire, presented the specimens to butterflies in the field. Females ignored the ultraviolet-reflective male wings—but males became greatly intrigued. Clearly, both female and male Checkered Whites make use of sexual differences in color in order to discriminate potential mates from individuals of their own sex.

Some female butterflies are also picky about color when choosing a mate from among many suitors. Diane C. Wiernasz of the University of Houston investigated this behavior in the Western White, *P. occidentalis,* a butterfly closely related to the Checkered White. She released virgin females into a field and captured males that successfully courted them. These males had darker markings at the tips of their forewings than did rejected suitors. And Wiernasz was able to make males unattractive to virgin females by using white paint to reduce the size of the crucial dark markings. This is the only study of its kind that we have, but it demonstrates that some females discriminate among males on the basis of subtle differences in color.

Females that prefer colorful males may be rewarded with the youngest and healthiest mates. To test this idea, my group and I spent a hot, humid summer with Orange Sulphur butterflies in Arizona alfalfa fields. Studies from the 1970s had shown that female Orange Sulphurs find the ultraviolet reflectance of male wings attractive—but as a male's wings lose scales with age, his ultraviolet color diminishes. We wondered if aging reduces a male's seductive charms. Our suspicions were confirmed when we found that virgin females indeed preferred males with intact wings to males with worn wings—a choice apparently driven by color, ensuring a younger mate.

Good Chemistry

Once a male and a female butterfly have noticed one another, courtship begins in earnest. The male's goal is to induce the female to alight and remain still for mating, which sometimes lasts an hour or more. In some species the female must also move her abdomen out from between her hindwings to grant the male access. Butterfly biologists have studied the ritual that precedes actual copulation in only a few dozen of the roughly 12,000 species of butterfly, but it seems clear that, for butterflies, what humans might think of as scent can be a language of love. The vocabulary of this language is chemical.

The best-understood case of nonvisual butterfly communication involves the Queen butterfly, *Danaus gilippus*. Males of this species produce pheromones, compounds designed to elicit specific reactions—of sexual interest in this case—from other butterflies. These pheromones disseminate from brushlike structures, called hair pencils, found at the end of the abdomen in males only. Hair pencils have a particularly large surface area for their small volume and are thus highly efficient at distributing chemicals. As a male flies up and down in front of a female, he touches her antennae with

his protruding hair pencils, thereby depositing pheromones. The female responds to this chemical signal by alighting and remaining still while the male copulates with her.

Many species of butterfly probably use pheromones in courtship. Males often possess features reminiscent of the Queen's hair pencils, such as patches of unusual scales on the wings and brushlike structures on the thorax. Like hair pencils, these scales and hairs have large relative surface areas that would presumably enhance pheromone distribution. And for the family of butterflies classified as Sulphurs, special scales on the male's generally bright yellow or orange wings do indeed emit compounds that may affect female behavior.

Some species of butterfly have evolved ritualistic courtship displays that could expose females to male pheromones, just as the up-and-down flight of the Queen allows hair pencils to touch antennae. A male Grayling, *Hipparchia semele,* for example, will alight directly in front of a female and catch her antennae between his wings. He bows slowly forward, rubbing the female's antennae against a patch of scales suspected of carrying pheromones. The male Barred Sulphur, *Eurema daira,* perches next to a female and waves his forewing up and down, dragging the edge of his wing along her antennae with each sweep. The male Gulf Fritillary, *Agraulis vanillae,* sits next to a female and claps his wings open and closed; the female's nearest antenna is often caught between the male's wings, where it touches brushlike scales.

Elaborate interactions such as these are not the norm in the butterfly world, however. In fact, courtship in most species is fleeting—lasting less than 30 seconds and consisting mostly of the male fluttering about the female. A more representative courtship may be that of the Little Yellow, in which the male buffets the female for a few seconds before alighting and attempting to copulate. This simple activity may be sufficient to waft pheromones onto the female's antennae, making her agreeable to mating.

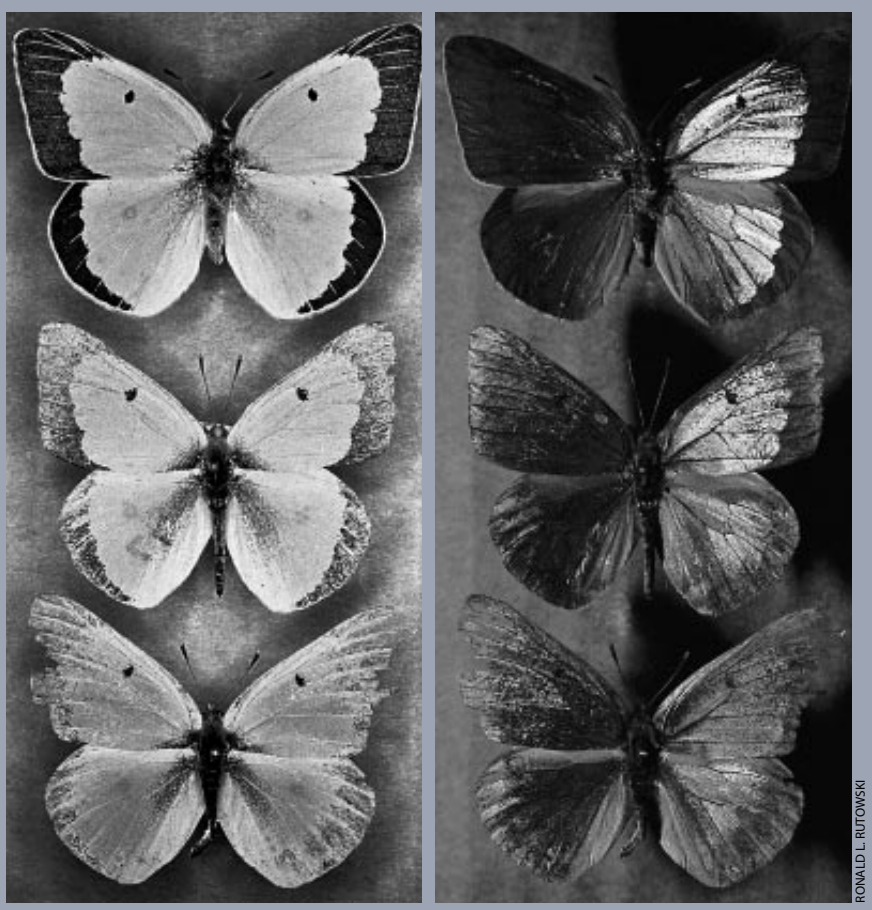
Despite the charming nature and eager efforts of the male, some females remain indifferent to any of these attentions. Females that have recently mated successfully can be most obstinate. These females will take defensive measures to discourage an unwanted suitor. If perched, they will flap their wings rapidly; if flying, they will flee, sometimes shooting dozens of feet upward in a maneuver called ascending flight. If the spurned male is persistent, the resulting aerial courtship can last several minutes. Just as a tale of dramatic conflict may be more compelling than one of tiresome harmony, these conspicuous rejections often attract more butterfly watchers than do the more fleeting courtships that lead to mating.

Location, Location, Location

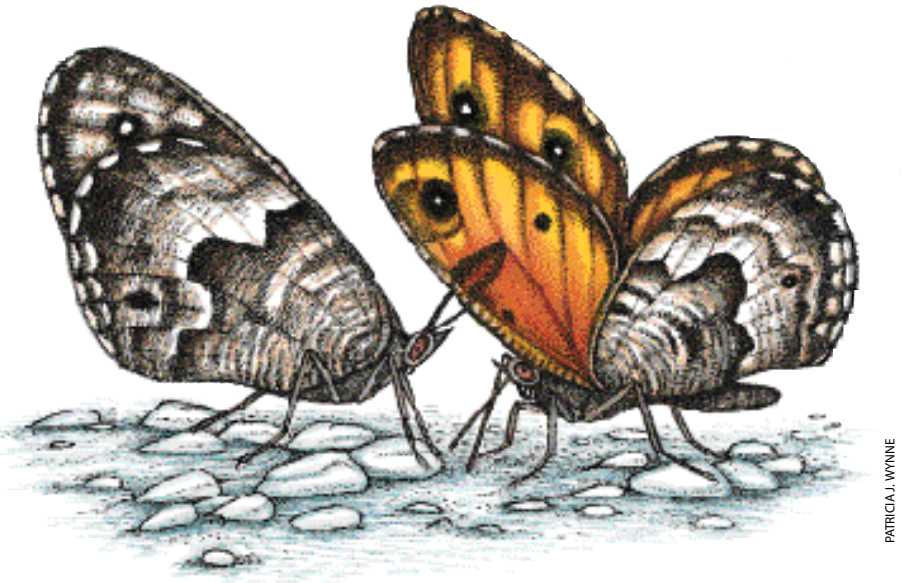
Gaudy wings, smooth moves and pheromones do a male butterfly no good if he cannot find a female butterfly on whom to practice his seduction. Males of many butterfly species adopt a search-on-the-fly strategy, wandering the landscape looking for mates. Often they investigate likely areas, such as plants where females tend to lay their eggs or sites where virgin butterflies emerge from their cocoons.

Males of the Empress Leilia species, *Asterocampa leilia,* however, use a highly systematic approach. Because that species' larvae feed and pupate on desert hackberry trees and because the females mate but once in their short lives, the males actually stake out that vegetation in search of young virgins. A few hours after dawn, just when the females emerge from their cocoons and become ready to fly for the first time, the males begin their watch.

Early in the day the males perch on the ground in open, sunny spaces near the trees. This early-morning sunbathing probably allows them to keep an eye out for other butterflies while keeping their bodies warm enough to give chase. (Because they cannot regulate body temperature internally, butterflies grow sluggish if the ambient temperature is too cold.) Later in the morning the males move up into the trees to exactly the average plane of flight of Empress Leilia females, about one meter above the ground. My students and I have ob-



LOSS OF SCALES WITH AGE (*top to bottom*) diminishes the ultraviolet reflectance of the male Orange Sulphur and renders him less attractive to females.



PATRICIA J. WYNNE

BOWING DISPLAY of the Grayling (*Hipparchia semele*) brings a female's antennae into contact with brushlike scales on the male's wings. These scales may produce chemicals that induce the female to accept his advances.

served that even when the male butterflies are perched at a tilt, they hold their heads so that their eyes are looking horizontally out of the tree. This orientation seems to ensure that their area of greatest visual acuity—which lies in a band at the equator of the visual field—coincides with the plane of likeliest female flight.

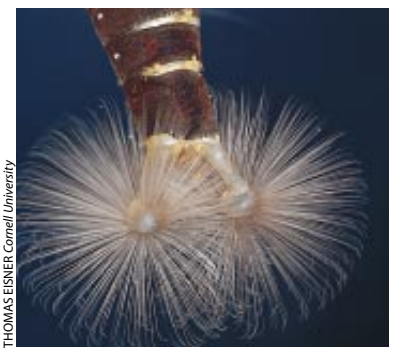
Male Empress Leilias guard their territory jealously for an hour or two. They will take off after any approaching object, whether butterfly, bird or tossed rock. These vigilant males pursue females or chase away encroaching males before returning to the same perch. Many species of butterfly show even greater territoriality, laying claim to mating grounds characterized more by geography than by resources such as hackberry trees: bare spaces, sunny spots, ravines and gullies, and especially hilltops.

We can only guess why male butterflies (and, indeed, other insects) seek these territories. Sunny patches may attract females to stop and warm themselves; a bare spot might make a good vantage point for visual contact. Among the most intriguing of territories is the hilltop. The virgin females (but not mated females) in some species do tend to fly uphill, but the riddle of cause and effect in the evolution of hill-topping behavior remains unsolved.

The elements of male butterfly courtship, from attractive wing color to enticing pheromones to auspicious environs, seem geared toward ensuring as many successful matings as possible. Even a male's preference for young females has a logical basis, as the youngest females have a better statistical chance to survive long enough to produce many of his offspring.

For males, a strong imperative, related to the impulse to send their genetic material into the next generation, is to prevent their mate from mating again. Male butterflies actually make a substantial contribution to females during copulation, passing along a large quantity of nutrients. This nutrient store, called the spermatophore, can be as much as 6 to 10 percent of the male's body weight; a male cannot afford such an investment in a female who will use his competitor's sperm to fertilize her eggs [see "Glandular Gifts," by Darryl T. Gwynne; Scientific American, August 1997]. In fact, evolution has come

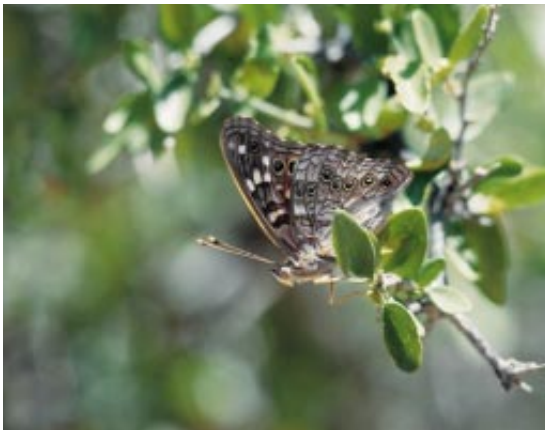
PATRICIA J. WYNNE



THOMAS EISNER *Cornell University*

ATTRACTIVE CHEMICALS are disseminated by brushlike structures called hair pencils on the male Queen butterfly (*Danaus gilippus*) (*top butterfly at far left*). These pheromones are produced from chemical precursors that males obtain by sucking at plants such as Crotalaria (*below*).

RONALD L. RUTOWSKI



PHOTOGRAPHS BY RONALD L. RUTOWSKI

HIGHLY SYSTEMATIC APPROACH for finding mates is adopted by Empress Leilia (*Asterocampa leilia*) males: they stake out the hackberry trees where females are likely to appear newly emerged from the pupal stage or later to lay eggs. Early in the morning males perch on the ground in a sunny spot where they can both keep a lookout and warm up (*above*). Eventually they move into the trees (*right, top* and *bottom*)—to exactly the typical height of the flight of the females.

up with a mechanism that favors the male that has succeeded in mating first. The presence of the spermatophore in the female's reproductive tract causes her to be unresponsive to further sexual advances. Experimental evidence supports this conclusion: artificially filling a virgin's reproductive tract renders her uninterested in mating, while cutting the nerves to this area in a mated female restores her sexual interest. Another male technique for barring other suitors from his mate is less elegant—he leaves a plug that obstructs the reproductive tract.

Females face different evolutionary pressures. They often get but one chance to mate and must therefore be highly selective. By accepting only the fittest male, a female can assure her own offspring a quality genetic endowment, and she might also secure for herself a more generous spermatophore—which most likely helps her live longer and, in turn, lay more of her eggs. Male colors, pheromones and displays may allow females to judge a suitor's overall fitness and success in life. We suspect that chemical signals indicate the quality of a male's diet: the crucial mating pheromone of male Queen butterflies, for instance, is produced only when the males have fed at certain plants. And vibrant colors can signal younger, healthier individuals.

As with human beings, some of the attributes and behaviors of butterfly courtship are quite elaborate, whereas others are fairly pedestrian. Intricate or simple, courtship and mating remain the mechanism by which survival and evolution take place. Whether a butterfly watcher takes in a swarming colony of Monarchs mating in the mountains of central Mexico or a dalliance between two alfalfa butterflies in a backyard, the observer is fortunate enough to be watching the results of, and the continuing course of, evolution. SA

The Author

RONALD L. RUTOWSKI has studied butterfly mating behavior for almost 25 years. After receiving his Ph.D. at Cornell University in 1976, he joined the faculty of Arizona State University, where he is a professor and co-director of the Biology and Society Program in the department of biology. When not chasing butterflies, he enjoys playing the violin, making beer, and bicycling.

Further Reading

THE EVOLUTION OF INSECT MATING SYSTEMS. R. Thornhill and J. Alcock. Harvard University Press, 1983.

THE BIOLOGY OF BUTTERFLIES. Edited by P. Ackery and R. Vane-Wright. Academic Press, 1984.

THE DEVELOPMENT AND EVOLUTION OF BUTTERFLY WING PATTERNS. H. F. Nijhout. Smithsonian Institution Press, 1991.

THE EVOLUTION OF MATE-LOCATING BEHAVIOR IN BUTTERFLIES. R. L. Rutowski in *American Naturalist,* Vol. 138, No. 5, pages 1121–1139; November 1991.

QUESTIONS ABOUT BUTTERFLY BEHAVIOR: THE CASE OF THE EMPRESS LEILIA. Ronald L. Rutowski in *American Butterflies,* Vol. 2, pages 20–23; May 1994.

SEXUAL DIMORPHISM, MATING SYSTEMS, AND ECOLOGY IN BUTTERFLIES. R. L. Rutowski in *Evolution of Mating Systems in Insects and Arachnids.* Edited by Jae C. Choe and Bernard J. Crespi. Cambridge University Press, 1997.

Léon Foucault

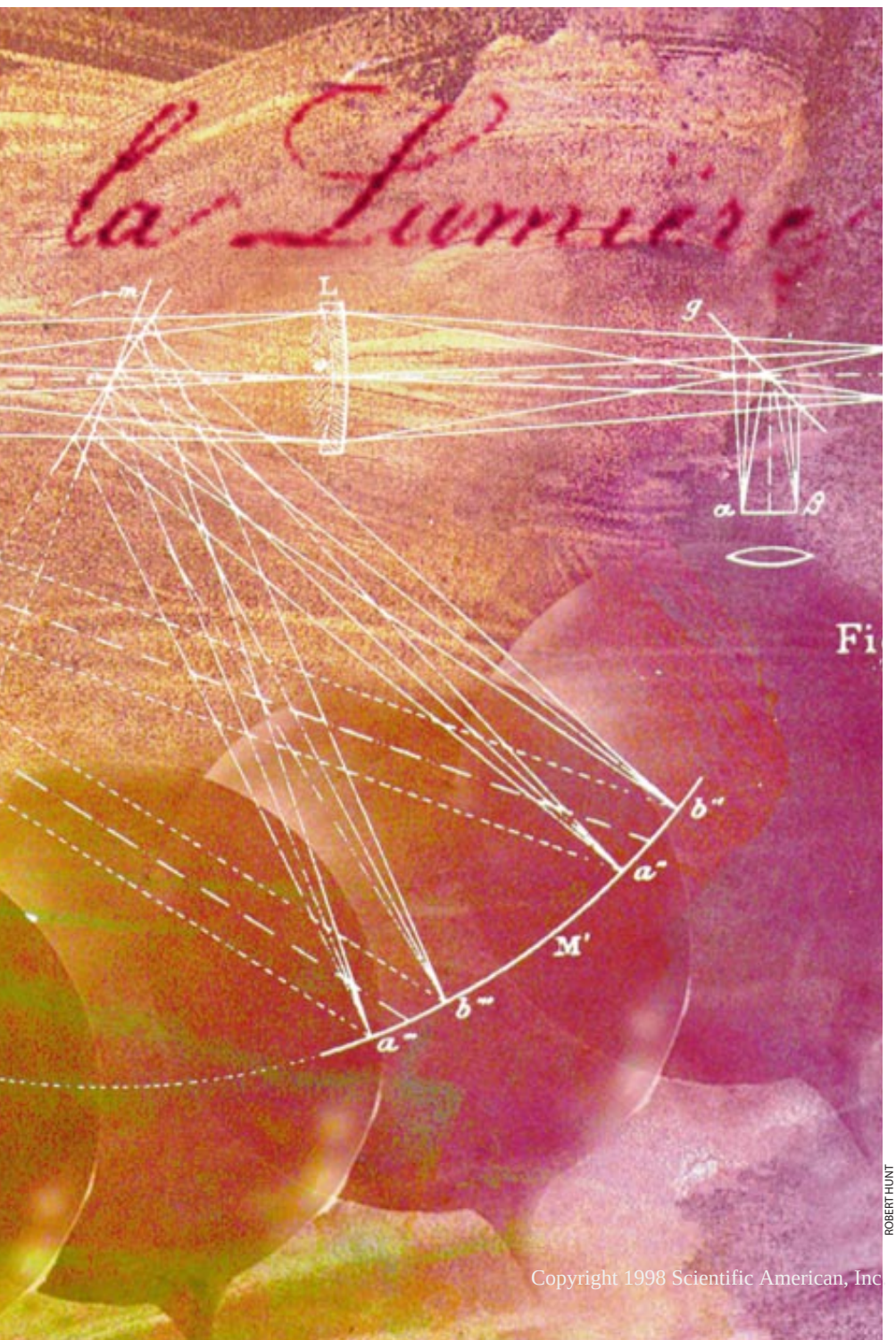
Celebrated for his pendulum experiment in 1851, Foucault also produced decisive evidence against the particle theory of light, invented the gyroscope, perfected the reflecting telescope and measured the sun's distance

by William Tobin

In the mid-1800s, scholars knew that the earth turned on its axis, but one important, anticipated proof was missing. Since the times of Galileo and Isaac Newton, people were aware that the rotation of the earth should influence the motion of moving bodies (as seen from observers on the ground), just as the spinning of a carousel affects a child's jump from one wooden horse to another. But for more than 200 years, natural philosophers had been trying to obtain evidence of this phenomenon. They dropped weights from towers or down wells. They even considered firing cannonballs straight upward in hopes of detecting slight deviations of the flight paths of those moving objects from the vertical. Subtle deflections would have offered further proof of the earth's rotation. Yet all their efforts were in vain.

Then, in 1851, a young French physicist stunned the world. Léon Foucault showed that a simple pendulum, through its continuous periodic swinging, would accumulate the tiny effects of the earth's rotation. Witnessed by observers at the Panthéon in Paris, the plane of motion of Foucault's pendulum veered slowly clockwise, revealing the counterclockwise rotation of the ground below. The experiment caused a sensation. Here, finally, was direct dynamical proof that the earth indeed spun on its axis.

Although posterity remembers Foucault for his pendulum, he made many other notable contributions to 19th-century science. He provided incontestable evidence against the particle theory of light, invented the gyroscope (which supplied additional proof of the earth's rotation), perfected the reflecting telescope and measured the solar distance. His work, a fascinating mixture of the pure and the applied, merits examination because it is characterized by extraordinary precision and insight. Interestingly, whereas some of his peers thought him the preeminent French physicist of the time, others considered him little more than a gifted tinkerer.



ROBERT HUNT

LÉON FOUCAULT (1819–1868) was responsible for many significant advances in 19th-century science. Curiously, although some of his peers considered him the greatest French physicist of that era, critics thought him a dilettante.

Virtuoso Hands

Jean-Bernard-Léon Foucault was a Parisian. He was born in Paris in 1819 and, apart from a few childhood years in Brittany, lived his entire life in the city, dying there in 1868. His father was a publisher and bookseller noted for numerous volumes of French history. The Foucaults were well-off. The family owned several properties and was able to send young Léon to the Collège Stanislas, a prestigious secondary school. It seems that the adolescent Foucault, described as a delicate youth, was a mediocre student who required a private tutor.

Nevertheless, Foucault had a great gift: his hands. He filled his teenage leisure hours by constructing model boats, telegraphs and steam engines. Hoping to capitalize on his dexterity, he entered the Paris medical school in 1839 with plans of becoming a surgeon. But the sight of blood revolted him, and he soon renounced medicine.

Around this time, Foucault became fascinated with daguerreotypy. In this first practical form of photography, images were recorded on iodized, silver-covered plates and developed with mercury

vapor. Though revolutionary, the process had one drawback: exposures could easily take an hour. But in 1841 Hippolyte Fizeau, a classmate of Foucault from the Collège Stanislas, discovered that bromine sensitized the plates, and Foucault developed a technique for applying the toxic vapor uniformly. The improvement dramatically reduced the exposure time, making portraiture feasible.

Foucault and Fizeau had much in common: they were just five days apart in age, and both would abandon medical studies for careers in physics. Soon after their work with bromine, they began collaborating. In 1844, at the request of François Arago, secretary of the French Academy of Sciences and director of the Paris Observatory, they took the first successful daguerreotype of the sun. The image clearly showed the solar disk to be brighter at the center than at the edges, confirming visual observations and refuting the suggestion by Dutch astronomer and physicist Christian Huygens that the sun was a liquid ball.

Foucault also investigated daguerreotypy for medical applications, working with Alfred Donné, who had been one of his medical lecturers. Donné was studying milk and other bodily fluids with the microscope. In 1845 their atlas of 80 micrographs, most of which Foucault took, marked the first medical use of the daguerreotype.

That same year Foucault succeeded Donné in reporting the Monday meetings of the Academy of Sciences for the *Journal des Débats*, an influential Paris newspaper. In his articles Donné had attacked Arago for turning the academy into a personal fiefdom. Foucault's writing was usually less outspoken, but his criticisms were sometimes sharp. He would pen these columns with varying regularity for more than 15 years.

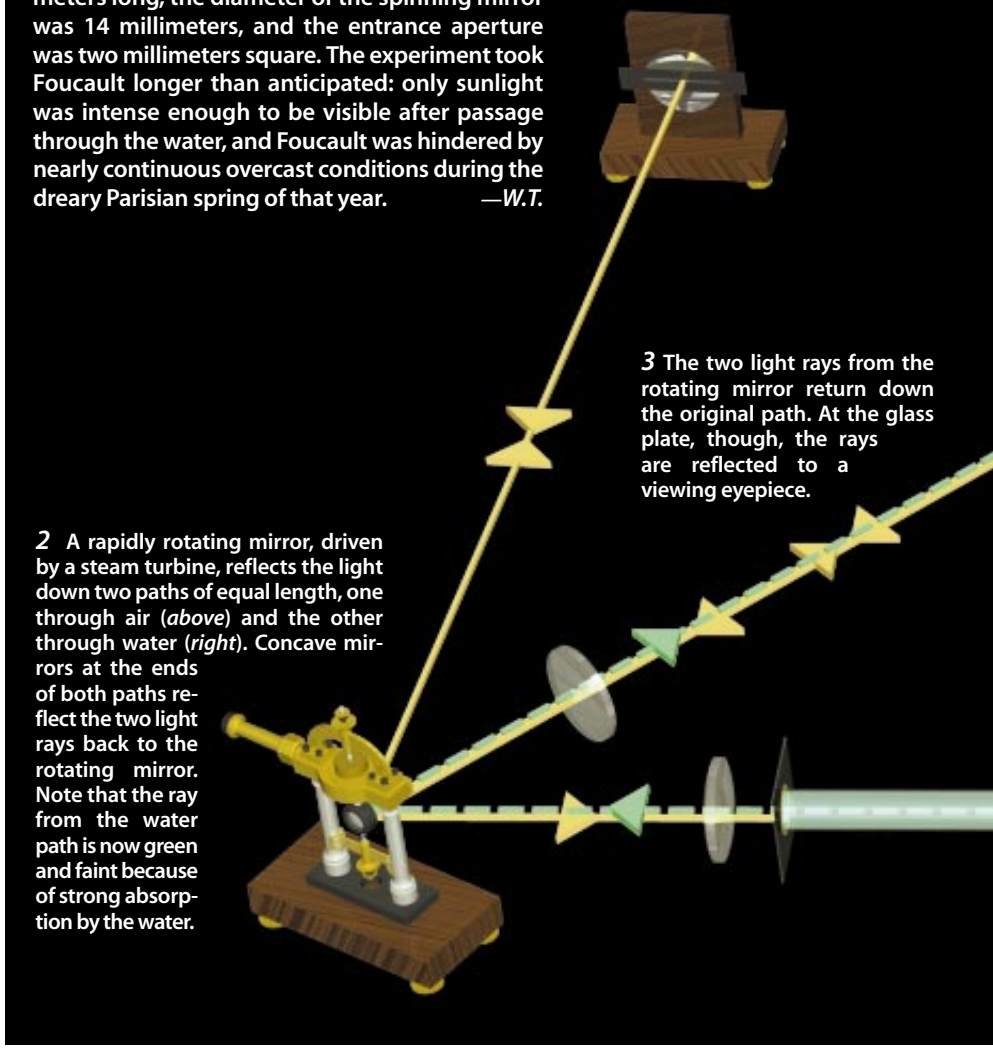
Waves versus Particles

Foucault's first major discovery came in 1850. His work provided indisputable evidence in the long-standing debate over the nature of light. Since the early 1800s the wave theory of Robert Hooke and Huygens had been gaining favor over the particle hypothesis of René Descartes and Newton. The discoveries of interference and polarization had supplied suggestive evidence for light waves. But corpuscularists always managed to patch up their theory to agree with experimental results.

A decisive test was needed, and Arago

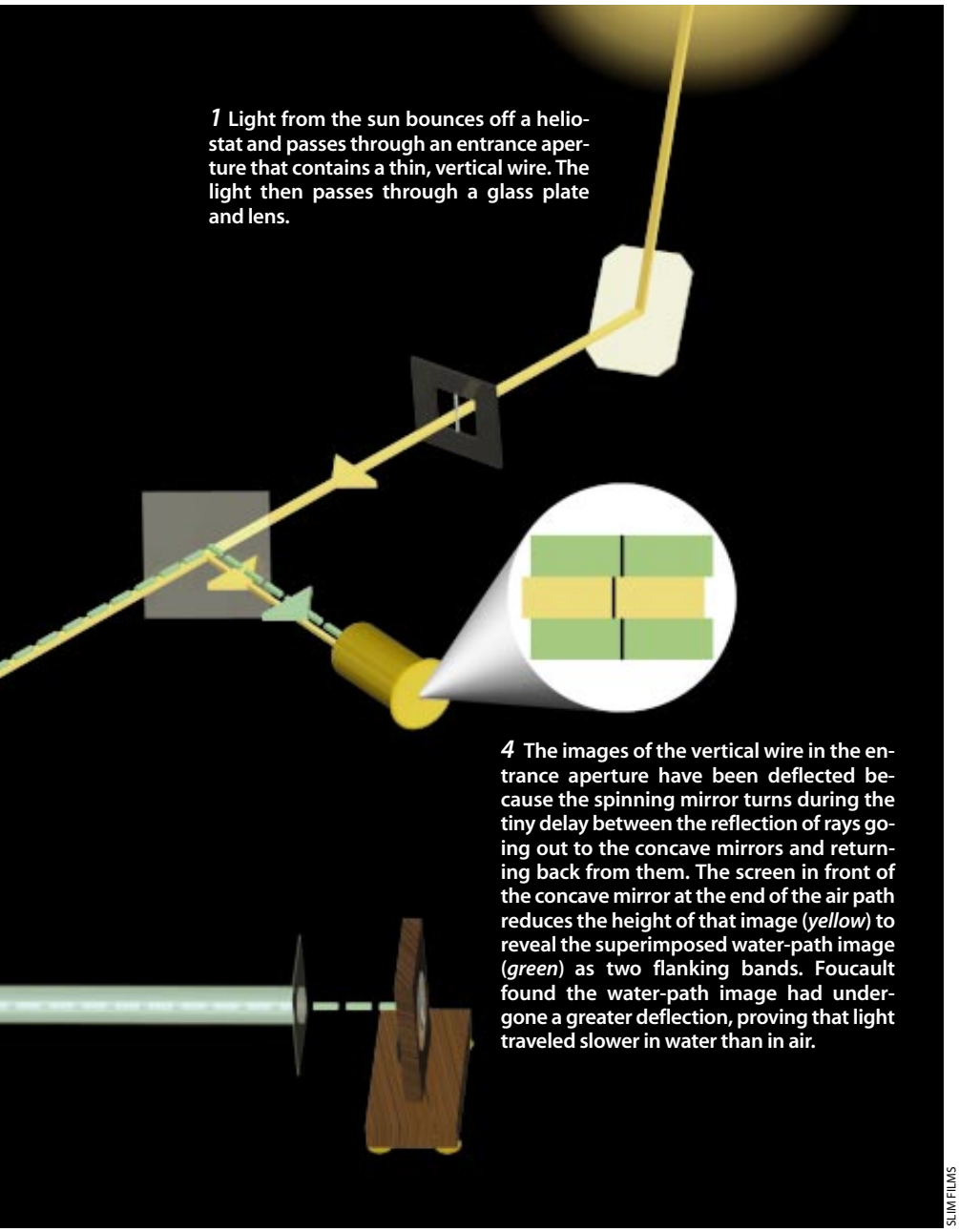
Air-Water Experiment

In 1850 Léon Foucault proved that light traveled faster in air than in water, in stark disagreement with corpuscular theory. The diagram is schematic: in reality, the water tube was three meters long, the diameter of the spinning mirror was 14 millimeters, and the entrance aperture was two millimeters square. The experiment took Foucault longer than anticipated: only sunlight was intense enough to be visible after passage through the water, and Foucault was hindered by nearly continuous overcast conditions during the dreary Parisian spring of that year. —*W.T.*



PARIS OBSERVATORY

ROTATING MIRROR was a crucial apparatus for Foucault. In addition to the experiment depicted here (*diagram above*), he used a spinning mirror in 1862 to measure the speed of light and the distance to the sun. The device shown in this photograph was used in that work. Foucault abandoned steam power and drove the mirror with compressed air from special bellows supplied by his neighbor, a prolific organ builder. The author has blown into the turbine of this device, which is currently at the Paris Observatory. The mirror spindle is still beautifully balanced and turns as smoothly as it must have for Foucault.



SLIM FILMS

had devised one. When light passes from air into water, the rays refract, or bend, at the boundary between the materials. The two theories explained this phenomenon in very different ways. According to the particle theory, light corpuscles were accelerated by the interface, and so they traveled faster in water than in air. The wave hypothesis, requiring wavefront continuity, predicted that the water speed was less. Thus, the nature of light could be settled unequivocally—or so Arago thought—by comparing the velocities of light in air and water.

The problem, of course, was the tremendous speed of light. Arago's concept called for parallel air and water paths along which light from a single source was shone. At the end of the paths, a rapidly spinning mirror would bounce the two beams. The slower beam, arriving later, would be reflected at a greater angle. Arago tried the test with a clockwork-driven mirror that spun up to 2,000 times per second, but he had difficulty resolving the fast-moving reflections. Soon after, diabetes dimmed Arago's eyesight, and he was forced to cease his experiments.

Foucault and Fizeau picked up the reins. They realized that the reflections could be made stationary—and observable—if concave mirrors were used to return the rays to an additional reflection at the rotating mirror and thence to a fixed eyepiece, as shown in the box on these two pages. But at this juncture, the two men, after collaborating so successfully for almost a decade, had a falling-out. It is unclear why. Possibly they argued over how best to drive the rotating mirror. Foucault's somewhat rigid character might not have helped. Donné would later write, "Foucault was not what one would call lovable: he had neither the suppleness of character nor the desire to please necessary for one to be considered agreeable by the world." Whatever the reason, Foucault and Fizeau worked independently, vying to be the first to coax a result from Arago's idea. Fizeau continued using the clockwork mirror, whereas Foucault relied on a little steam turbine to rotate his mirror 800 times per second.

Foucault won the race, showing in April 1850 that water slowed the passage of light. Seven weeks later Fizeau confirmed the result. The corpuscular theory was dead, and Foucault's experiment won him widespread acclaim in addition to France's Legion of Honor.

Physicists, of course, would later realize that Foucault's experiment was not quite as crucial as it then seemed. The demolition of corpuscular theory did not mean that waves provided a full description of light's behavior. Quantum mechanics would be needed to explain such phenomena as the photoelectric effect, black-body radiation and lasing.

Inspired by a Lathe

Foucault's next accomplishment was his most renowned. Inspiration for the experiment struck when he happened to twang a steel rod that was clamped in the chuck of a lathe. He noticed that although the rod rotated with the chuck, the plane of vibration did not. This effect, which even today causes surprise [*see top illustration on next page*], arises because inertia keeps objects in the same state of motion unless they are disturbed by an external force. Perceiving that the chuck and the rod might be analogous to the earth and a pendulum, Foucault deduced that the rotation of the earth would manifest itself to observers on the ground via a slow veering of the swing plane of the pendulum.

Working in the cellar of the house he shared with his mother, Foucault constructed his first pendulum with a five-kilogram (11-pound) brass bob attached to a two-meter (six-and-a-half-foot) wire. During an initial trial, on January 3, 1851, the wire snapped. But five days

later Foucault succeeded: the swing plane of his pendulum turned.

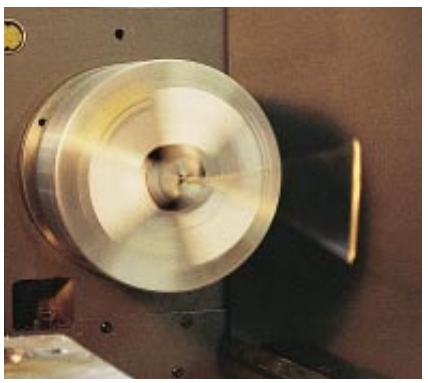
At Arago's invitation, Foucault installed a pendulum with an 11-meter wire at the Paris Observatory. Announcing his discovery to the Academy of Sciences a few days later, Foucault asserted that the pendulum's swing plane would seem to veer by 360 degrees a day at the earth's geographic poles, whereas elsewhere the rate would be slower by a factor of the sine of the location's latitude. (Thus, there would be no veering at the equator.) To derive the sine factor, Foucault wrote, "one must resort either to analysis or to considerations of mechanics and geometry which are outside the limits of this note."

With the support of the president of the Republic, Louis-Napoléon Bonaparte (later the emperor Napoléon III), the pendulum was scaled up once more for a spectacular public demonstration. During the spring of 1851, Parisians flocked to see a 28-kilogram bob swinging on a 67-meter wire under the lofty and elegant dome of the Panthéon. The advance of the swing plane of the pendulum was made apparent by a stylus, attached to the underside of the bob, which plowed through piles of damp sand at each extremity of the swing. Air resistance gradually damped the oscillation, and the pendulum needed to be restarted after five to six hours. But during that interval, the swing plane had veered the expected 60 to 70 degrees clockwise. The experiment was soon repeated around the world. In Rio de Janeiro, on the other side of the equator, the swing plane moved counterclockwise, as expected.

Foucault was wise not to try to justify the sine factor publicly. In the subsequent months, dozens of savants attempted to explain the pendulum's motion either geometrically or analytically. In general, proponents of one approach hotly contested the interpretations of partisans of the other. The sine factor arises because the pendulum is not completely free to swing in space: except at the geographic poles, its point of suspension sweeps out a circle, and the direction of gravity forever changes as the earth rotates. The consequence is that the pendulum follows the stars [*see box on opposite page*]. In analytical terms, the veering results from the Coriolis force that is perceived in a rotating frame—the same force that results in a sideways knock to the child leaping around a carousel.

For Foucault, the pendulum's veering posed no philosophical problem: it indicated the absolute nature of space. Decades later the Moravian physicist Ernst Mach would challenge this view by speculating that it was the distant stars that both caused inertia and established the unaccelerated, or "inertial," frames of reference against which accelerations and rotations could be gauged. It would not be possible, Mach asserted, for a planet to rotate in an otherwise empty universe, and a Foucault pendulum on such a planet would not veer.

Mach's ideas influenced Albert Einstein considerably, but the general theory of relativity takes a middle course. Its inertial frames do not extend indefinitely across space (they are valid only locally),

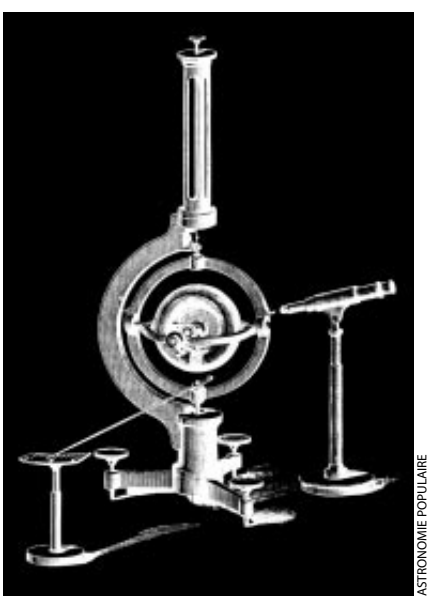


WILLIAM TOBIN

STUDIO CNAM; MUSÉE DES ARTES ET MÉTIERS-PARIS

SMITHSONIAN INSTITUTION

SMITHSONIAN INSTITUTION



ASTRONOMIE POPULAIRE

ROTATION OF THE EARTH occupied Foucault's attention in 1851–1852. The idea for his pendulum came to him when he observed a twanged rod gripped in the chuck of a spinning lathe (*top*). The rod rotates with the chuck, but the plane of its vibration stays fixed because no force acts that can move it. Foucault, who was only 31 years old at the time (*middle*), inferred that the lathe and rod would be analogous to the earth and a pendulum (*opposite page*). For further proof of the earth's rotation, Foucault invented the gyroscope (*bottom left* and *middle*). The slow drift of the rotor could be observed with a microscope (*bottom right*).

but within them rotation is absolute. According to general relativity, though, a spinning body such as a planet will tug the inertial frame around very slightly, and a Foucault pendulum should not be so firmly locked to the distant stars. At the earth's poles, this frame dragging—also known as the Lense-Thirring effect—amounts to about one fifth of an arc second a year and may recently have been detected by laser range measurements of orbiting satellites.

Because the sine factor caused much confusion and misunderstanding, Foucault worked on demonstrating latitude-independent motion. A year later he produced the necessary device, which he named the gyroscope. The axis of his spinning gyroscope was completely free to maintain its orientation in space, and its slow drift with respect to the earth could be observed with a microscope. Through most of the 20th century, mechanical gyroscopes would give cardinal service in surveying and navigation.

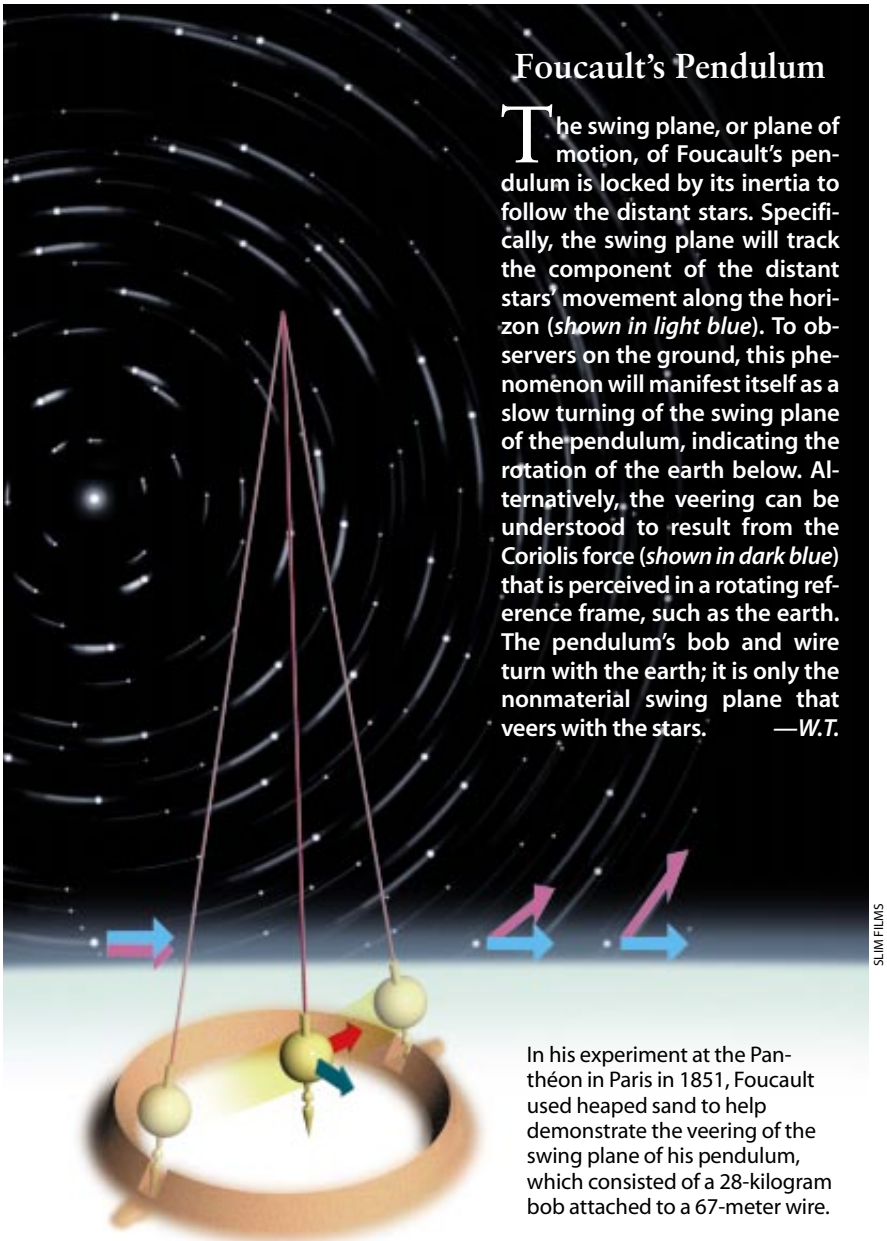
Reflection over Refraction

For all his past success—and despite the fact that he was already in his mid-30s—Foucault had yet to land an official post. Personality may have been a factor. "His airs of a mandarin with three pigtails do not please everyone," noted one contemporary. Nevertheless, in 1855 Foucault was appointed the official physicist of the Paris Observatory, which, following Arago's death, was headed by Urbain Le Verrier, famed for his prediction that had led to the discovery of the planet Neptune. Needing to update the observatory's instruments, Le Verrier asked Foucault to construct a refracting telescope with a giant lens of 74-centimeter diameter. Foucault soon concluded, however, that a reflecting telescope with a large mirror would be a better choice.

At the time, Lord Rosse was building such telescopes in Ireland, but his instruments used metal mirrors that were heavy, difficult to shape and troublesome to maintain because their surfaces tarnished. Glass, as Newton had known, was a more promising mirror material, provided it could be made sufficiently reflective. Fortunately, Foucault was aware that the noxious mercury amalgams applied to the back of looking glasses were being replaced with a layer of silver deposited by the chemical reduction of silver-nitrate solutions. In fact, he had used such a layer on his spinning

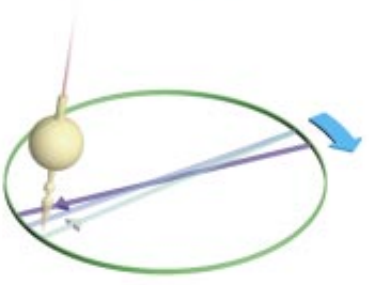
Foucault's Pendulum

The swing plane, or plane of motion, of Foucault's pendulum is locked by its inertia to follow the distant stars. Specifically, the swing plane will track the component of the distant stars' movement along the horizon (*shown in light blue*). To observers on the ground, this phenomenon will manifest itself as a slow turning of the swing plane of the pendulum, indicating the rotation of the earth below. Alternatively, the veering can be understood to result from the Coriolis force (*shown in dark blue*) that is perceived in a rotating reference frame, such as the earth. The pendulum's bob and wire turn with the earth; it is only the nonmaterial swing plane that veers with the stars. —*W.T.*

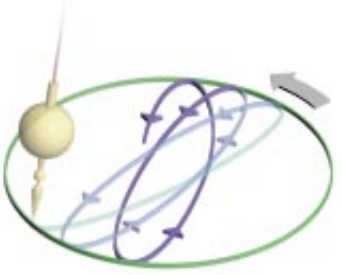


SLIM FILMS

In his experiment at the Panthéon in Paris in 1851, Foucault used heaped sand to help demonstrate the veering of the swing plane of his pendulum, which consisted of a 28-kilogram bob attached to a 67-meter wire.



In the Northern Hemisphere the plane will veer slowly clockwise (*shown greatly exaggerated*), revealing the counterclockwise rotation of the ground below.



The pendulum must be constructed and launched with extreme precision, otherwise the trajectory could loop (*shown greatly exaggerated*), masking the effect of the earth's rotation.

Léon Foucault

mirror for his earlier air-water experiment. The silver layer was very thin and could be applied to the *front* of telescope mirrors without altering their focusing properties. Furthermore, once the silver tarnished, it could be replaced easily without damaging the optical quality of the underlying glass surface. The metalized-glass reflecting telescope had become a technological possibility.

Foucault made satisfactory 10- and 22-centimeter reflecting telescopes, but the then hit-or-miss methods of polishing optical surfaces were inadequate for larger diameters. After failing at five separate attempts to fabricate a 42-centimeter mirror, Foucault devised three techniques for examining optical surfaces. The most powerful was the knife-edge test: a sharpened blade partly eclipses the reflected image of a point source in such a way that any polishing flaws are shown in exaggerated relief, thus enabling the optician to identify and eliminate the imperfections.

Foucault's biggest telescope had an 80-centimeter mirror and was installed at the Marseilles Observatory under the clear Mediterranean skies. The path was now paved for the giant reflecting telescopes of the future.

WILLIAM TOBIN

Measuring the Solar Distance

For Le Verrier, the discovery of Neptune was only a preface to his life's work—the Newtonian analysis of the motions within the solar system. Through heroic calculations Le Verrier refined the estimates of planetary masses. He also deduced that the distance to the sun was about 3 percent smaller than previously believed. At that time, the physical value of the speed of light was derived solely from astronomical measurements, one of which was the solar distance. Consequently, Le Verrier estimated that the true speed of light in a vacuum was also about 3 percent less than the then best estimate of 308,000 kilometers per second.

For help in proving this prediction, Le Verrier again turned to Foucault and asked the observatory physicist to convert his 1850 differential speed-of-light experiment into an absolute one. To do so, Foucault extended the light path to 20 meters by using a relay of five concave, front-silvered mirrors. A clock-driven rotating toothed wheel helped to set the spinning mirror stroboscopically to precisely 400 turns per second. Foucault's result, announced to the Academy of Sciences in September 1862, was 298,000 kilometers per second, in agreement with Le Verrier's prediction and within 1 percent of the modern value.

Physics textbooks today sometimes refer to Foucault's measurement, mainly because the method was later adopted by Albert A. Michelson in the U.S. But Foucault's purpose for doing his experiment is often misrepresented. At the time, people did not know that the speed of light was a physical constant of prime importance. (James Clerk Maxwell's electromagnetic theory of light was not published until 1865, and Einstein had yet to be born.) The sole issue in 1862 was the sun's distance, deemed by Le Verrier the "most critical question of modern astronomy."

The speed-of-light experiment was Foucault's last project of consequence. The remainder of his life was devoted mostly to committee work and to a return to an earlier interest—mechanical governors. He was seeking a universal regulator that would make his fortune. But success proved elusive: industrial requirements were too varied, and his designs were dynamically unstable.

The Life of a Cat

How to summarize Foucault the scientist? "Not everyone considered Foucault to be a true physicist," Donné wrote, "…because Foucault had not studied all branches of physics, [and] he would possibly not have been able to

WILLIAM TOBIN

TELESCOPE with a 20-centimeter-diameter mirror (*left*) was made by Foucault using a specialized process of applying a thin layer of silver onto the front of a glass mirror. He developed techniques that enabled the fabrication of much larger mirrors, opening the way for the giant reflecting telescopes of the future. Foucault's largest telescope (*above*) boasted an 80-centimeter silvered-glass mirror and was completed in Paris in early 1862. It was soon transferred to the clearer southern skies of Marseilles Observatory, where it would serve for 100 years. Recently the telescope was classified as a French historical monument.

give a complete course in elementary physics.... For many, he was an amateur. Indeed, Foucault himself used this word, glorifying in it: 'We are amateurs..., amateurs in the true sense..., we devote ourselves to those aspects of science towards which our instinct leads us.' "

Foucault's strength, in two words, was *experimental precision:* even today, building a Foucault pendulum is no trivial matter. Typically, the veering occurs at only $^{1}/_{20,000}$ the rate of swinging. Indeed, the motion of the swing plane of many a purported Foucault pendulum results not from the earth's rotation but from sloppiness that manifests itself in a looping trajectory.

London's Royal Society was one of many learned organizations that recognized Foucault's merits when it awarded him its most prestigious honor, the Copley Medal, in 1855. But recognition from his peers at home did not come so easily. In France the scientific tradition was analytical and formal, and although Foucault had astonishing mechanical intuition and insight, his mathematical skills were poor. His doctoral thesis on his 1850 air-water measurements contained errors of algebra that he was incapable of correcting during the oral examination.

Nor was Foucault always able to reach higher abstractions. Studying the electric arc in 1849, he reported that it could both absorb and emit the yellow Fraunhofer rays, and he even suggested that the study of stellar spectra might profit astronomy. But it was the German physicist Gustav Kirchhoff, not Foucault, who a decade later formulated clear empirical rules describing the absorption and emission of radiation.

Finally, Foucault had not been trained by the accepted schools for the scientific elite, the École Polytechnique and the École Normale Supérieure. Perhaps for these reasons, and because of resentments engendered by his newspaper articles, it was only in 1865, on the sixth attempt that Foucault was finally elected to the Academy of Sciences, long after his reputation had been established.

Two years later, in July 1867, Foucault began to suffer from symptoms of paralysis. He died within seven months, at the age of 48, from what was possibly rapidly progressing multiple sclerosis, rather than overwork and worry, as was assumed at the time.

Donné's obituary of his friend revealed some of Foucault's many facets: "When I had introduced him into the intimacy of my own family, he was often felt to be disagreeable and annoying, most especially because he never lost his temper but defended his opinions in a tone that was always cool and calm. This calm was a great strength in certain circumstances, and I have seen him—he, of weak and fragile appearance—put the most formidable opponents to flight.... [But] he was never unwilling to explain the most abstract physical laws to laymen—and above all to laywomen."

Despite that last comment, Foucault never married. Wrote Donné: "He led, one might say, the life of one of those cats in good houses who install themselves in the best place in the bedroom or living room.... When some visitor called who did not interest him, he would retreat to a corner with his notebook and omnipresent pencil." Still, Donné continued, "if he was not supple, his fidelity in friendship withstood all tests." Indeed, at his funeral Foucault was mourned by a large circle of friends.

Foucault's name would later be cast into the ironwork of the Eiffel Tower. He was not a genius of Newton's or Einstein's ilk, but posterity rightly remembers this remarkable son of the City of Light. SA

The Author

WILLIAM TOBIN first became interested in Léon Foucault's life more than 10 years ago, while working at the Marseilles Observatory, where Foucault's largest telescope is preserved. Tobin is currently Senior Lecturer in Astronomy at the University of Canterbury in New Zealand. His astrophysical research focuses on variable stars in the Magellanic Clouds. Tobin obtained his bachelor's degree in physics from the University of Cambridge and his Ph.D. in astronomy from the University of Wisconsin–Madison. He is at work on a full-length biography of Foucault. This article was written during sabbatical leave at the Institute of Astrophysics in Paris.

Further Reading

FOUCAULT'S INVENTION OF THE SILVERED-GLASS REFLECTING TELESCOPE AND THE HISTORY OF HIS 80-CM REFLECTOR AT THE OBSERVATOIRE DE MARSEILLE. William Tobin in *Vistas in Astronomy,* Vol. 30, No. 2, pages 153–184; 1987.

PERFECTING THE MODERN REFLECTOR. William Tobin in *Sky & Telescope,* Vol. 74, No. 4, pages 358–359; October 1987.

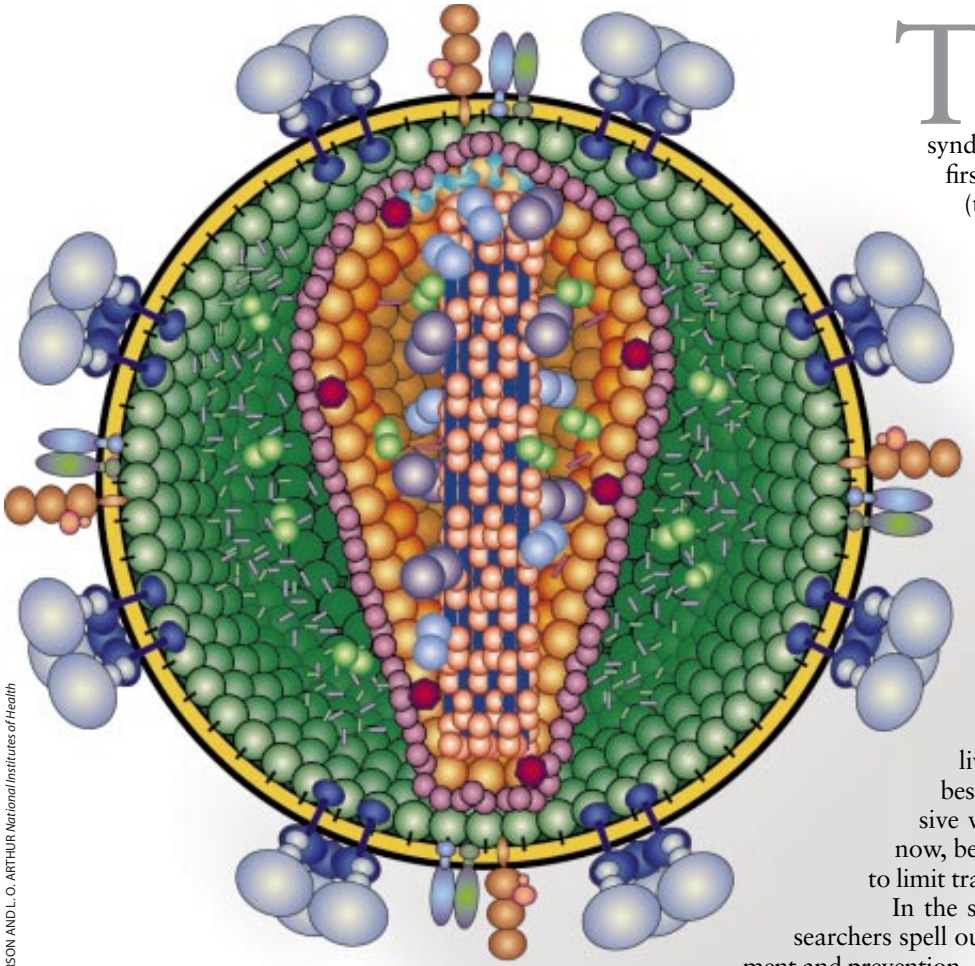
FRENCH DAGUERREOTYPES. Janet E. Buerger. University of Chicago Press, 1989.

TOOTHED WHEELS AND ROTATING MIRRORS: PARISIAN ASTRONOMY AND MID-NINETEENTH CENTURY EXPERIMENTAL MEASUREMENTS OF THE SPEED OF LIGHT. William Tobin in *Vistas in Astronomy,* Vol. 36, No. 3, pages 253–294; 1993.

FOUCAULT, HIS PENDULUM AND THE ROTATION OF THE EARTH. W. Tobin and B. Pippard in *Interdisciplinary Science Reviews,* Vol. 19, No. 4, pages 326–337; 1994.

SPECIAL REPORT

Defeating AIDS: What Will It Take?



L. E. HENDERSON AND L. O. ARTHUR *National Institutes of Health*

Ten years ago, when SCIENTIFIC AMERICAN published an issue devoted to AIDS (the acquired immunodeficiency syndrome), scientists knew that the disease, first identified in 1981, was caused by HIV (the human immunodeficiency virus). The virus had been uncovered within three years after AIDS itself was first recognized. But physicians were only beginning to understand that in the absence of an effective treatment, the virus was almost always fatal. And they were just starting to think of AIDS not as a discrete disorder but as the end stage of an unremitting HIV infection.

Today aggressive treatment can preserve health and prolong life for HIV-positive patients with access to optimal care. But the treatments are imperfect, costly and demanding. Further, they are unavailable to the overwhelming majority of HIV-infected people, most of whom live in the developing world. Clearly, the best solution is prevention by an inexpensive vaccine, but no vaccine is in sight. For now, behavioral change remains the main way to limit transmission of the virus.

In the special report that follows, leading researchers spell out the vexing challenges to better treatment and prevention—for children as well as adults. They also point the way to solutions, outlining emerging ideas for overcoming those obstacles and, ultimately, for conquering HIV completely. —*The Editors*

HIV 1998: The Global Picture

by Jonathan M. Mann and Daniel J. M. Tarantola

Worldwide, the populations most affected by the AIDS virus are often the least empowered to confront it effectively

In 1996, after more than a decade of relentless rises, deaths from AIDS finally declined in the U.S. The drop appears to have stemmed mainly from the introduction of powerful therapies able to retard the activity of HIV, the virus responsible for AIDS. Other economically advantaged nations, including France and Britain, have documented declines as well. But the trend in industrial countries is not representative of the world as a whole.

Further, the international pandemic of HIV infection and AIDS—composed of thousands of separate epidemics in communities around the globe—is expanding rapidly, particularly in the developing nations, where the vast majority of people reside. Since the early 1980s more than 40 million individuals have contracted HIV, and almost 12 million have died (leaving at least eight million orphans), according to UNAIDS, a program sponsored by the United Nations. In 1997 alone, nearly six million people—close to 16,000 a day—acquired HIV, and some 2.3 million perished from it, including 460,000 children.

This grim picture reflects some other unpalatable facts. Over the years, resources devoted to battling the pandemic have been apportioned along societal lines. Although more than 90 percent of HIV-infected people live in developing nations, well over 90 percent of the money for care and prevention is spent in industrial countries. This disparity explains why the new HIV-taming therapies, costing annually upward of $10,000 per person, have had no impact in the developing nations; by and large, these countries lack the infrastructure and funds to provide the medicines. In a few locales in the developing world, notably in parts of Uganda and Thailand, public health campaigns seem to be slowing the rate of infection. Yet those are the exceptions: in most other places, the situation is worsening.

HIV is spreading especially quickly in sub-Saharan Africa and in Southeast Asia. The region below the Sahara now houses two thirds of the globe's HIV-infected population and about 90 percent of all infected children. In areas of Botswana, Swaziland and several provinces of South Africa, one in four adults is afflicted; in many African countries, life expectancy, which had been rising since the 1950s, is falling. Unprotected heterosexual sex accounts for most of HIV's spread in sub-Saharan Africa, but the problem is compounded by contamination of the blood supply. At least a quarter of the 2.5 million units of blood administered in Africa (mostly to women and children) is not screened for the AIDS virus.

In Southeast Asia the epidemic is dominated by India (with three to five million HIV-infected individuals) and Thailand. It is now also raging in Burma and is expanding further into Vietnam and China.

Mirroring the dichotomy between the developing and the industrial nations, certain populations within nations are suffering a disproportionate number of infections. Epidemiologists have been dismayed to uncover a societal-level factor influencing the distribution: groups whose human rights are least respected are most affected. As epidemics mature within communities and countries, the brunt of the epidemic often shifts from the primary population in which HIV first appeared to those who were socially marginalized or discriminated against before the epidemic began.

Those who are discriminated against—whether because of their gender, race or economic status or because of cultural, religious or political affiliations—may have limited or no access to preventive information and to health and social services and may be particularly vulnerable to sexual and other forms of exploitation. And later, if they become infected, they may similarly be denied needed care and social support. Stigmatization therefore pursues its course unabated, deepening individual susceptibility and, as a result, collective vulnerability to the spread of HIV and to its effects.

As a case in point, 10 years ago in the

HIV'S GLOBAL EFFECT is devastating. Sub-Saharan Africa and South and Southeast Asia are home to the greatest number of HIV-infected people (*map*). The disease is also spreading fastest in those areas (*graph*). In the U.S., as in most nations, populations that encounter the most discrimination often face the highest risk of acquiring AIDS (*bar chart*). For instance, in 1996 black men and Hispanic men were, respectively, 51 and 25 times more likely than white women to be diagnosed.

U.S., whites accounted for 60 percent of AIDS cases and blacks and Hispanics for 39 percent. By 1996, 38 percent of new cases were diagnosed in whites and 61 percent in blacks and Hispanics. Further, between 1995 and 1996, the incidence of AIDS declined by 13 percent in whites but not at all in blacks.

Social marginalization, manifested in lack of educational and economic opportunity, also magnifies risk in the developing world. In Brazil, for example, the bulk of AIDS cases once affected people who had at least a secondary school education; now more than half the cases arise in people who have attended only primary school, if that. Moreover, in much of the world, women (who account for more than 40 percent of all HIV infections) have low social status and lack the power to insist on condom use or other safe-sex practices; they will be unable to protect themselves until their social status improves. Recognizing the societal-level roots of vulnerability to HIV and AIDS, UNAIDS has incorporated advancement of human rights in its global-prevention strategy.

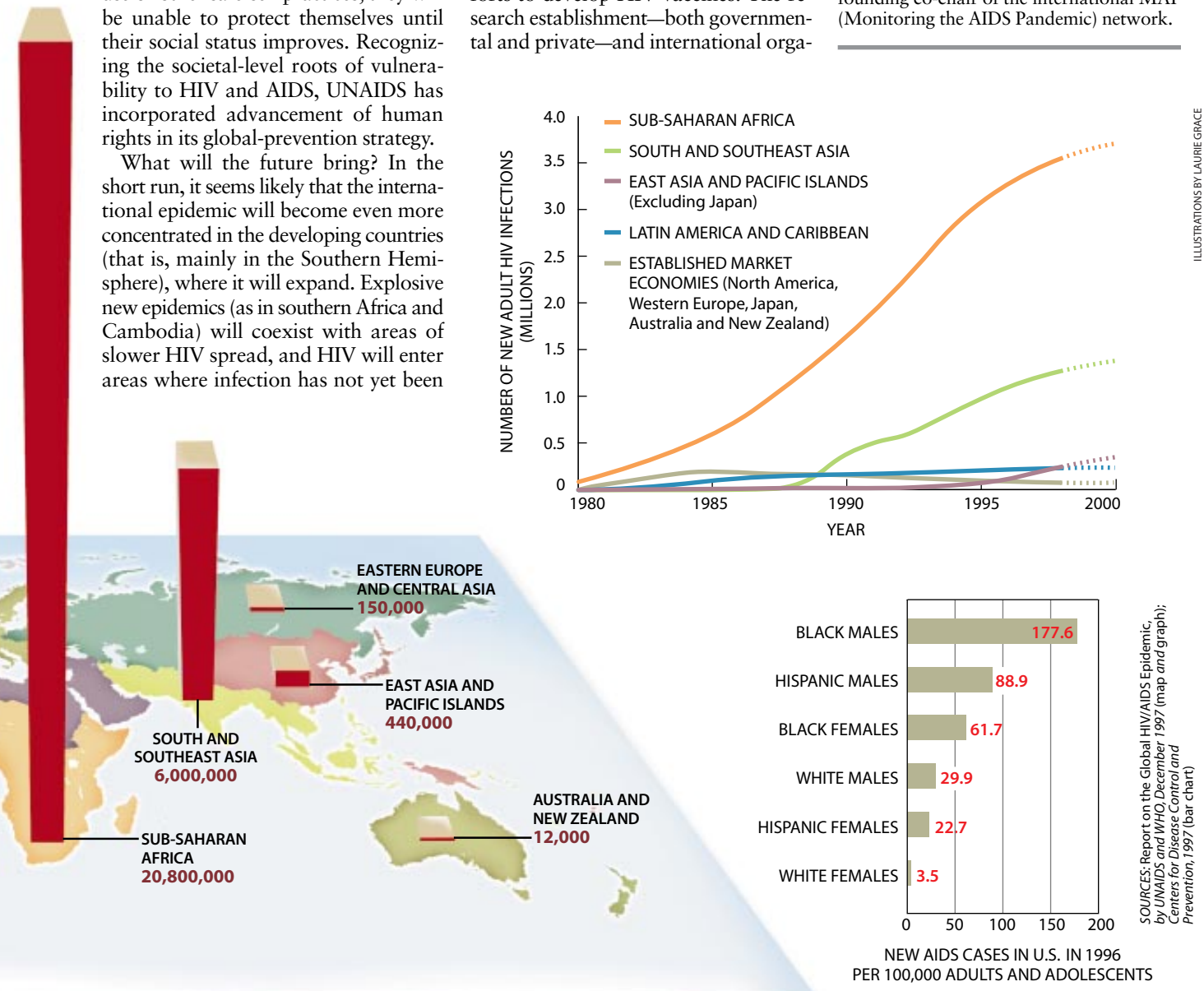
What will the future bring? In the short run, it seems likely that the international epidemic will become even more concentrated in the developing countries (that is, mainly in the Southern Hemisphere), where it will expand. Explosive new epidemics (as in southern Africa and Cambodia) will coexist with areas of slower HIV spread, and HIV will enter areas where infection has not yet been detected. And the already overwhelming burden of care will increase enormously.

In the industrial nations the epidemic will slow, at least for some populations, but it will take a higher toll on socially marginalized groups. The cost of care will rise substantially, as more and more infected individuals receive aggressive therapy. In both the Southern and Northern hemispheres, efforts to address societal forces that enhance vulnerability to HIV infection and AIDS will most likely proceed slowly, countered by significant resistance from social elites.

Control of the pandemic will require the extensive broadening of prevention programs. But given the large numbers of people at risk for infection and the difficulty of effecting behavioral change, the expansion of prevention programs will have to be coupled with greater efforts to develop HIV vaccines. The research establishment—both governmental and private—and international organizations must give highest priority to finding a vaccine and making it available to those who need it most: the marginalized populations who are bearing the brunt of the global HIV and AIDS pandemic. SA

The Authors

JONATHAN M. MANN and DANIEL J. M. TARANTOLA are co-editors of the 1996 report *AIDS in the World II* and, until recently, were colleagues at the François-Xavier Bagnoud Center for Health and Human Rights at the Harvard School of Public Health. In January, Mann, formerly director of the center and founding director of the WHO Global Program on AIDS, became dean of the School of Public Health at Allegheny University of the Health Sciences in Philadelphia. Tarantola is director of the International AIDS Program at the center and founding co-chair of the international MAP (Monitoring the AIDS Pandemic) network.



ILLUSTRATIONS BY LAURIE GRACE

SOURCES: Report on the Global HIV/AIDS Epidemic, by UNAIDS and WHO, December 1997 (map and graph); Centers for Disease Control and Prevention, 1997 (bar chart)

Improving HIV Therapy

by John G. Bartlett and Richard D. Moore

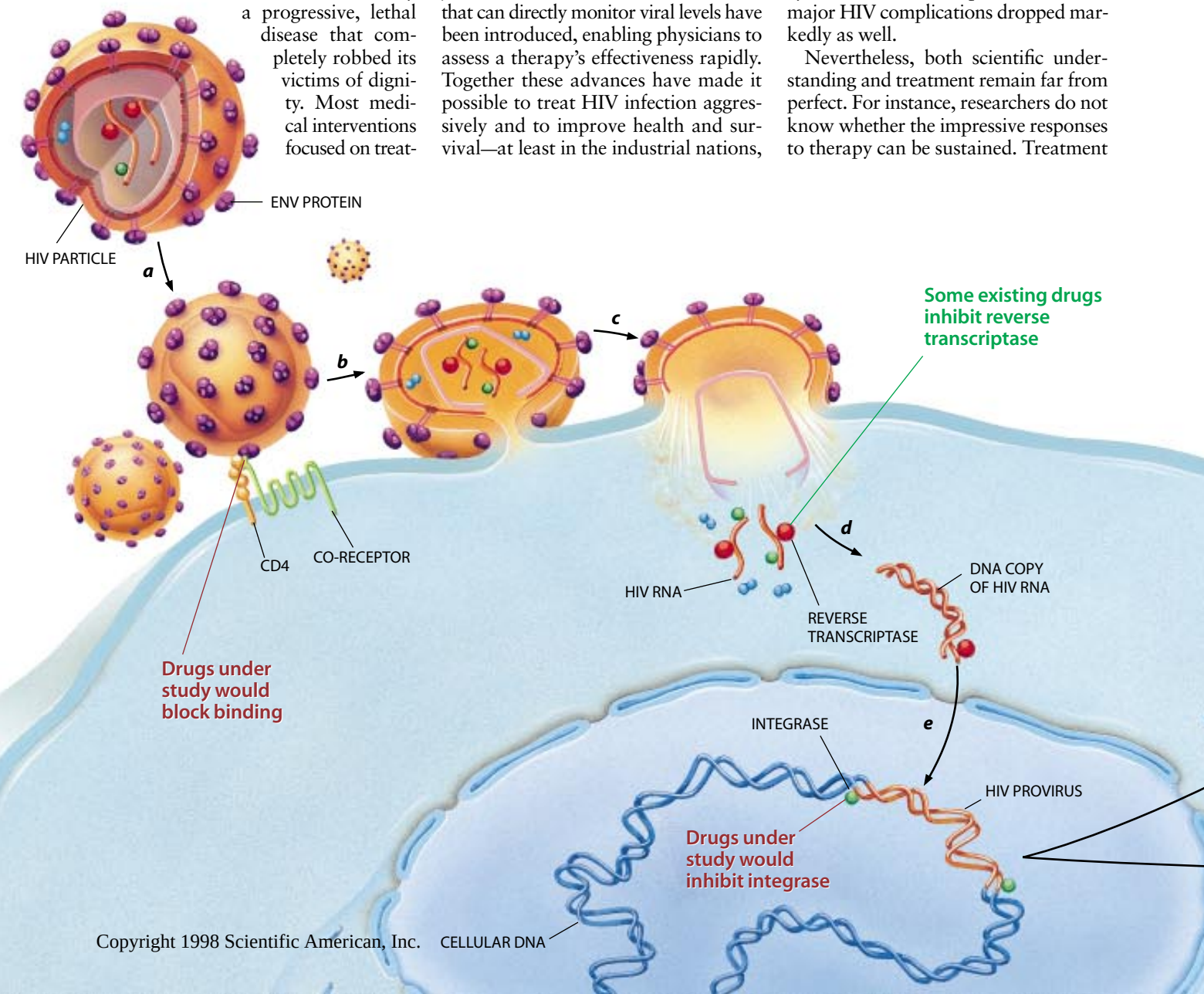
A vaccine would certainly be ideal for preventing infection by HIV and thus for avoiding AIDS—the late stage of HIV infection, when immunity is severely impaired. Yet the near-term prospects for a vaccine are poor, and people who contract the virus need care. For the immediate future, then, many scientists are concentrating on improving therapy.

Until only a few years ago, HIV infection was everyone's worst nightmare—it was almost invariably a progressive, lethal disease that completely robbed its victims of dignity. Most medical interventions focused on treating pneumonias and other serious "opportunistic" infections that stemmed from immune failure, not on controlling HIV itself.

Since late 1995, however, several related advances have led to a profound shift in the prospects for most patients who receive treatment today. Notably, scientists have gained a much fuller understanding of how HIV behaves in the body and a better sense of how to shackle it. Two classes of potent drugs have joined the anti-HIV arsenal, and tests that can directly monitor viral levels have been introduced, enabling physicians to assess a therapy's effectiveness rapidly. Together these advances have made it possible to treat HIV infection aggressively and to improve health and survival—at least in the industrial nations, where the intensive therapy is now widely available.

Initially, the signs of a sea change were anecdotal: extraordinary but increasingly frequent tales of people who were plucked from the brink of death to resume vigorous, productive lives. More recently, statistics have borne out the anecdotes. Between the first half of 1996 and the first half of 1997, deaths from AIDS in the U.S. declined by 44 percent. In roughly the same period the frequency of HIV-related hospitalizations and major HIV complications dropped markedly as well.

Nevertheless, both scientific understanding and treatment remain far from perfect. For instance, researchers do not know whether the impressive responses to therapy can be sustained. Treatment



Today's optimal treatments can work magic, but they are costly and onerous and do not work for everyone. What might the future bring?

is burdensome and costly, which puts it out of reach of certain patients. In addition, some fraction of people who receive the best available care respond poorly. For such reasons, the search continues for ways to make therapy more universally effective and accessible.

The ultimate goal, of course, is a cure. Investigators are unsure whether that aim is feasible. But many of us are cautiously optimistic that we are, at last, beginning to accumulate the weaponry needed to manage HIV as a bearable, chronic disorder, somewhat akin to diabetes or hypertension.

Recommendations for optimal therapy seek to halt viral replication indefinitely—something inconceivable just three years ago. To meet this target, patients must usually take three or even four carefully selected drugs twice or more a day, exactly as prescribed. These general guidelines, and more specific recommendations, derive from current knowledge of HIV's activity in untreated patients.

How HIV Harms

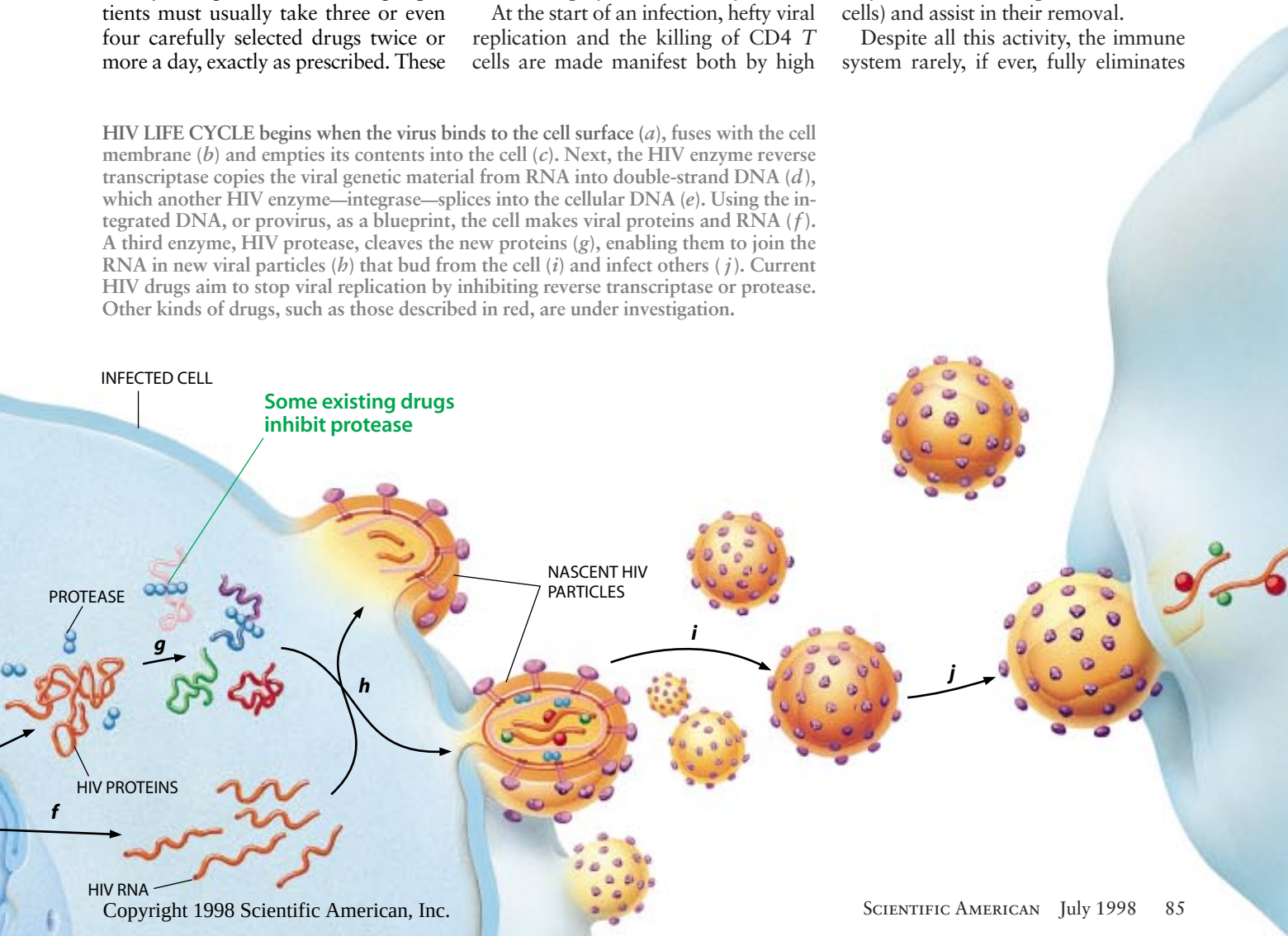
The virus spreads from one person to another usually through sexual intercourse, direct exposure to contaminated blood, or transmission from a mother to her fetus or suckling infant. In the body, HIV invades certain cells of the immune system—including CD4, or helper, *T* lymphocytes—replicates inside them and spreads to other cells. (These lymphocytes, named for the display of a molecule called CD4 on their surface, are central players in immunity.)

At the start of an infection, hefty viral replication and the killing of CD4 *T* cells are made manifest both by high levels of HIV in the blood and by a dramatic drop in CD4 *T* cell concentrations from the normal level of at least 800 cells per cubic millimeter of blood. About three weeks into this acute phase, many people display symptoms reminiscent of mononucleosis, such as fever, enlarged lymph nodes, rash, muscle aches and headaches. These maladies resolve within another one to three weeks, as the immune system starts to gain some control over the virus. That is, the CD4 *T* cell population responds in ways that spur other immune cells—CD8, or cytotoxic, *T* lymphocytes—to increase their killing of infected, virus-producing cells. The body also produces antibody molecules in an effort to contain the virus; they bind to free HIV particles (outside cells) and assist in their removal.

Despite all this activity, the immune system rarely, if ever, fully eliminates

TOMO NARASHIMA

HIV LIFE CYCLE begins when the virus binds to the cell surface (*a*), fuses with the cell membrane (*b*) and empties its contents into the cell (*c*). Next, the HIV enzyme reverse transcriptase copies the viral genetic material from RNA into double-strand DNA (*d*), which another HIV enzyme—integrase—splices into the cellular DNA (*e*). Using the integrated DNA, or provirus, as a blueprint, the cell makes viral proteins and RNA (*f*). A third enzyme, HIV protease, cleaves the new proteins (*g*), enabling them to join the RNA in new viral particles (*h*) that bud from the cell (*i*) and infect others (*j*). Current HIV drugs aim to stop viral replication by inhibiting reverse transcriptase or protease. Other kinds of drugs, such as those described in red, are under investigation.



the virus. By about six months, the rate of viral replication reaches a lower, but relatively steady, state that is reflected in the maintenance of viral levels at a kind of "set point." This set point varies greatly from patient to patient and dictates the subsequent rate of disease progression; on average, eight to 10 years pass before a major HIV-related complication develops. In this prolonged, chronic stage, patients feel good and show few, if any, symptoms.

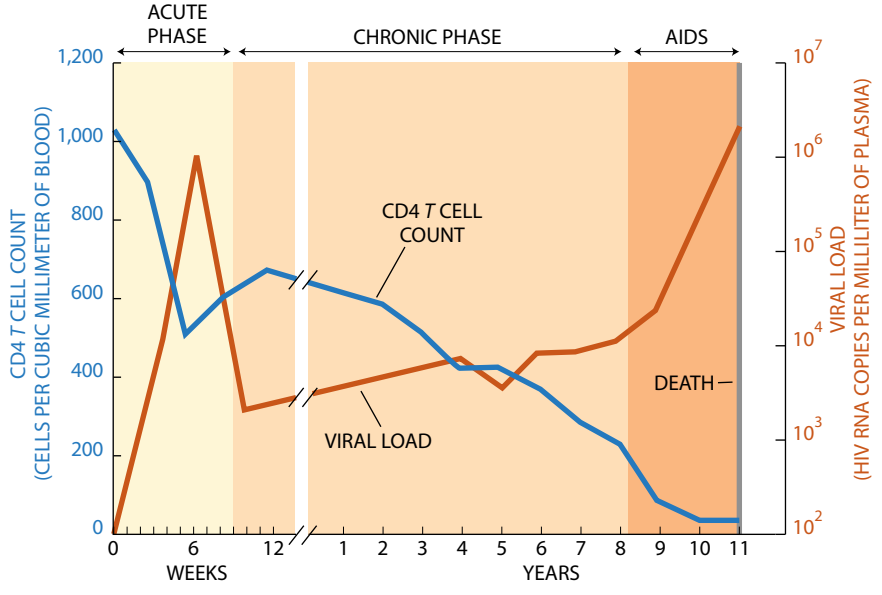
Their apparent good health continues because CD4 *T* cell levels remain high enough to preserve defensive responses to other pathogens. But over time, CD4 *T* cell concentrations gradually fall. When the level drops below 200 cells per cubic millimeter of blood, people are said to have AIDS.

As levels dip under 100, the balance of power shifts away from the immune system. HIV levels skyrocket, and microbes that the immune system would normally control begin to proliferate extensively, giving rise to the potentially deadly opportunistic infections that are the hallmarks of AIDS (such as *Pneumocystis carinii* pneumonia and toxoplasmosis). Once such diseases appear, AIDS frequently becomes lethal within a year or two. (Opportunistic infections sometimes occur before the CD4 *T* cell level falls under 200; in that case, the appearance of the infections leads to a diagnosis of AIDS, regardless of the CD4 *T* cell level.)

Although patients typically survive HIV infection for 10 or 11 years, the course can vary enormously. Some die within a year after contracting the virus, whereas an estimated 4 to 7 percent maintain fully normal CD4 *T* cell counts for eight years or more and survive beyond 20 years.

At the cellular level, scientists also know how HIV invades and destroys CD4 *T* lymphocytes. The virus gains access to the interior of these cells (and certain other cell types) by binding to CD4 itself and to another molecule, a "co-receptor," on the cell surface. Such binding enables HIV to fuse with the cell membrane and to release its contents into the cytoplasm. Those contents include various enzymes and two strands of RNA that each carry the entire HIV genome: the genetic blueprint for making new HIV particles.

One of the enzymes, reverse transcriptase, copies the HIV RNA into double-strand DNA (a property that qualifies HIV as a "retrovirus"). Then a second enzyme, integrase, helps to splice the HIV DNA permanently into a chromosome in the host cell. When a *T* cell that harbors this integrated DNA (or provirus) becomes activated against HIV or other microbes, the cell replicates and also unwittingly begins to produce new copies of both the viral genome and viral proteins. Now another HIV enzyme—a protease—cuts the new protein molecules into forms that are packaged with the virus's RNA genome in new viral particles. These particles bud from the cell and infect other cells. If enough particles form, they can overwhelm and kill the cell that produced them.



BRYAN CHRISTIE

SOURCE: Anthony Fauci et al. in Annals of Internal Medicine, *Vol. 124; 1996*

NATURAL COURSE OF HIV INFECTION in a typical untreated patient begins with a sharp rise of virus in the blood (*orange line*) and a consequent drop in CD4 *T* cells (*blue line*), the immune cells most damaged by HIV. The immune system soon recovers somewhat, however, and keeps HIV levels fairly steady for several years. Eventually, though, the virus gains the upper hand. AIDS is diagnosed when the CD4 *T* cell level drops below 200 cells per cubic millimeter of blood or when opportunistic infections (reflecting failed immunity) arise, whichever happens first.

Data's Message: Stop Viral Growth

All approved anti-HIV, or antiretroviral, drugs attempt to block viral replication within cells by inhibiting either reverse transcriptase or the HIV protease. Two classes inhibit reverse transcriptase and thus forestall genetic integration. Agents in one of these classes, the nucleoside analogues, resemble the natural substances that become building blocks of HIV DNA; when reverse transcriptase tries to add the drugs to a developing strand of HIV DNA, the drugs prevent completion of the strand. This group includes the first anti-HIV drug—zidovudine (AZT), which was introduced in 1987—and its close chemical relatives. Nonnucleoside reverse transcriptase inhibitors, composed of other kinds of substances, constitute the second class of antiretrovirals. A third class, the protease inhibitors, blocks the active, catalytic site of the HIV protease, thereby preventing it from cleaving newly made HIV proteins.

The basic course of untreated HIV infection has been known for a while, but recent work has been filling in some missing pieces. It is these results that have convinced physicians of the urgent need to halt viral replication as completely as possible.

At one time, for example, crude technology suggested that HIV actually infected few CD4 *T* cells and replicated only weakly for a long time. This view implied that most of the lost *T* cells disappeared via a mechanism other than wild HIV proliferation, which in turn meant that drugs able to block viral reproduction might not interfere much with disease progression until shortly before AIDS set in. Now it is clear that HIV replicates prolifically from the start. HIV levels remain fairly stable for several years only because the body responds for a time by manufacturing extraordinary numbers of CD4 *T* cells.

Anti-HIV Drugs Now on the Market*

Generic Name (Other Names)	Typical Dosage	Some Potential Side Effects
Reverse Transcriptase Inhibitors: Nucleoside Analogues		
Didanosine (Videx, ddI)	2 pills, 2 times a day on empty stomach	Nausea, diarrhea, pancreatic inflammation, peripheral neuropathy
Lamivudine (Epivir, 3TC)	1 pill, 2 times a day	Usually none
Stavudine (Zerit, d4T)	1 pill, 2 times a day	Peripheral neuropathy
Zalcitabine (HIVID, ddC)	1 pill, 3 times a day	Peripheral neuropathy, mouth inflammation, pancreatic inflammation
Zidovudine (Retrovir, AZT)	1 pill, 2 times a day	Nausea, headache, anemia, neutropenia (reduced levels of neutrophil white blood cells), weakness, insomnia
Pill containing lamivudine and zidovudine (Combivir)	1 pill, 2 times a day	Same as for zidovudine
Reverse Transcriptase Inhibitors: Nonnucleoside Analogues		
Delavirdine (Rescriptor)	4 pills, 3 times a day (mixed into water); not within an hour of antacids or didanosine	Rash, headache, hepatitis
Nevirapine (Viramune)	1 pill, 2 times a day	Rash, hepatitis
Protease Inhibitors		
Indinavir (Crixivan)	2 pills, 3 times a day on empty stomach or with a low-fat snack and not within 2 hours of didanosine	Kidney stones, nausea, headache, blurred vision, dizziness, rash, metallic taste in mouth, abnormal distribution of fat, elevated triglyceride and cholesterol levels, glucose intolerance
Nelfinavir (Viracept)	3 pills, 3 times a day with some food	Diarrhea, abnormal distribution of fat, elevated triglyceride and cholesterol levels, glucose intolerance
Ritonavir (Norvir)	6 pills, 2 times a day (or 4 pills, 2 times a day if taken with saquinavir) with food and not within 2 hours of didanosine	Nausea, vomiting, diarrhea, abdominal pain, headache, prickling sensation in skin, hepatitis, weakness, abnormal distribution of fat, elevated triglyceride and cholesterol levels, glucose intolerance
Saquinavir (Invirase, a hard-gel capsule; Fortovase, a soft-gel capsule)	6 pills, 3 times a day (or 2 pills, 2 times a day if taken with ritonavir) with a large meal	Nausea, diarrhea, headache, abnormal distribution of fat, elevated triglyceride and cholesterol levels, glucose intolerance

*As of April 1998

SLIM FILMS

DRUG CHOICES today are varied and growing. Yet the most effective treatments (usually two nucleoside analogues and one or two protease inhibitors) can be demanding and complex—too much so for some patients. All require remembering many pills a day, some swallowed on an empty stomach, some not. The medicines can also produce side effects and often cannot be taken with certain anti-HIV or other medications, such as painkillers, antidepressants or agents that ease nausea. The regimen having the fewest pills—eight—uses indinavir and Combivir. More cumbersome plans that are widely used include ones combining ritonavir and saquinavir with Combivir (14 pills total) or combining saquinavir with didanosine and stavudine (24 pills).

In addition, investigators have learned that in untreated patients the strength of the initial immune response (in the acute stage) apparently exerts a decisive influence on the rate of progression to AIDS. Patients who display strong CD8 *T* cell activity, and who thus achieve greater suppression of viral replication and a lower viral set point, progress more slowly than do individuals who mount a weaker fight.

It also seems that strong immunological activity in the acute phase of infection helps to preserve the body's later ability to manufacture the subset of CD4 *T* cells that specifically react to HIV. Once those cells are lost, they may not return, even if subsequent treatment stops viral replication and gives the immune system a chance to shore up its overall CD4 *T* cell numbers.

Finally, at any stage, viral levels correlate with prognosis. Many studies specifically relating those levels to disease progression suggest that patients whose viral concentrations fall into the undetectable realm and stay there are most likely to avoid progression to AIDS.

In aggregate, such findings teach that the amount of virus in the system plays a major role in determining a patient's fate. That is why therapy must aim to shut down viral reproduction. For the vast majority of patients, whose immune systems cannot fight HIV adequately unaided, aggressive drug therapy offers the best chance for long-term well-being. It appears, however, that even patients who respond well to therapy (in whom replication stops) will have to continue taking the medicines for several years and perhaps indefinitely. The reason, as will be seen, has to do with the presence in the body of HIV-sequestering havens that are not eradicated by antiretroviral therapy.

The Elements of Optimal Therapy

Theory and clinical trials indicate that the best way to achieve maximum viral suppression is HAART: highly active antiretroviral therapy. At the moment, HAART usually consists of triple therapy, including two nucleoside analogues and a protease inhibitor (at a cost for the medicines of about $10,000 to $12,000 a year). For an up-to-date list of recommended combinations, readers can consult *Guidelines for the Use of Antiretroviral Agents in HIV-Infected Adults and Adolescents,* issued by the

How Drug Resistance Arises

by Douglas D. Richman

When anti-HIV therapy fails to keep HIV levels suppressed, the cause is often viral resistance to at least one of the drugs being administered. How does such insensitivity develop?

At its root, resistance is mediated by mutations in viral genes. The genome (complete set of genes) in any HIV particle serves as the blueprint for enzymes and other proteins needed to make new copies of the virus. Current anti-HIV drugs, or antiretrovirals, bind to specific HIV enzymes and impede their activity. Gene mutations can alter these enzymes in ways that diminish this binding or otherwise undermine the pharmaceutical attack. Some medicines, including the nonnucleoside reverse transcriptase inhibitors, can be undone by a single mutation; others, such as the protease inhibitors, require multiple mutations in a single genome.

Ironically, use of an antiretroviral can actually promote proliferation of drug-resistant HIV variants. The reason relates to HIV's rapid and relatively sloppy replication. Unchecked, HIV makes roughly 10 billion new viral particles every day in an infected individual. Yet it does so without great attention to accuracy. Hence, the genome of each new particle is likely to differ from the "parent" genome in at least one spot. Coupled with HIV's prolific rate of reproduction, this inaccuracy means that every mutation able to cause or contribute to drug resistance is likely to be generated daily in one or another HIV particle. In other words, even if a patient has never been treated, any compound that is delivered will encounter some HIV variant that is already resistant or is on its way to accumulating the full set of mutations needed for resistance.

Now consider what happens when a patient takes an antiretroviral. For argument's sake, let's say this medicine can be thwarted only by five mutations in a genome. The drug will bar reproduction by variants that are still sensitive to the agent, but HIV subpopulations that are somewhat insensitive will continue to proliferate to an extent. (In advance of therapy, most variants probably will carry no more than two resistance mutations.) Over time, the ongoing replication will enable at least some of the semiresistant variants to acquire another resistance mutation. If the drug is still present, these forms of HIV will outpace their more susceptible cousins and will have the opportunity to acquire the other needed genetic alterations. At that point, they will evade the antiretroviral entirely and grow unchecked.

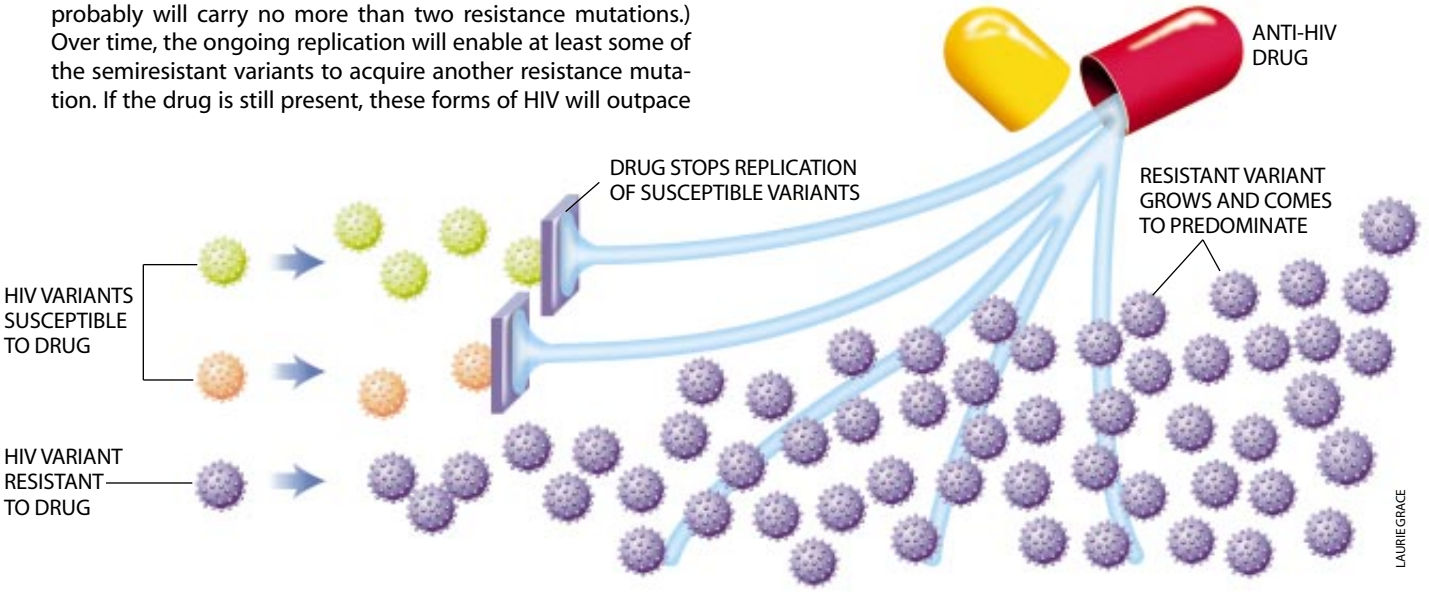
In other words, use of an anti-HIV agent will "select for" the development and growth of resistant strains unless the agent blocks all viral proliferation. Similarly, if resistant virus is present when a drug is first delivered, the compound will permit growth of the resistant population and will fail quickly.

Because resistance can occur readily, and because no single pharmaceutical on the market is powerful enough to suppress HIV on its own, physicians no longer treat patients with single drugs (monotherapy). They often choose combinations that both maximize potency and reduce the likelihood of resistance. For example, physicians try to avoid anti-HIV drugs the patient has taken before, under the assumption that past use will have helped establish insensitive HIV populations.

Having selected a therapy, doctors must also stay on the lookout for signs of emerging resistance. In general, if virus is detectable in the blood after four to six months of therapy—whether because of inadequate drug potency, insufficient drug absorption or poor adherence to the drug regimen—its presence warns that replication is occurring and that resistance has developed or is likely to follow.

Tests still under development may improve the ability to tailor therapies to individual patients. One kind will examine the genetic makeup of the HIV variants in individuals, to determine which resistance mutations they carry. The other will assess the extent of resistance to particular drugs. With such information, physicians will be able to select antiretrovirals that are least likely to meet resistance. Also, if viral levels in a patient begin to rise, the tests should reveal whether resistance is at fault and should help physicians to identify alternative therapies having the best chance of success.

DOUGLAS D. RICHMAN is professor of pathology and medicine and co-director of the AIDS Research Institute at the University of California at San Diego School of Medicine. He is also director of the Research Center for AIDS and HIV Infection at the San Diego Veterans Affairs Medical Center.



SINGLE DRUG "SELECTS FOR" growth of resistant virus. In the absence of medication, many HIV variants coexist in a patient (*left*). If a replication-blocking drug is given, it will halt replication of susceptible variants (*top*) but will essentially be ignored by resistant versions (*bottom*), which will continue multiplying after the drug-sensitive versions have stopped.

U.S. Department of Health and Human Services (on the World Wide Web, see http://www.hivatis.org/trtgdlns.html).

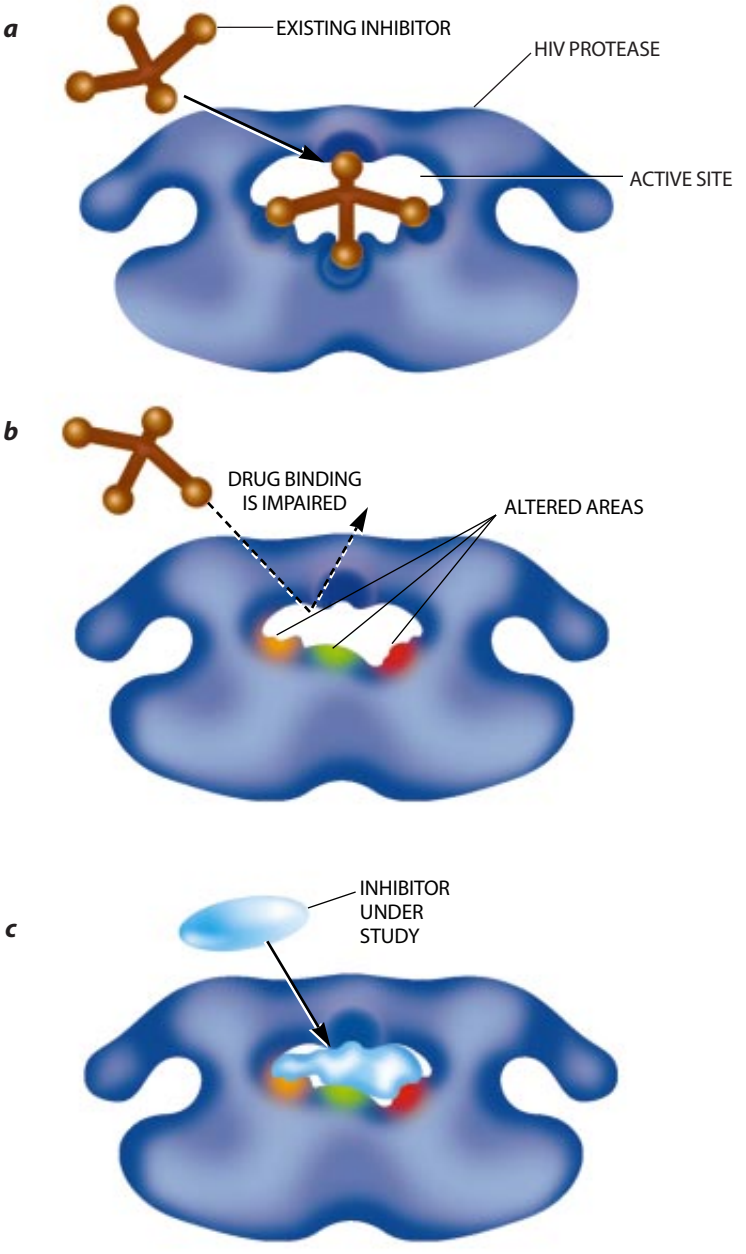
Doctors prescribe the combination of two nucleoside analogues and a protease inhibitor most often because this configuration was the first shown to work well. Other kinds of drug cocktails that incorporate newer drugs and may be more powerful or simpler to use are being evaluated as first-line options as well. Among those are two protease inhibitors or one protease inhibitor combined with a nonnucleoside reverse transcriptase inhibitor; drugs in both those classes are more potent than zidovudine and other nucleoside analogues. Physicians are also examining the value of combining four or more agents.

Multiple-drug regimens make sense for a couple of reasons. If one drug fails to block viral replication, a second may pick up the slack. And if both of those attacks fail, the third drug should provide extra insurance.

In addition, as Douglas D. Richman explains on the opposite page, HIV inevitably becomes resistant, or unresponsive, to antiretrovirals that fail to suppress viral replication completely. Because no single drug on the market can achieve such suppression on its own, any agent given alone will eventually be rendered useless, sometimes within weeks. Moreover, HIV strains that become resistant to one drug often become insensitive to other drugs in the same class (a phenomenon known as cross-resistance), eliminating those drugs as alternatives. The virus should have much more difficulty becoming resistant to treatment if it is assaulted by a mix of compounds that together end viral replication fully.

Given that viral levels in the blood correlate with time to AIDS and years of survival, physicians monitor those levels as a window to a therapy's effectiveness. That way, treatments unable to control the virus can be detected before immune failure results, in time to take corrective measures.

Viral levels are assessed by viral-load assays, which count copies of HIV RNA in a milliliter of plasma (the cell-free part of blood); the number of viral particles is half the RNA count [*see next page*]. Current tests are sensitive to RNA concentrations of 500 or more copies per milliliter and above. Assays now becoming available are sensitive to as few as 50 copies per milliliter. Within eight weeks after therapy starts, viral levels should drop by at least 10-fold.



STRATEGY UNDER STUDY aims to overcome HIV resistance to existing protease inhibitors. Those drugs bind to the active site of the protease enzyme, thus preventing the enzyme from functioning (*a*). HIV can evade this effect by altering parts of the enzyme (*bright colors in b*) so that the drug cannot attach effectively. A new drug in development (*light blue in c*) has been designed to be flexible enough to bind to—and inhibit—even altered forms of HIV protease. All drawings are highly schematic.

By six months from the initiation of treatment, levels should be undetectable and remain so thereafter.

In clinical trials the results of triple therapy have been phenomenal. For instance, in patients who had CD4 *T* cell counts between 200 and 500 cells per cubic millimeter and had not been treated before (and so were unlikely to harbor resistant virus), 75 to 85 percent achieved viral loads below 500 RNA copies per milliliter within 24 to 100 weeks, and 60 to 75 percent of the patients achieved viral loads below 50 copies.

The "real life" experience has also been good but less so. Almost identical results from clinics at San Francisco General Hospital and the Johns Hopkins Medical Institutions reveal that approximately 50 percent of patients given triple-drug therapy achieved the goal of viral loads below 500 copies per milliliter at six to 52 weeks after the start of treatment.

Viral-Load Tests Provide Valuable Answers

by John W. Mellors

In the early 1990s tests that could accurately detect the amount of HIV in a patient's blood finally became available. These viral-load assays have since revolutionized understanding of HIV's behavior and have helped define new principles of therapy. The assays directly measure viral RNA per milliliter of blood plasma; each HIV particle contains two strands of RNA, so the level of actual virus is half the RNA count.

The tests enabled researchers to show that HIV never undergoes a period of slow growth. From the start, the collective viral population in a patient generates many billions of new HIV particles a day, resulting in destruction of millions of CD4 *T* lymphocytes—the immune cells most depleted in infected patients. The body tries to compensate for the loss by making new CD4 *T* cells, but the immune system remains under constant siege and eventually fails to keep up.

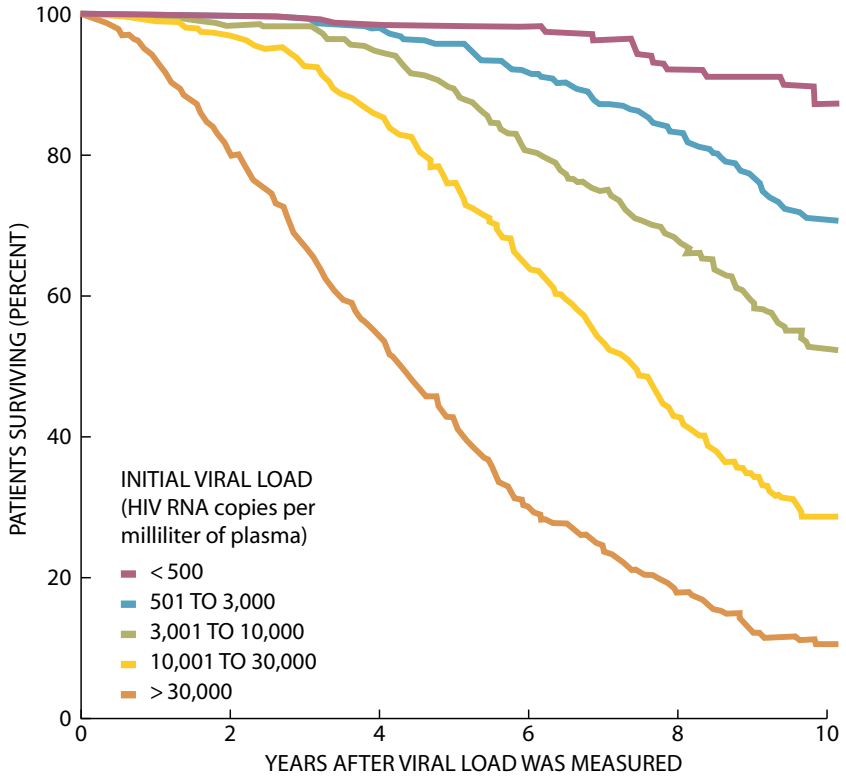
In 1996, as part of the federally supported Multicenter AIDS Cohort Study (MACS), my colleagues and I measured virus in stored plasma samples collected from about 1,600 untreated HIV-infected men and traced the fate of those patients. We found striking differences in prognosis, depending on the level of virus (*graph*). For example, 70 percent of the men whose viral load was greater than 30,000 copies per milliliter died within six years of the test, the average survival time being 4.4 years. In contrast, fewer than 1 percent of patients whose viral load was below 500 copies per milliliter died in six years, and the average survival time was more than 10 years. These results established that viral load critically influences the rate of disease progression. They also suggested that lowering viral levels as much as possible for as long as possible with therapy is essential to prolonging life.

Subsequent studies using viral-load tests have both confirmed this concept and changed how new therapies are evaluated. Until recently, investigators assessed potential treatments by comparing the incidence of AIDS or death in test subjects and a control group, often having to wait years for definitive results. After the viral-load assays became available, several large studies demonstrated that measures of viral load, often after just weeks of treatment, were valid indicators of whether a therapy could slow progression of HIV infection. For example, treatments that lowered the load by 75 to 90 percent within eight to 24 weeks reduced by 50 to 65 percent the likelihood of progressing to AIDS within a year. Viral-load measures have therefore replaced assessment of clinical outcome in therapeutic trials, and routine monitoring of viral levels has been incorporated into medical practice.

Recent studies have suggested a refinement of the admonition to keep viral levels low. Current treatment guidelines aim to maintain viral load below 500 copies per milliliter (the limit of detection for the tests usually used today). Failure to reach that level is associated with breakthrough of drug-resistant virus and loss of control over HIV replication. Yet trials applying more sensitive tests indicate that depressing viral load below 50 copies per milliliter offers better insurance against resistance. Moreover, such a reduction is probably necessary to halt viral replication everywhere—in lymph nodes (where the replication rate is higher than in the blood) and also in other body compartments. I believe, therefore, that viral loads below 50 should become the new goal if the more sensitive tests (now used only in research) are made readily available. Not all physicians agree, however, in part because this target may be more difficult to reach, particularly for patients in whom initial therapy has failed.

A major new challenge for therapy is finding ways to eliminate HIV from infected resting CD4 *T* cells, which do not produce viral particles but harbor the genetic blueprints for doing so in the future. Current antiretroviral drugs cannot eliminate these HIV reservoirs. To develop such therapies and monitor their effects, investigators will undoubtedly require a new generation of viral-load tests—ones that can accurately measure the virus hiding in resting cells. Fortunately, efforts to create those urgently needed assays are under way.

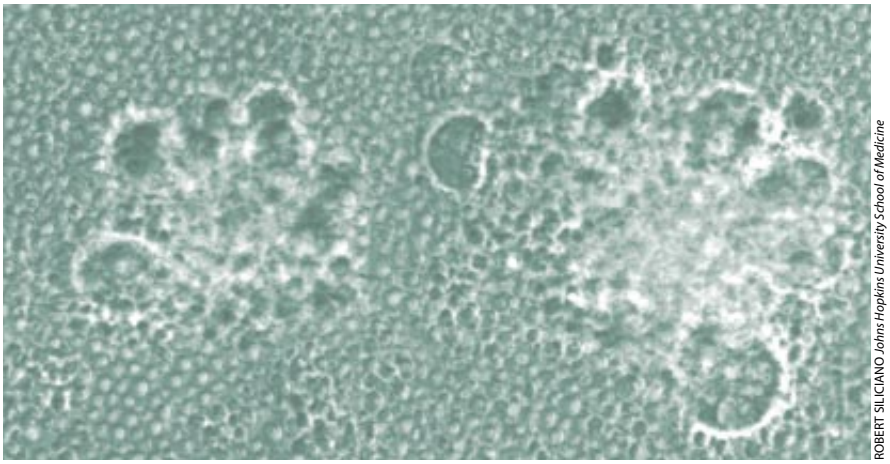
JOHN W. MELLORS is associate professor of medicine and director of the HIV/AIDS programs at the University of Pittsburgh Medical Center. He also serves as a staff physician at the Veterans Affairs Pittsburgh Healthcare System.



TEN-YEAR SURVIVAL in 1,604 HIV-infected men depended on the patients' viral load at the start of the 10-year period. The men were followed in the era before effective therapy for HIV was introduced; their viral load was assessed retrospectively, from stored blood samples. This finding was among the first to establish the major influence of viral load on disease progression.

Although the "field" results have not matched those of the clinical trials, a 50 percent success rate is still 100 percent better than could be achieved just a few years ago. And that imperfect level of success apparently has been enough to slow disease progression in many cases. Since HAART was introduced, the HIV clinics at San Francisco General and at Johns Hopkins have witnessed a 50 to 80 percent decrease in hospitalizations for HIV-related problems and a 50 to 70 percent drop in the incidence of major AIDS-related opportunistic infections.

The less stellar results in the community are not surprising. Patients cared for in clinics are much more heterogeneous than are trial participants. Many begin therapy in a relatively late stage of HIV infection, when their viral burden may be too large to overcome with existing therapy and when their immune system may be beyond rescue. Most clinic patients, moreover, have had prior exposure to antiretroviral drugs and so probably harbor virus resistant to one or more drugs in the regimen.



ROBERT SILICIANO *Johns Hopkins University School of Medicine*

TWO CLUSTERS consist of *T* cells that have fused and are dying because formerly quiescent HIV in a hidden "reservoir"—in a single, inactive (resting) *T* cell—was stimulated to replicate. The viral particles released from that cell then spread to and destroyed healthy cells. To cure HIV, investigators will have to find a way to eradicate potentially deadly reservoirs, which are not eliminated by existing drugs.

The Down Side of HAART

What is more, HAART has many drawbacks that can hinder adherence unless a patient is very committed and organized. Patients must swallow at least eight, and often 16 or more, anti-HIV pills a day (along with any other needed medicines, such as those meant to prevent specific opportunistic infections or to control pain). They must remember which ones have to be consumed with food, which on an empty stomach and which cannot be taken together or with other kinds of pills. Even the most efficient individuals can become confused or forget doses; a number of clinic patients are homeless, demented or dependent on illicit drugs or otherwise lacking in the supports that would help them adhere strictly to a medication schedule.

Patients missing the encouragement and support provided in clinical trials may also be more likely to skip doses or to give up on therapy when they encounter unpleasant side effects. The unwanted effects of antiretrovirals can range from rash, nausea, diarrhea and headache to anemia, neuropathy (painful or numb feet), hepatitis and possibly diabetes. Certain effects, if severe enough, can be dangerous or intolerable for anyone, but others are potentially bearable.

Whatever the reasons for missing doses, the consequences can be incomplete suppression of viral replication and the emergence of resistance. HAART, as some say, is unforgiving of nonadherence. Once resistance sets in, the patient cannot "start fresh" with the same plan; it will not work. Regrettably, incomplete adherence to treatment plans accounts for about half of treatment failures. No one knows the minimum number of missed pills needed to promote resistance.

We should note that in some patients who apparently follow therapy strictly, the virus fails to reach, or to stay at, undetectable levels. This virological failure may occur because the prescribed drug cocktail is not potent enough to halt viral replication fully in those particular individuals. Their bodies, for whatever reason, may break down the drugs too rapidly or may block their complete dissemination, resulting in too little medication reaching infected cells. Also, when viral-load measures are below 500 copies per milliliter, some people will truly have no viral replication occurring, but others will have some unknown degree of activity. This small amount of activity may be enough to allow resistance to arise.

Unsettled Issues

Experience with HAART in the past few years has been instructive, but it has yet to resolve a number of questions—among them, Should everyone be treated as soon as they are diagnosed?

Experts agree that the ideal time to institute therapy is during the acute stage, when patients have the best chance of preserving their immune defenses. Few patients are diagnosed at that stage, however. At the other end of the spectrum, people who already have symptoms, or who have a CD4 *T* cell count below 200 per cubic millimeter, need treatment, too; they probably will not live long without intervention.

Authorities on HIV care also favor offering HAART to a third group: symptomless patients with a CD4 *T* cell count between 200 and 500 cells per cubic millimeter or a viral-load measure exceeding 10,000 to 20,000 RNA copies per milliliter. Without treatment, patients in this third category have at least an 8 percent chance of progressing to an AIDS-defining opportunistic infection within three years and at least a 25 percent chance of reaching that unwelcome benchmark within six years.

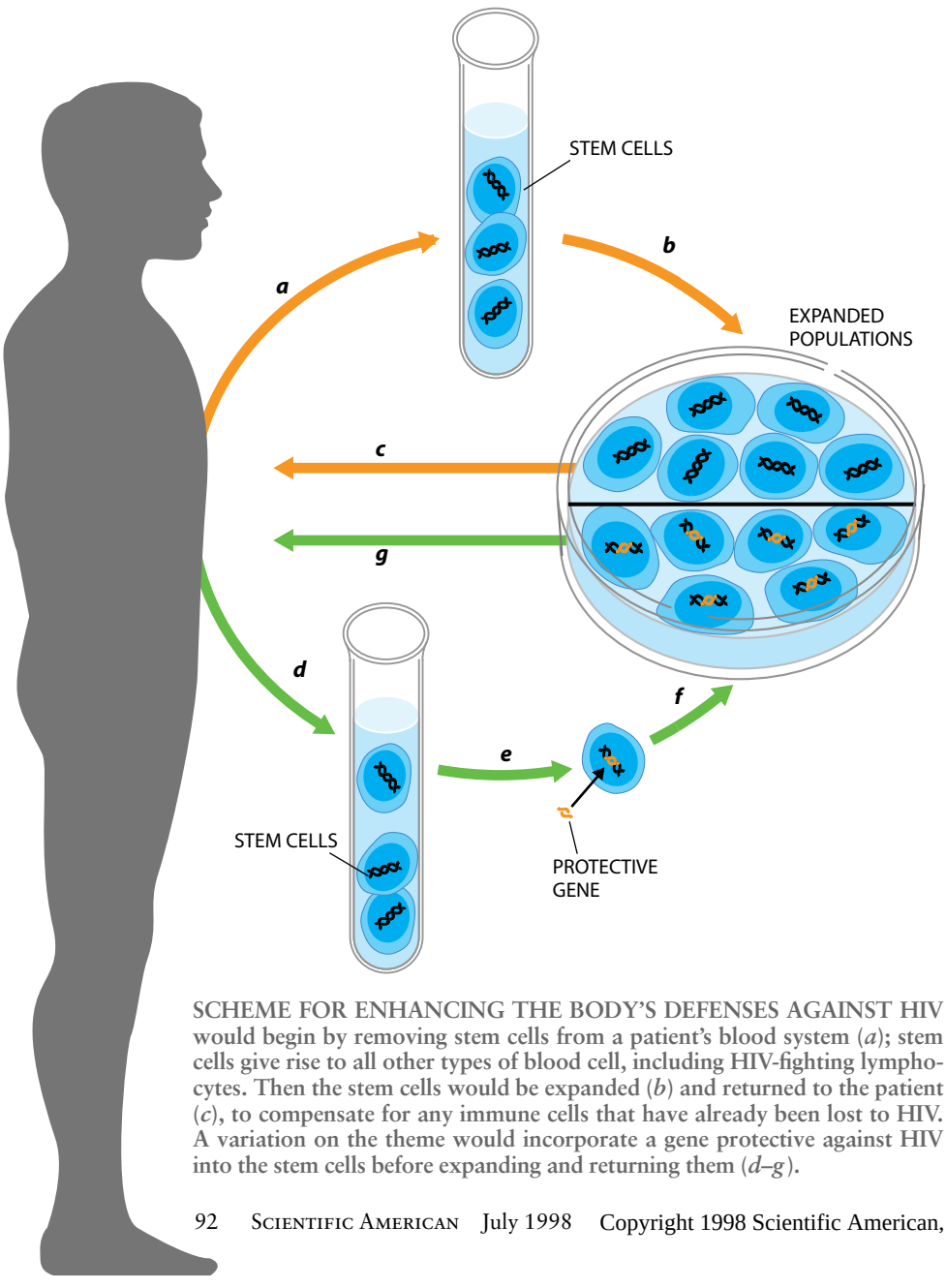
The *T* cell and viral-load thresholds, however, are fairly arbitrary. Because patients in the third group feel perfectly well, certain of them may be quite reluctant to embark on a demanding treatment plan, with its side effects and constant reminder of the disease. Further, if resistance arises and treatment fails, the patients will be left with restricted options later on. Some will therefore choose to delay therapy until tests show signs of progression or until simplified regimens or new drugs with fewer side effects become available. Most patients with better numbers—CD4 *T* cell counts over 500 and viral burdens below 10,000 to 20,000—will decide to wait as well.

For patients in whom HAART does

not quash detectable viral replication, the issue becomes, What next? Unfortunately, no second-string, or salvage, strategy seems to succeed as frequently as does HAART given to patients who have never been treated with antiretroviral drugs. Physicians generally respond by helping patients who had difficulty following the initial therapy to overcome obstacles to adherence. They then switch the entire regimen to one composed of all new drugs, taking care to avoid agents that are likely to encounter resistance.

Experience also offers limited sage advice for what to do when salvage regimens do not halt detectable viral replication. In people whose immunity has collapsed, persisting with antiretroviral treatment might do little good. For others, continuation of the regimen may be worthwhile. As the clinics at San Francisco General and Johns Hopkins have found, virological failure (incomplete viral suppression) is not the same as clinical failure (the emergence of HIV-related complications). And a few investigations have shown that even when viral levels are climbing, combination therapy can sometimes restrain HIV replication somewhat; a technically failing therapy, then, may nonetheless help maintain or elevate CD4 *T* cell levels and thus might buy patients valuable time.

Even those who do very well on HAART from the beginning pose puzzles for investigators. Scientists remain unsure whether apparently successful therapy can fully reconstitute the immune system. So far only partial restoration seems to be common, with treatment raising CD4 *T* cell levels by an average of 100 to 200 cells per cubic millimeter. The mix of CD4 *T* cell subsets may remain abnormal as well—that is, the expanding population may not recognize as many pathogens or may be less effective at combating infections than was the original *T* cell population.



SCHEME FOR ENHANCING THE BODY'S DEFENSES AGAINST HIV would begin by removing stem cells from a patient's blood system (*a*); stem cells give rise to all other types of blood cell, including HIV-fighting lymphocytes. Then the stem cells would be expanded (*b*) and returned to the patient (*c*), to compensate for any immune cells that have already been lost to HIV. A variation on the theme would incorporate a gene protective against HIV into the stem cells before expanding and returning them (*d*–*g*).

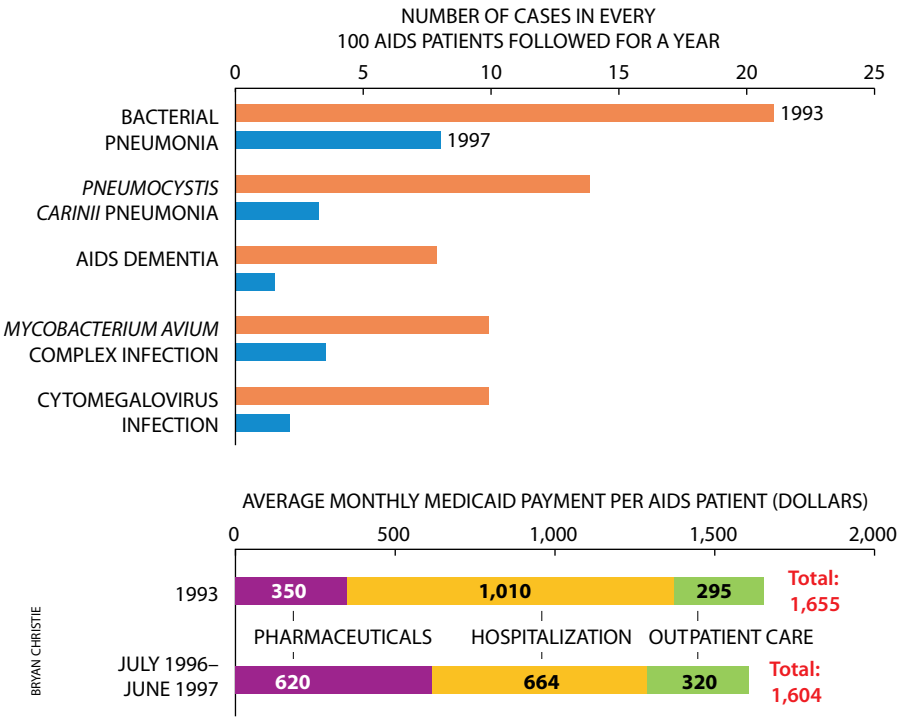
Can Reservoirs Be Eliminated?

Looking ahead, everyone hopes that initially successful HAART regimens will keep working year after year. When HAART fully stops HIV replication, it seems to prevent the development of resistance and thus should work indefinitely. But whether fact will bear out theory remains to be seen.

Investigations into whether currently available drug combinations can cure HIV infection are less encouraging. In resting (nondividing) CD4 *T* cells, HIV can apparently survive in a latent state, as an integrated provirus that produces few or no viral particles. These resting cells can be prodded to churn out new particles if something causes them to become active again. When HAART was first introduced, some scientists expressed hope that resting *T* cells would die out quickly on their own, thereby eliminating the perpetual threat of a HIV resurgence. New findings, however, suggest that certain of the cells may endure, and pose a threat, for years.

Not surprisingly, investigators are considering ways to incite the immune system to eliminate these reservoirs. To destroy the cells, the immune system has to "see" them. Normally when a cell is infected by a virus, the cell displays pieces of the microbe, or antigens, on the cell surface; then *T* cells that bear receptors for the antigens orchestrate an attack. But cells carrying a quiescent provirus in their DNA do not display HIV antigens.

One strategy for overcoming this problem is to administer substances that would purposely activate the resting cells, inducing them to make HIV particles, display antigens and arouse a killing immune response. Treatment with HAART, meanwhile, could prevent the virus from colonizing other cells. The risk, obviously, is that the strategy would backfire, allowing the virus to gain the upper hand. And even if the plan worked, other protected havens for HIV, such as brain neurons, might continue acting as lifelong sanctuaries for proviral HIV. (White blood cells known as macrophages can serve as reservoirs,



VALUE OF AGGRESSIVE HIV THERAPY, which came into wide use in 1996, is reflected in declining frequencies of several opportunistic infections in AIDS patients (as defined by CD4 *T* cell levels) at the HIV clinic of Johns Hopkins University (*top*). Although costs per patient did not decline much (*bottom*), less money had to be spent on hospital care—another sign that patients were healthier.

too, but they are relatively short-lived.)

Because current treatment plans seem unlikely to cure HIV infection within several years and because of HAART's various drawbacks—including side effects, complicated regimens and resistance—the search is on for additional drugs. The first entries to reach the market will add new choices within existing drug classes, and some will combine existing drugs into a single pill or will otherwise reduce the complexity or toxicity of current regimens. And at least two custom-designed protease inhibitors in clinical trials appear to work against strains of HIV that are resistant to existing protease inhibitors.

Other Therapeutic Ideas

Other agents under consideration would stall HIV replication in new ways. Some would block the integrase enzyme from inserting HIV DNA into a cell's chromosomes. A different strategy attempts to knock zinc out of a protein that needs the metal in order to draw HIV RNA into new particles.

Teams are also exploring ways to stop infected cells from making critical viral proteins. One such approach deploys so-called antisense DNA to inactivate two genes (*tat* and *rev*) that normally give rise to proteins needed for the efficient manufacture of other viral proteins. In rhesus macaques exposed to the simian version of HIV, this therapy limited viral replication and CD4 *T* cell depletion.

Many investigators are attempting to interdict HIV's entry into cells. Recall that to fuse with, and gain entry into, a target cell, the virus must attach both to CD4 and to a co-receptor on the cell surface. Initial attempts to interfere with binding to CD4 were disappointing, but new possibilities have recently come to light, and many groups are examining compounds that might sheath the HIV-docking site on co-receptors to keep HIV at bay [see "In Search of AIDS-Resistance Genes," by Stephen J. O'Brien and Michael Dean; Scientific American, September 1997].

A number of scientists are focusing on the other part of the equation: the immune system. They are trying to augment the body's remaining forces or to restore lost powers. For instance, some patients are receiving low doses of a biological molecule called interleukin-2, which enhances the proliferation of *T* lymphocytes. It is also hoped that interleukin-2 can help force immature "stem cells" to spawn a full repertoire of fresh new cells of the immune system, including *T* lymphocytes and antibody-producing cells that can recognize and eliminate HIV.

Immune-reconstitution efforts are taking additional forms as well—such as harvesting stem cells from an HIV-infected patient, multiplying them in the laboratory and then returning the enlarged cell population. A twist on this scheme would supply the infused cells with a gene that would protect them from later succumbing to HIV.

Other approaches being pondered might kill HIV-infected cells without relying on the immune system. One would deliver a virus that had been genetically altered to enter only HIV-producing cells, leaving HIV-free cells untouched. It would enter by attaching to viral proteins displayed on a colonized cell and would then destroy the cell.

As time goes by, patients infected with HIV are sure to have more and more options for therapy. Specific recommendations may change from time to time, but the principles of therapy, now grounded in solid research, will remain steadfast: until there is a cure or a vaccine, controlling HIV replication offers the best chance for a long, productive life.

The fact that such insight and the tools to suppress HIV reproduction are available is both astonishing and wonderful. The therapeutic advancement achieved since late 1995 has few parallels in the history of medicine, save perhaps for the revolution sparked by the introduction of penicillin. Just three years ago those of us who cared for HIV patients mainly devoted our time and energy to easing symptoms and preparing patients to die. Now we help patients to live. The war against HIV is far from over, but the success of aggressive therapy is certainly a victory to be savored. SA

The Authors

JOHN G. BARTLETT and RICHARD D. MOORE are colleagues at the Johns Hopkins University School of Medicine. Bartlett is professor of medicine, chief of the division of infectious diseases and director of the AIDS service at the university. He is also co-chair of the panel that wrote and updates the U.S. guidelines for using anti-HIV drugs in adults and adolescents. Moore is associate professor of internal medicine, infectious diseases and clinical pharmacology at Johns Hopkins and is director of the program in pharmacoepidemiology and pharmacoeconomics there.

When Children Harbor HIV

by Catherine M. Wilfert and Ross E. McKinney, Jr.

HIV infection is particularly difficult to combat in the young

Children infected with HIV face more obstacles to staying well than adults do. The virus tends to behave more aggressively in the young, leading more rapidly to the immune dysfunction of AIDS and to death. Fewer anti-HIV drugs are available for children (those younger than 13 years), and pediatricians have inadequate child-based data on which to construct treatment plans. Moreover, the therapeutic strategies thought best from a medical standpoint can be difficult for families to carry out.

HIV-infected children almost always contract the virus from their mothers, during gestation or through breast-feeding. Roughly two thirds acquire the virus at delivery or in the days leading up to birth. Worldwide, more than a million children are infected now, and about 1,500 babies join the afflicted every day, primarily in developing nations.

HIV's aggressiveness in children becomes evident quickly. Infants typically become symptomatic during the first year of life—much faster than adults, who commonly feel fine for several years. Similarly, a large fraction of infected babies—up to 16 percent—die before their fourth birthday because of rapid destruction of the immune system and development of many of the same opportunistic infections that can fell grown-ups. Very few adults succumb that quickly.

Beyond becoming sicker more quickly, children can also show effects not observed in older patients. HIV can invade the young brain early in the disease course. Because this organ is still immature and evolving in children, the invasion may impair intellectual development and motor functioning and lead to coordination problems. These effects can be permanent. Adults may suffer brain dysfunction, too, but usually not until much later.

In addition, physical growth may be retarded (both height and weight), even in the absence of other symptoms. Effective antiviral therapy brings the growth rate back to normal, however—an observation that allows changes in height and weight to serve as indirect indicators of whether a treatment plan is controlling the virus.

All children contract more bacterial infections than adults do, simply because they have not had the years of exposures that build up immunity. When the immune system is depressed by HIV, vulnerability increases. Consequently, about 20 percent of pediatric AIDS victims battle serious, recurring bacterial infections, such as meningitis or pneumonia. Some undergo repeated bouts of viral infections, such as chicken pox, that rarely recur in their peers.

Scientists now have some sense of why HIV becomes destructive more swiftly in youngsters. Children often have high viral loads—large concentrations of HIV RNA (viral genes) in their blood—an indication that the amount of infectious virus in the system is also high. Adults generally have comparably elevated viral loads soon after contracting HIV, but the levels drop within months, whereas those in infants remain elevated for years. Persistently high virus burdens, which can be harder to reduce than small ones, are associated with faster disease progression and shorter survival.

CHILD TAKES ANTI-HIV medicine through a needleless, syringelike device. Children have fewer drug choices, in part because they often need liquids; certain anti-HIV compounds taste terrible or are difficult to formulate as syrups.

JEFF SCHEID *Gamma Liaison Network*

Treatment Dilemmas

Aside from the inherent challenges of treating patients who harbor large amounts of HIV in their system, therapy for children can be problematic for other reasons. Logic suggests that the optimal treatment approaches for adults should work best in children. In other words, children, like adults, should ideally receive combinations of three or four different drugs, including one or more from the class known as the protease inhibitors. (Single anti-HIV drugs have limited value when used on their

own, and they promote viral resistance that undermines the drugs' effectiveness.) Yet this understanding is difficult to translate into effective action.

Part of the problem is that even in the industrial nations, where health care systems and medical supplies are extensive, fewer drugs are available for children than for adults. Because many children cannot swallow pills, they often need liquids or syrups, but certain anti-HIV compounds are insoluble in water or taste sickening.

Further, drug companies historically have waited to test medications on children until the agents proved effective in adults. Aside from delaying drug approvals, the paucity of studies in children has meant that pediatricians frequently resort to guesswork in choosing drug dosages for children, who must be given amounts calibrated to body size. Fortunately, the tide is changing. New anti-HIV drugs are being tested more rapidly in children, and ongoing studies are finally examining three- and four-agent drug regimens analogous to the aggressive treatment plans of adults.

If the problems of drug availability and dosing were resolved, the intensive regimens would still be extraordinarily demanding. Three or more agents must be given two or three times a day, every day, 365 days a year. And doses cannot be missed, or HIV can become resistant to the medications.

Anyone who has had to treat a generally well infant for a simple ear infection knows how easy it is to forget a dose. Sadly, families of HIV-positive children are often less equipped than most to cope with the stringent demands of intensive anti-HIV therapy. They may be poor, disorganized or headed by a single mother who is ailing herself. Families can also be reluctant to tell others about a child's HIV status. They may therefore allow children to skip having medicine at day care or school. Health professionals who treat these children may thus face the awful task of considering with a family whether aggressive combination therapy can be handled or whether a less optimal drug regimen should be used.

Now a new therapeutic quandary has emerged. In the past, treatment for infants was sometimes delayed for weeks or months after birth. Recent findings in adults suggest, however, that beginning combination therapy promptly after diagnosis is better for limiting viral replication and for establishing immunity.

That conclusion in itself is not especially problematic. But the research also implies that starting powerful therapy immediately at birth—when HIV is usually contracted—can potentially prevent infection or significantly delay progression to AIDS in infected infants. Unfortunately, physicians often cannot confirm infection until perhaps two weeks after delivery, when the virus reaches detectable levels in the blood. To best protect affected infants, physicians might have to prescribe intensive therapy for all neonates born to HIV-positive women, including for the majority who normally evade HIV infection. In view of the drawbacks of aggressive anti-HIV therapy—complexity, toxicity and expense—many doctors are understandably reluctant to take that course.

Prevention: For Too Few

As always, prevention is the best medicine. The approach used to block mother-to-infant (vertical) transmission in the industrial nations grew out of a joint U.S.-French clinical trial called the Pediatric AIDS Clinical Trials Group Protocol 076. In that landmark 1994 study, HIV-infected women took zidovudine (AZT) orally from as early as the 14th week of pregnancy. During delivery, they were given the drug intravenously. The newborns were started on zidovudine immediately and kept on it for six weeks. Whereas 26 percent of infants born to untreated mothers acquired HIV infection, only 8 percent of infants in the treatment group suffered that fate. Neither mothers nor infants showed any untoward effects from the therapy.

Since 1994, widespread application of the 076 protocol has dramatically reduced mother-child transmission in the U.S. and parts of Europe. To protect women as well as their babies, a group of experts recommends that combination therapy including zidovudine should be offered to pregnant women, who must, however, be informed of gaps in data on safety, toxicity and improved efficacy.

Exactly how zidovudine blocks vertical transmission is a mystery. Given alone, the drug will not lower viral concentrations profoundly. Hence, although women with lower viral loads are less likely (albeit still able) to pass the virus to their babies, the medicine is not helping by markedly reducing viral levels in this case. The drug might work primarily by blocking infection during delivery, when virus-laden maternal secretions can infect the mucosal surfaces of the infants' eyes, mouth and gastrointestinal tract and gain access to the bloodstream.

The picture in the developing nations is considerably bleaker than in the U.S. Because women in those countries typically have limited or no access to prenatal care, the expensive 076 protocol has not been feasible. A simpler intervention is therefore essential—ideally consisting of a single dose of medicine at labor to the mother and, possibly, one to her baby. If the treatment were inexpensive enough, it could be given without testing for HIV, which would be valuable because in many areas testing is too expensive or leads to stigmatization. (A woman infected with HIV may be ostracized or abandoned by her husband if her condition is revealed, even when he is the source of the disease.)

Studies are under way in the developing world to assess new preventive approaches [see "Coping with HIV's Ethical Dilemmas," on page 106]. Some involve shorter courses of zidovudine. Others rely on different drugs.

Earlier this year one of those trials—in Thailand—found that oral zidovudine given for an average of 25 days at the end of pregnancy and throughout labor and delivery can reduce the transmission rate by half. Unfortunately, both the reduced course of zidovudine and the HIV testing that has to precede it are still beyond the grasp of many people.

Physicians who provide care for children or pregnant women infected with HIV continue to be frustrated by troubling contradictions. Methods proved to limit mother-child transmission are available but are too complicated and costly for universal use in the countries with the most pediatric HIV disease. In industrial countries, aggressive therapeutic regimens akin to those now helping adults could extend life for youngsters, but many children lack the supports they require for adhering to those treatment programs. In rich and poor nations alike, the need for simpler approaches to prevention and treatment is urgent. SA

The Authors

CATHERINE M. WILFERT and ROSS E. McKINNEY, JR., are at Duke University Medical Center, where Wilfert is professor emerita of pediatrics and microbiology and McKinney is acting chief of the Division of Pediatric Infectious Diseases and associate professor. Wilfert is also scientific director of the Pediatric AIDS Foundation.

Preventing HIV Infection

by Thomas J. Coates and Chris Collins

Altering behavior is still the primary way to control the epidemic

With a vaccine still years away, the only broadly applicable way to prevent new HIV infections is to change behaviors that enable transmission of those infections—especially behaviors relating to sex and drug injection.

Because most people simply will not choose celibacy, realistic public health workers have focused on encouraging adoption of safer sexual practices, notably condom use. That people can be persuaded to employ safer sex is well illustrated by the experience of San Francisco's gay community in the 1980s. Perhaps 8,000 individuals became infected with HIV in both 1982 and 1983. That figure declined to 1,000 a decade later and is now estimated at less than 400 a year. The primary reason for this decline was a precipitous decrease in unprotected anal intercourse as a result of education about safer sex practices.

Targeted education aimed at a particular at-risk community is a prime way to persuade people to engage in preventive practices. In San Francisco, information about HIV transmission and safer sex was made available in the media and at centers of gay society, such as churches, gay organizations and clubs. Programs aimed at the commercial sex industry have greatly lowered the risks of HIV transmission for both worker and client. In Thailand the Ministry of Public Health has attempted to inspire 100 percent condom use in brothels. It provides condoms and advocates safer sex practices through the media. From 1991 to 1995 the number of men who wore condoms when with prostitutes at brothels rose from 61 to 92.5 percent. HIV infection rates among Thai army conscripts fell from 12.5 percent in 1993 to 6.7 percent in 1995 because condom use was effective and also in part because fewer men employed prostitutes.

Testing and follow-up counseling reduce risk behaviors among both infected and uninfected individuals, as has been documented in a large study in three developing countries. A number of research efforts involving "discordant" heterosexual couples—where only one partner is HIV positive—have shown that counseling following a positive test can strongly protect the uninfected partner. In Rwanda, condom use in discordant couples who received counseling increased from 3 to 57 percent. In Congo (the former Zaire), the increase was from 5 to 77 percent.

MARILYN HUMPHRIES *Impact Visuals*

PROTESTERS attempt to educate the public. Federally funded studies have shown that needle exchanges help to lower HIV transmission without increasing drug use.

Comprehensive sex education in schools can promote safer sex while actually decreasing sexual activity among young people. A review of 23 school programs found that teens who received specific information and training about how to insist on condom use were less likely to engage in sex. Those who did have sex had it more safely and less frequently than those not exposed to AIDS-specific educational material. Adolescents not yet sexually active who receive information about HIV have their first sexual experiences later in life and have fewer sex partners than students who receive HIV information after they have begun having sex.

Peer influence and community action are excellent complements to more general education. In one investigation, researchers identified popular opinion leaders among gay men in small cities. The researchers trained these "trendsetters" to endorse safer sex practices among their friends and acquaintances. The number of men engaging in unprotected anal intercourse dropped by 25 percent in only two months; condom use went up 16 percent; and 18 percent fewer men had more than one sex partner. In two similar cities without such peer influences, no changes occurred. In another trendsetter study the rate of unprotected intercourse among young gay men fell by more than 50 percent. Such a drop in risk behavior could actually reduce the HIV transmission rate enough to stop the epidemic in that population.

Advertising and marketing can also change a community norm, making condoms more acceptable. A mass-media campaign advocating safer sex in Congo caused condom sales to increase from 800,000 in 1988 to more than 18 million by 1991. A local survey found that those claiming they practiced mutual fidelity went from 29 to 46 percent in a one-year period. An aggressive marketing campaign aimed at 17- to 30-year-olds in Switzerland saw condom use with casual sex partners rise from 8 to 50 percent between 1987 and 1991. Among 17- to 20-year-olds the figure went from 19 to 73 percent. Critics of the frank presentation of sexually oriented materials decry the potential for increasing rates of sexual activity. The Swiss study, however, found such rates to be unchanged—only the safety of the practices increased.

Easing access to condoms is another way to increase their use—both by giving them away and by making them

less embarrassing to buy. A study published in 1997 in the *American Journal of Public Health* reported that when condoms were made available in high schools, usage went up without an increase in number of sex partners or a lowering of the age of initiation of sexual activity. At a drug-abuse treatment center, condoms were almost five times as likely to be taken from private rest rooms as from a public waiting area. Clearly, the perception of privacy encourages the acquisition of condoms.

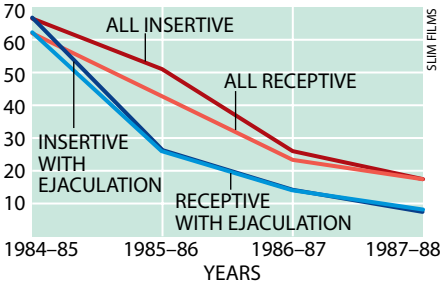
Physician-patient dialogue may also help reduce risk behaviors, but doctors have squandered valuable opportunities. A recent study found that only 39 percent of adolescents ever talked with their doctors about avoiding HIV, and only 15 percent discussed their sex lives; however, almost 75 percent said they would trust their doctors with information about their sex habits, and up to 90 percent said they would find it helpful to talk about sex with a doctor. Ninety-four percent of physicians ask about smoking habits; frank discussions of sex are no less appropriate in a doctor-patient setting.

Drug treatment should be a first-line approach to reducing risk for HIV and other infectious diseases in intravenous drug users. Substitution strategies, such as methadone treatment for heroin addiction, clearly reduce transmission of HIV through needle sharing.

Access to clean needles can help protect those still using injection drugs. Exchange programs, despite the controversies they elicit, have been shown to lower the risk of viral infection in many studies worldwide. Six U.S. government–funded studies have found that needle exchanges help to reduce HIV transmission without leading to greater drug use. Some jurisdictions have expanded beyond needle exchange. In 1992 Connecticut began a model program in which pharmacists were permitted to sell and individuals were allowed to possess up to 10 syringes without medical prescriptions. Among users who reported ever sharing syringes, sharing dropped from 52 to 31 percent, and street purchases of syringes dropped from 74 to 28 percent. Fears of encouraging drug abuse have proved unfounded: many studies have shown that needle availability does not increase the use of illegal drugs.

Direct outreach to drug users is also effective. A program supported by the National Institute on Drug Abuse followed 641 injection addicts, consistently encouraging them to seek treatment to get off drugs and use safe injection methods in the interim. After four years, 90 individuals had contracted HIV, only half the statistical expectation for that population.

PERCENTAGE OF GAY AND BISEXUAL MEN ENGAGING IN UNPROTECTED ANAL INTERCOURSE IN SAN FRANCISCO



SOURCE: Ekstrand et al. in *American Journal of Public Health*, Vol. 80, No. 973; 1990

DECLINE IN RISK BEHAVIOR was significant and fast because of targeted education. Reaching individual cultural groups in a manner acceptable to those populations pays prevention dividends.

What Does Not Work

One-time exposure to information is less successful than interventions that teach skills and reinforce positive behaviors repeatedly. Young people in particular need to learn exactly how to use condoms and how to be assertive about demanding their use before they will modify their behavior significantly.

A single message is insufficient to reach the multiple diverse communities grappling with the AIDS epidemic. Educational approaches must be tailored to fit the ethnicity, culture and sexual preference of a given population. The San Francisco example of outreach aimed at the gay community attests to the success of this focused approach.

Abstinence-only programs do a disservice to America's youth. Congress recently approved $250 million for five years of sex education restricted to discussions of abstinence alone. Such efforts cater to a political agenda more than any societal realities—two thirds of high school seniors say they have had intercourse. Educational programs, while encouraging abstinence, must provide the knowledge and means to protect the young from HIV.

Coercive measures to identify people with HIV or their sexual partners are likely to backfire. In this age of promising HIV therapies, it is important that infected individuals enter care as soon as possible after diagnosis. Early therapy can also prevent pregnant mothers from passing HIV to their children. But mandatory testing and the threat of coercive measures to identify sexual contacts undermine faith in and comfort with the health care system. A 1995 survey in Los Angeles found that 86 percent of those responding would have avoided an HIV test if they knew their names would be given to a government agency. Expanding opportunities for anonymous and confidential testing can bring more people into care and counseling.

Settling for the status quo is also a threat to prevention. Investigators need to continue to develop and refine interventions that can reach groups at risk. Women in particular need approaches that protect them from infected partners. With access to female condoms, their rates of sexually transmitted diseases are lower than when only male condoms are available. Better microbicides would likewise protect women whose partners are unwilling to practice safer sex.

Prevention is in many ways a less exciting topic than the development of wonder treatments or a vaccine. Yet effective behavioral and policy interventions are the best tools available to address an epidemic in which 16,000 people become infected worldwide every day. Concerted research on HIV vaccines must continue. Yet even when a vaccine is available, it most likely will not confer 100 percent protection on all those vaccinated. Distribution of the vaccine to everyone in need is another obstacle on the road to full protection.

Therefore, behavioral intervention will continue to play a role in bringing the global HIV epidemic under control and is indeed a matter of life and death. As June Osborn, former chair of the National Commission on AIDS, has said, "If we do preventive medicine and public health right, then nothing happens and it is very boring. We should all be praying for boredom." SA

The Authors

THOMAS J. COATES and CHRIS COLLINS were colleagues at the Center for AIDS Prevention Studies at the University of California, San Francisco. Coates remains the director of the center and of the AIDS Research Institute there. He has advised the National Institutes of Health's Office of AIDS Research, the World Health Organization and the U.S. Congress. Collins, a former policy researcher at the AIDS center, now writes on AIDS prevention policy from Washington, D.C.

HIV Vaccines: Prospects and Challenges

by David Baltimore and Carole Heilman

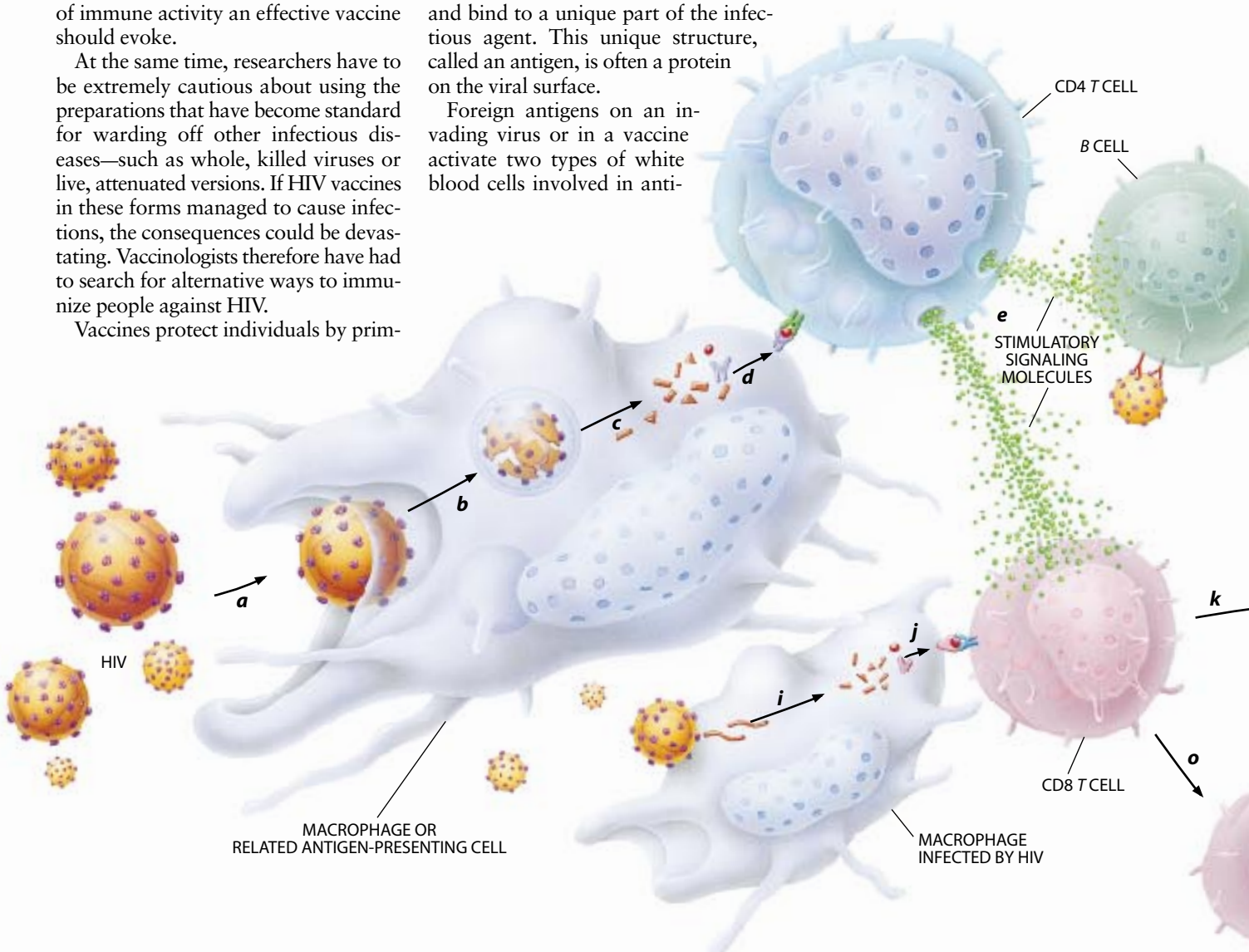
Scientists know more about HIV—the human immunodeficiency virus that causes AIDS—than any other virus. Yet designing a vaccine able to protect against it remains as much of a challenge today as when the virus was discovered. Part of the problem is that, unlike the body's response to most acute viral infections, the natural immune response does not destroy HIV. This failure makes it difficult for investigators to know what type of immune activity an effective vaccine should evoke.

At the same time, researchers have to be extremely cautious about using the preparations that have become standard for warding off other infectious diseases—such as whole, killed viruses or live, attenuated versions. If HIV vaccines in these forms managed to cause infections, the consequences could be devastating. Vaccinologists therefore have had to search for alternative ways to immunize people against HIV.

Vaccines protect individuals by priming the immune system to recognize disease-causing organisms when they are encountered. In the case of HIV, a successful vaccine should be able to eliminate incoming virus and destroy quickly any cells that become infected.

Most vaccines activate what is called the humoral arm of the immune system, stimulating formation of protective antibodies: molecules that mark free virus (which circulates outside cells) for destruction. The antibodies recognize and bind to a unique part of the infectious agent. This unique structure, called an antigen, is often a protein on the viral surface.

Foreign antigens on an invading virus or in a vaccine activate two types of white blood cells involved in antibody manufacture. After contacting antigens, cells known as *B* lymphocytes mature and produce antibodies. In addition, helper, or CD4, *T* lymphocytes direct *B* cells to manufacture more antibodies or to take the form of memory *B* cells. The memory cells do not produce antibodies immediately but respond vigorously to subsequent exposures. Following vaccination, the long-term production of small amounts of antibody and the persistence of memory

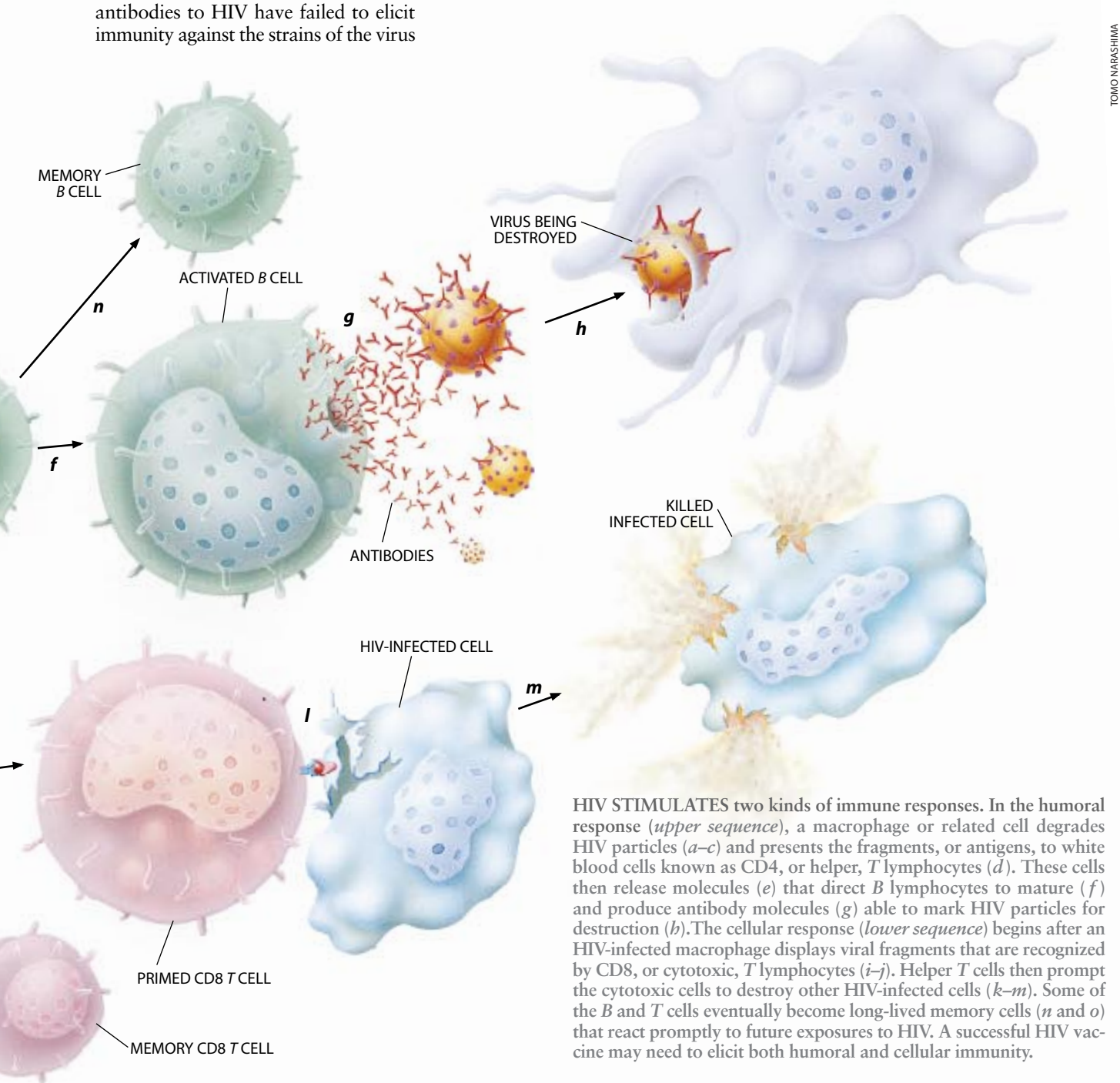


Unlike vaccines for many viruses, those for HIV may have to go beyond generating antibodies. Devising approaches that will fully activate the immune system is far from simple

cells allow the body to mount a rapid defense if ever it encounters the virus.

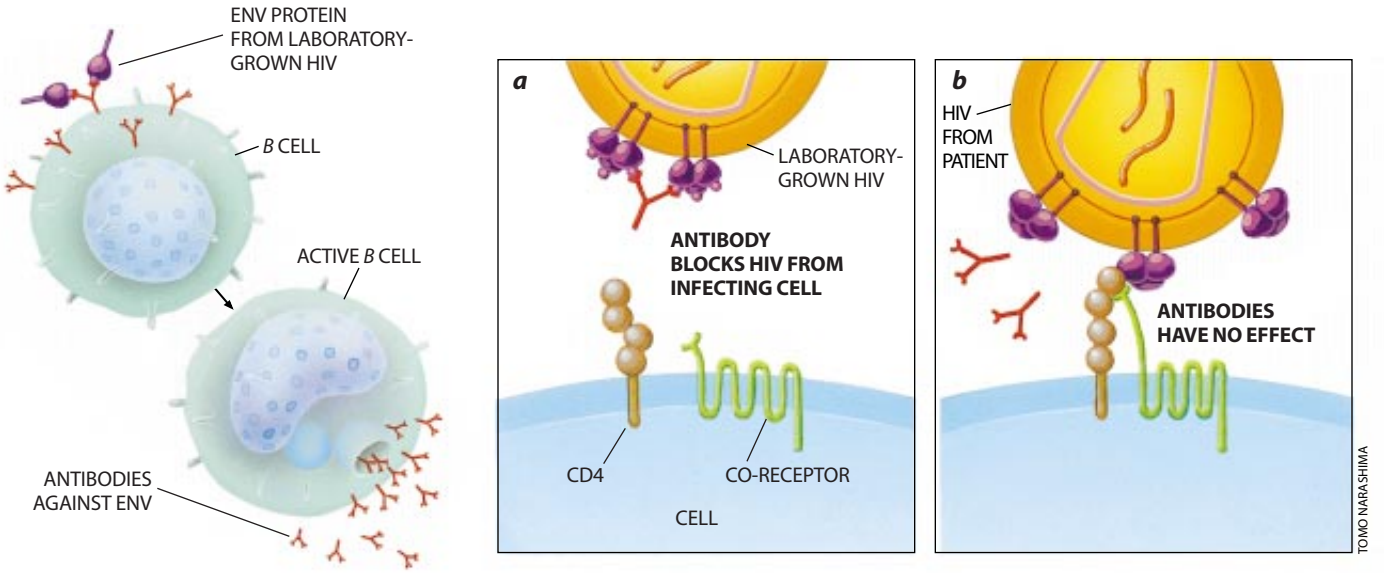
No vaccines have been designed specifically to stimulate the other arm of the immune system, known as the cellular component. But many AIDS researchers are working on just that aim because, thus far, vaccines designed to generate antibodies to HIV have failed to elicit immunity against the strains of the virus commonly found in infected patients.

In cellular immunity, activated white blood cells called cytotoxic *T* lymphocytes (CD8 *T* cells) multiply and cruise through the bloodstream and tissues, searching for and eliminating virus-infected cells. Some also become memory cells, ready to leap into action after a later exposure to a pathogen. Unlike antibodies, cytotoxic *T* lymphocytes recognize infected cells, rather than the infectious agent itself. Like *B* cells in the humoral arm of the immune system, however, cytotoxic *T* cells are activated in part by signals from helper *T* cells. In



TOMO NARASHIMA

HIV STIMULATES two kinds of immune responses. In the humoral response (*upper sequence*), a macrophage or related cell degrades HIV particles (*a–c*) and presents the fragments, or antigens, to white blood cells known as CD4, or helper, *T* lymphocytes (*d*). These cells then release molecules (*e*) that direct *B* lymphocytes to mature (*f*) and produce antibody molecules (*g*) able to mark HIV particles for destruction (*h*).The cellular response (*lower sequence*) begins after an HIV-infected macrophage displays viral fragments that are recognized by CD8, or cytotoxic, *T* lymphocytes (*i–j*). Helper *T* cells then prompt the cytotoxic cells to destroy other HIV-infected cells (*k–m*). Some of the *B* and *T* cells eventually become long-lived memory cells (*n* and *o*) that react promptly to future exposures to HIV. A successful HIV vaccine may need to elicit both humoral and cellular immunity.



PURE ENV PROTEIN, isolated from virus grown in the laboratory, has been studied as a vaccine. The protein successfully induced *B* lymphocytes to make antibodies that recognized the Env protein (*left panel*). Further, the antibodies prevented laboratory-grown HIV from infecting cultured cells (*a*), perhaps by blocking binding to cell-surface receptors or by enhancing elimination of the virus. Disappointingly, though, those antibodies have not been able to bar infection by virus isolated directly from patients (*b*).

the long run, the most effective HIV vaccines may well be the ones that stimulate both the humoral and cellular arms of the immune system, generating antibodies and activated cytotoxic *T* cells.

Efforts to design an HIV vaccine that maximizes production of antibodies or stimulation of cytotoxic *T* cells have been hampered by a lack of basic knowledge about how the immune system functions. Until investigators can learn how to induce the body to generate and maintain memory cells and cytotoxic *T* cells, those attempting to develop HIV vaccines will have to rely on a certain amount of trial and error, hoping to hit on an approach that will work.

The Antibody Approach

Vaccines that stimulate the production of protective antibodies have proved successful for combating diseases such as poliomyelitis, measles and influenza. At present, the most extensively tested HIV vaccine candidates contain some part of the envelope protein (Env), the molecule that coats the surface of the virus. Because the virus uses Env as a kind of key for gaining entry to human cells, generating antibodies that attach to the business end of this protein should prevent HIV from binding to and infecting cells.

The Env protein, also called gp160, is actually an association of two units: gp120, a sugar-shrouded protein that juts out of the virus membrane and interacts with receptors on the surface of human *T* lymphocytes, and gp41, the small protein that anchors gp120 to the membrane. Both gp120 and gp160 have been tested as HIV vaccine candidates in human volunteers.

In tests, the proteins elicited the production of antibodies, a result that raised hopes that they might form the basis for an effective HIV vaccine. Further, the resulting antibodies effectively neutralized live HIV in a test tube, blocking its ability to infect cultured human lymphocytes.

Unfortunately, the antibodies only recognized strains of HIV that were similar to those used to generate the vaccines. The gp120 and gp160 proteins in the preparations were made from HIV strains that had been cultivated in the laboratory. The antibodies elicited against proteins from such lab-adapted virus strains were ineffective at neutralizing HIV strains isolated directly from infected patients; the isolates were quite able to infect cultured cells.

Why did the antibodies fail to neutralize the HIV obtained directly from patients? The structure of the Env protein in laboratory-grown strains appears to be somewhat looser than that of the surface protein in patient isolates; those in isolates are folded tightly. Antibodies to laboratory strains of HIV may recognize parts of the Env protein that are not normally exposed in viruses from patients, probably because the recognition sites are buried within the more folded form of the protein. Antibodies to laboratory-grown virus, then, would not "see" their targets on HIV isolated from patients.

Researchers are currently developing vaccines based on surface proteins prepared from patient isolates. Such preparations may present Env in the conformation found in patients. Yet even these vaccines may not work. The Env protein on such isolates may be very densely packed and highly camouflaged by sugars. As a result, *B* cells may be unable to find many antigens and so may produce relatively few kinds of antibodies. Such an outcome would be consistent with the finding that people who are infected with HIV generally produce a limited repertoire of antibodies that react with the surface of HIV.

When Env binds to a cell, the protein changes its shape somewhat. A vaccine that duplicates the conformation adopted by gp120 as it attaches to receptors on the cell surface may succeed best at raising antibodies able to block HIV from infecting human cells.

Individuals who are infected with HIV but remain healthy and keep viral replication in check may offer some hope for guiding the design of an effective HIV vaccine. Some of these long-term survivors make a very small amount of antibody, which, when isolated, can neutralize HIV from patient isolates. Further, those antibodies can neutralize viruses from many different patient isolates—a necessity for an AIDS vaccine that will be effective against a broad spectrum of HIV strains. Unfortunately, even these antibodies may not be the whole answer. Tests of cells in culture indicate that the antibodies must be present at surprisingly high concentrations to block HIV entry into cells effectively.

Pure protein vaccines may not be the

best way to stimulate antibody production: in isolation, gp120 does not appear to have a precise conformation, and gp160 clumps into an ineffective aggregate. To get around these difficulties, researchers are currently testing two different vaccine strategies designed to present the Env proteins in a more natural conformation.

One plan of attack involves using whole, killed virus particles. This disabled form of HIV, incapable of multiplying, might present the immune system with more natural forms of Env proteins. With a better target, *B* cells might produce a better quality and a higher quantity of protective antibody.

Making a killed-virus vaccine requires a rigorous inactivation procedure, because residual virus and even residual viral genetic material could potentially be dangerous. Harsh treatment makes the vaccine less effective, however; the inactivation process can cause HIV to shed its weakly attached gp120. Many researchers have therefore been moving away from this design, although the gp120 stability problem may ultimately be solvable.

Env proteins can also be presented to the immune system embedded in "pseudovirions," artificial structures that resemble virus particles. These empty lipid shells could be made to carry nothing but gp160. Pseudovirions would be safer than whole, killed virus, because they lack the genes that could propagate HIV infection. Unfortunately, pseudovirions are very difficult to manufacture and produce in a stable form. Researchers hope, however, to have sturdier versions ready for safety trials in humans shortly.

Recruiting Cytotoxic *T* Cells

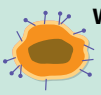
Different vaccine strategies are required to generate activated cytotoxic *T* lymphocytes. Although surface proteins or even whole, killed virus particles can elicit antibody production, they are poor stimulants of cellular immunity. Cytotoxic *T* cells recognize short pieces of foreign protein that appear on the surface of an infected cell. Infected immune cells generate these antigenic peptides as they digest samplings of viral proteins—surface proteins such as Env as well as the internal proteins that drive viral reproduction and assembly. A carrier protein then escorts the protein fragments to the cell membrane, where they are displayed on the outside of the cell.

For an HIV vaccine to stimulate cell-based immunity, it must direct selected cells to synthesize and display one or more peptides from the proteins normally made by the virus. These cells would trick the body into mounting an immune response against all cells displaying the viral peptides, including ones truly invaded by HIV.

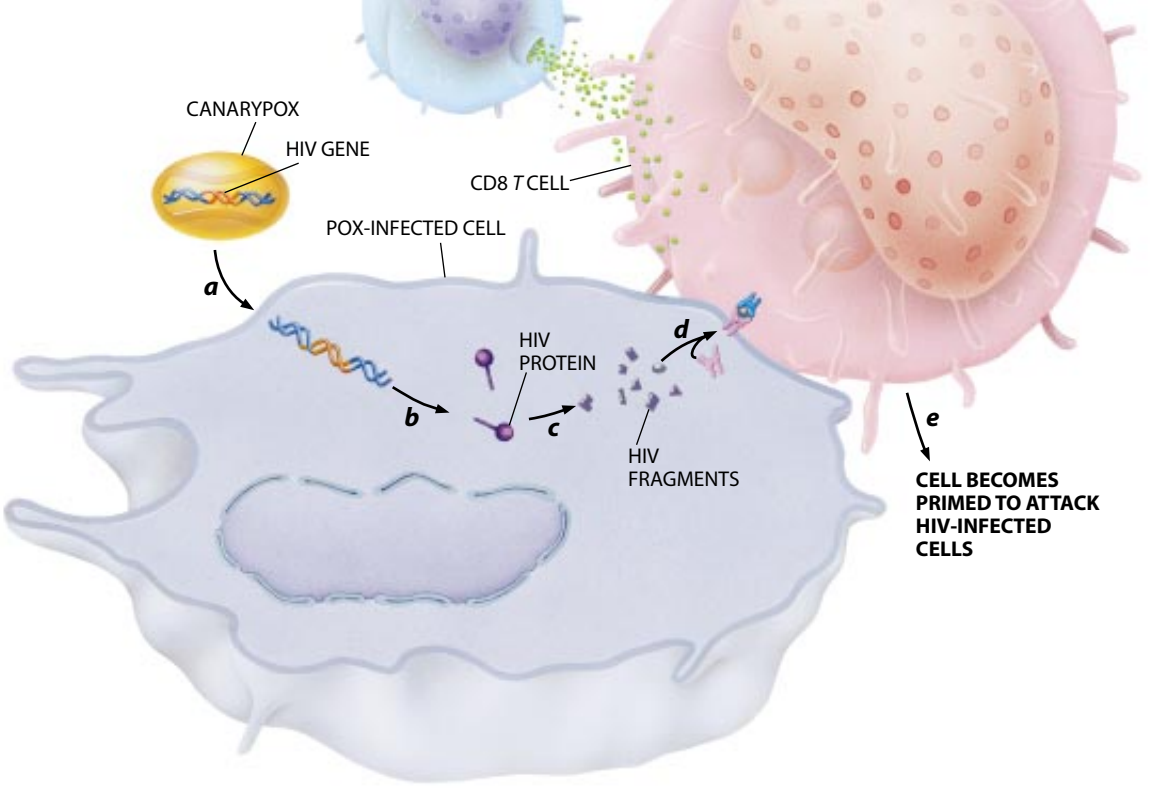
The Sabin vaccine against polio, which consists of a live poliovirus, turns out to evoke cytotoxic *T* cell activity against polio-infected cells, yet it does not cause polio, because the virus has been weakened in the laboratory by certain genetic mutations. So far, though, no mutations have been identified that will transform HIV into a vaccine that will be completely safe.

Investigators are, however, developing other methods for inducing cells to produce and display HIV proteins. One approach, construction of a so-called live vector vaccine, takes advantage of

Vaccine Strategies under Study

Vaccine Constituents	Status	Advantages	Disadvantages
Vaccines Eliciting Anti-HIV Antibodies			
 **Viral surface proteins,** such as gp120	In phase I and II trials, which examine safety	Safe and simple to prepare	Vaccine-elicited antibodies have failed to recognize HIV from patients
 **Whole, killed HIV**	Not under study in humans	Should present HIV surface proteins in a relatively natural conformation; simple to prepare	Slight risk that preparations might include some active virus; inactivated virus might shed its proteins and become ineffective
 **Pseudovirions** (artificial viruses)	Close to phase I trials	Present HIV surface proteins in a relatively natural conformation	Difficult to produce and to ensure long-term stability
Vaccines Eliciting Cellular Responses			
 **Live vector viruses** (non-HIV viruses engineered to carry genes encoding HIV proteins)	In phase II trials	Makers can control amount and kinds of viral proteins produced	Complicated to prepare; current vaccines elicit modest immune response
 **Naked DNA** containing one or more HIV genes	In phase I trials	Simple and inexpensive to prepare	Some worry that integration of HIV genes into human cells could harm patients
 **HIV peptides** (protein fragments)	In phase I trials	Simple to prepare	Do not elicit strong immune response
Vaccines Eliciting Antibody and Cellular Responses			
 **Combinations of elements,** such as pure gp120 protein plus canarypox vector	In phase II trials	Should stimulate both arms of the immune response at once	Complicated to prepare
 **Live, attenuated HIV**	Not under study in humans; being assessed in nonhuman primates	Most closely mimics HIV; may interfere with infectious HIV's ability to replicate	Virus could potentially cause disease

SLIM FILMS



TOMO NARASHIMA

RECOMBINANT CANARYPOX VACCINE is among those being studied as a way to elicit cell-based immunity against HIV. Such vaccines deliver HIV genes to human cells (*a*). The viral genes are translated into proteins (*b*), which are subsequently digested into fragments (*c*) and displayed on the cell surface (*d*). These fragments stimulate HIV-specific cytotoxic, or CD8, *T* lymphocytes, thereby priming them to kill any cells that may actually be infected with HIV (*e*).

the ability of different viruses to invade cells. Researchers insert selected HIV genes into a virus that is not harmful and then allow the benign virus, or vector, to deliver the DNA to cells in the body. Because genes are the blueprints for proteins, the infected cells will produce HIV proteins. These viral proteins are then chopped and shipped to the cell surface, where they can attract the attention of wandering cytotoxic *T* lymphocytes. The *T* cells, in turn, should multiply in response to the antigenic stimulation and stand ready to kill any cells that actually become infected with HIV.

The most extensively tested live vector vaccines are based on the canarypox virus. This nonpathogenic relative of the smallpox virus enters human cells but is incapable of assembling new viral particles. Researchers have engineered canarypox viruses to deliver the genes that direct the production of Env and gp120 and a variety of nonsurface HIV proteins, such as Gag (the core protein) and protease.

To date, the canarypox vaccines tested in humans have proved safe and have elicited modest cytotoxic *T* cell–based immune responses. To stimulate a more vigorous immune response, researchers are developing viruses that will produce greater quantities or varieties of HIV proteins inside infected cells. Administering multiple doses of these vaccines may help generate and maintain high numbers of activated cytotoxic *T* cells.

Other researchers are looking into administering viral peptides—fragments of viral proteins—to induce an immune response. Because antigenic peptides derived from viral proteins activate cytotoxic *T* lymphocytes, perhaps peptides would work as a vaccine. Unfortunately, peptides by themselves do not elicit a strong immune response, cellular or antibody-based, in humans. The peptides may be degraded before they reach the target cells, or they may not be presented efficiently by the cells that encounter them. Peptide vaccines may benefit from the development of better adjuvants, materials delivered along with a vaccine that induce the immune system to respond more strongly.

A rather novel approach to eliciting a cellular immune response involves injecting "naked" HIV DNA—genetic material with no proteins or lipids to deliver or protect it. At one time, scientists believed that naked DNA would be degraded too rapidly to be effective as a vaccine. In reality, the DNA does get into cells and can direct the production of viral proteins. In animal studies in mice and nonhuman primates, DNA vaccines have successfully generated cytotoxic *T* lymphocytes that recognize HIV proteins. In some but not all experiments, the DNA vaccine protected animals from subsequent infection with HIV. Further studies in animals and humans are evaluating the safety and effectiveness of this approach.

Combination Strategies

The most effective strategies—and the ones that are furthest along in human testing—incorporate elements that will stimulate both arms of the immune response. For example, a patient might receive a vaccine containing a canarypox virus carrying the *Env* gene to stimulate cellular immunity. Months later the same patient might receive pure gp120 to elicit the generation of antibodies. This combination strategy is called a prime boost, because the canarypox vector primes the cytotoxic *T* cells, and the gp120 protein then strengthens, or boosts, the immune response by eliciting antibody production.

Early trials have demonstrated that humans vaccinated using such a combination strategy develop both humoral and cellular immunity. But the antibodies generated have been against laboratory-adapted HIV strains, and the cytotoxic *T* cell response has not been strong. The next generation of combination vaccines will use canarypox viruses that carry

more HIV genes capable of producing greater quantities of viral protein, and the boost may contain gp120 proteins made from HIV isolated from patients. Such vaccines are being produced and may soon be ready for testing in humans.

Many researchers also continue to look into developing a live, attenuated HIV vaccine. Because such a vaccine would closely mimic active HIV, it should theoretically be effective at inducing cellular immunity, antibody-based immunity and perhaps other unknown modes of protection. By systematically deleting genes critical for HIV replication, scientists hope to develop a variant of the virus that can elicit a strong immune response without giving rise to AIDS.

Recently a group of physicians volunteered to participate in the first clinical trial of a live, attenuated HIV vaccine. Such a protocol would allow researchers to monitor the volunteers' immune responses and study the long-term safety of the vaccine. The physician volunteers believe the value of testing this approach outweighs the potential risks to their health. Their plan remains highly controversial, and we and many other researchers feel that attenuated HIV viruses should be more fully investigated in nonhuman primates before any movement into human trials.

Monkeys and AIDS

Vaccines based on a live, attenuated simian immunodeficiency virus (SIV)—a relative of HIV that infects monkeys—have been tested in macaques and other nonhuman primates. Monkeys infected with pathogenic strains of SIV will develop an AIDS-like syndrome. By studying this monkey model, scientists are able to test live, attenuated vaccines for their safety and their ability to protect animals when they are challenged by subsequent exposure to pathogenic strains of SIV. Several different attenuated SIV vaccines have proved remarkably effective at suppressing the growth of a wild-type virus.

The basis of this immunity in macaques is unclear: animals that are effectively protected from SIV challenge do not necessarily have high levels of neutralizing antibodies or activated cytotoxic *T* lymphocytes. The protective effects may be a consequence of some combination of antibody, helper *T* cell and cytotoxic *T* cell activity, or the effects may derive from other aspects of immunity. Further work is needed to determine exactly how the SIV vaccines manage to confer protection.

Although initial studies suggested a high degree of safety for the live, attenuated SIV, extended and expanded safety studies are beginning to show increased numbers of vaccinated animals progressing to AIDS-like syndromes, even in the absence of exposure to wild-type virus.

FRED WHITEHEAD *Animals Animals*

RHESUS MACAQUE is a type of monkey being examined in vaccine studies. Animals receiving a live, attenuated simian version of HIV have been able to limit subsequent infection by the natural monkey virus. But in a worrisome finding, the vaccine itself has eventually caused disease in some monkeys.

The studies are now starting to look at a greater number of animals, but the results suggest that live, attenuated vaccines may not provide full, long-term immunity and may even cause disease. The findings also imply that investigators should proceed with caution before testing such vaccines in humans.

Prognosis

If the immune system in HIV-infected individuals cannot wipe out the virus, why should a vaccine that activates the same immune responses be expected to block infection? Vaccines may give the body an immunological "head start" by priming the immune system to attack HIV as soon as it appears, rather than taking time to initiate a defense from scratch. By doing so, vaccine-induced immunity may succeed in containing the virus where the naturally infected body does not.

At present, however, there is no proof that vaccination against HIV is possible, because no protective vaccine candidate has yet moved into Phase III trials, which are large-scale tests designed to evaluate effectiveness in humans. In addition, the wide genetic variability of HIV may reduce the utility of any vaccine under development, because HIV strains isolated from patients in different parts of the world have distinctly different structures in their Env and, to a lesser extent, other proteins. Whether these differences, or additional ones we have yet to appreciate, will significantly hamper vaccine development remains uncertain.

But there is hope. As the pathogenesis of HIV infection has become better understood, investigators have realized that if the virus can be kept at low concentrations in the blood, an infected person may never progress to AIDS. This insight is encouraging because it suggests that even a partially effective vaccine could be valuable in limiting the amount of virus in patients, thus potentially reducing their infectiousness and the symptoms they suffer.

It is unlikely that we will develop a vaccine suitable for wide-scale use in humans within the next five years. Even if the prime-boost combination approach appears to stimulate cellular immunity and generate good broad-spectrum antibodies, large clinical trials will still be needed to demonstrate its value. Those trials alone will take several years. In the meantime, researchers will continue to pursue every approach that might help the immune system combat HIV. SA

The Authors

DAVID BALTIMORE and CAROLE HEILMAN work together on the National Institutes of Health's AIDS Vaccine Research Committee, a group charged with reassessing the priorities of the vaccine initiative and identifying new and innovative areas of vaccine research. Baltimore, chairman of the committee since its formation in 1996, is president of the California Institute of Technology. Heilman, a microbiologist, is deputy director of the Division of AIDS at the National Institute of Allergy and Infectious Diseases in Bethesda, Md.

Avoiding Infection after HIV Exposure

by Susan Buchbinder

Treatment may reduce the chance of contracting HIV infection after a risky encounter

Suppose you are a doctor in an emergency room and a patient tells you she was raped two hours earlier. She is afraid she may have been exposed to HIV, the virus that causes AIDS, but has heard that there is a "morning-after pill" to prevent HIV infection.

Can you in fact do anything to block the virus from replicating and establishing infection? Would you respond differently to a patient with an HIV-positive sexual partner whose condom had broken during sex? How would you treat a patient who occasionally got drunk and had unprotected sex with someone of unknown HIV status?

Unfortunately, there are no easy answers. Only in the past few years have physicians had any data that could help resolve these dilemmas, and much remains unknown. Several studies suggest, however, that under certain circumstances, delivering anti-HIV medications after exposure to the virus—that is, giving postexposure prophylaxis (or PEP)—might prevent HIV infection.

JOHN CHIASSON *Gamma Liaison Network*

PHYSICIAN Mahlon Johnson, a pathologist, was infected with the AIDS virus in a workplace accident several years ago. His medicines are beside him. Today, treatment after an exposure to the virus can sometimes prevent infection.

Laboratory experiments indicate that administering anti-HIV drugs either before exposure or up to 24 hours afterward can protect some test animals from infection with HIV or a related virus that infects monkeys. The strength of the protection depends on the type and duration of the antiviral therapy.

Human studies also encourage hope that PEP can work. Transmission of HIV from women to their babies is reduced if the mothers are treated with the drug zidovudine (AZT) during pregnancy and labor and if the babies are treated immediately after birth. In this situation, however, protection may result in part from zidovudine circulating in the baby's blood at the time of exposure, rather than after exposure.

Further evidence comes from a study of health care workers exposed to a patient's HIV-infected blood through a needle stick or other accident that breaks the skin. The investigation found that health personnel who were exposed to HIV but did not become infected were more likely to have received PEP (with zidovudine) than were workers who did become infected. This finding, however, does not amount to proof that PEP reduces infection; methodological difficulties preclude drawing definitive conclusions.

Given the uncertain effectiveness of PEP, we physicians want to be sure that the treatment does not pose unacceptable risks to our patients. We know the short-term side effects from experience with HIV-infected patients, but we know almost nothing about the long-term consequences of using HIV-fighting drugs in people who may not in fact harbor the virus. Furthermore, we have only a very rough idea of how long PEP should be continued.

Despite the uncertainties, the Centers for Disease Control and Prevention have recommended that all medical personnel who have a significant exposure to HIV-infected blood or bodily fluids (such as through a needle stick) be counseled and, depending on the situation, be offered anti-HIV PEP. The drugs now recommended for this treatment are some of the same ones widely used to treat established HIV infection: zidovudine, lamivudine (3TC) and possibly also one of the newer drugs called protease in-

hibitors [*see table at right*]. These agents are known to be effective against HIV in combination, are widely available and have relatively few serious side effects. Alternative drugs may be given, depending on the likelihood that the source patient harbors resistant viruses and on other medical conditions the exposed worker may have.

Although the popular term "morning-after pill" might suggest an easy fix, the recommended prophylactic course involves taking the drugs at least twice a day for a total of four weeks. Patients who choose PEP must be ready for a number of temporary side effects, including headache, nausea, fatigue and anemia. They must also be willing to face both unknown long-term risks and the fact that HIV often becomes resistant to the drugs it encounters. If a patient becomes infected in spite of PEP, any resulting viral drug resistance would reduce the options available for treating the infection.

Should the recommendation made for accidentally exposed health workers also be made for people who are exposed to HIV through sex or through using contaminated needles to inject drugs? The unsatisfactory answer is: perhaps.

Some types of nonoccupational exposure can be just as dangerous as a needle-stick injury. The risk of becoming infected from a single instance of unprotected receptive anal sex with a man of unknown HIV status, for example, may be as high as that from a stick by a needle that has been used in a patient known to carry HIV. The one-time risk from using contaminated injection equipment appears to be comparable.

Yet the circumstances surrounding nonoccupational exposures differ in important ways from those that occur in health care settings. These circumstances may lessen the effectiveness of PEP or even increase the dangers.

Beyond Health Care Workers

Health care workers sustaining a needle-stick injury usually recognize their exposure immediately and can obtain PEP medications without having to pay for them. Moreover, their physicians will have access to the source patient's medical record, which means they can tailor treatment so it is most likely to be effective. For example, if a drug had previously been found to be ineffective at reducing virus levels in the source patient, then the exposed person would probably need to avoid taking it.

In contrast, individuals exposed through sexual contact or drug use may not know the HIV status of a partner, and the partner's medical record is often unavailable. Even when people immediately recognize a nonoccupational exposure to HIV, they may find it hard to arrange quickly for medical care and for payment to cover expensive drugs.

This barrier is a major problem, because the available evidence suggests that PEP must begin quickly to be effective. Indeed, treatment that starts within hours of exposure is likely to be more effective than treatment that starts a day or more later, and there is probably little to be gained from initiating PEP more than 72 hours after exposure.

Even if a nonoccupational exposure to HIV is unequivocal and the patient seeks help quickly, other factors may argue against treatment. Notably, if the patient cannot take the prescribed medications as directed or is exposed repeatedly to HIV during the period of PEP treatment, medication-resistant HIV may emerge (thus complicating future treatment). Yet many people find it difficult to adhere strictly to therapy, and some find it hard to practice safe sexual or intravenous drug-taking behavior consistently.

Other risks of PEP are less obvious. Merely knowing that it is available could lead some people to take more chances than they would otherwise. Thus, there is a real possibility that offering PEP for a nonoccupational exposure might in some cases increase the likelihood of infection. This consideration presumably does not apply to health care workers, who seem unlikely to become more careless just because PEP is available.

Several studies are in progress or being planned to clarify the benefits and drawbacks of PEP for people exposed nonoccupationally, so that health care providers can make sound recommendations. In the meantime, physicians have published some general guidelines.

Doctors confronted with a worried patient should assess the probability that HIV infection will occur, considering the type of exposure and the likelihood of infection in the partner. Counseling should be offered and referrals made as appropriate. Because HIV transmission during sexual assault has been reported, PEP is often recommended in such cases. Other isolated high-risk sexual exposures (such as condom breakages) may justify PEP, particularly for the receptive partner (for whom the risk of acquiring HIV is highest). Because multiple courses of therapy also multiply the risks of treatment, most physicians would not recommend PEP for persons who are exposed to HIV repeatedly.

PEP is at best a strategy of last resort in preventing HIV infection. The real hope for halting the worldwide spread of this deadly virus is still a comprehensive preexposure prevention program and—ultimately—an effective vaccine. SA

Typical Protocols for Postexposure Prevention

Drug	Type
Preferred two-drug regimen:	
Zidovudine (Retrovir, AZT)	Reverse transcriptase inhibitor
and	
Lamivudine (Epivir, 3TC)	Reverse transcriptase inhibitor
Alternative two-drug regimen:	
Stavudine (Zerit, d4T)	Reverse transcriptase inhibitor
and	
Didanosine (Videx, ddI)	Reverse transcriptase inhibitor
If source patient has advanced HIV disease, has a high viral load or has previously been treated with any drug in the selected two-drug regimen, consider adding:	
Nelfinavir (Viracept)	Protease inhibitor
or	
Indinavir (Crixivan)	Protease inhibitor

SLIM FILMS

DRUG REGIMENS may, if instituted quickly, reduce the chance of HIV infection after a risky exposure. Therapy is usually maintained for four weeks.

The Author

SUSAN BUCHBINDER is director of the HIV research section at the San Francisco Department of Public Health and an assistant clinical professor at the University of California, San Francisco. Buchbinder studies patients who carry the virus yet remain healthy over many years as well as people who resist HIV infection despite multiple exposures. She is involved in research on preventive HIV vaccines and on other approaches to reducing infections.

Coping with HIV's Ethical Dilemmas

by Tim Beardsley, *staff writer*

The issues are many and thorny

AIDS has challenged ethical precepts from the earliest days of the pandemic, and it continues to pose agonizing dilemmas in countries rich and poor. AIDS patients and others infected with HIV, the AIDS virus, face discrimination of various sorts the world over, and the high cost and long-term nature of treatment aggravate quandaries that occur to a lesser degree in other serious diseases. Major battles revolve, in particular, around ensuring access to therapies and testing new ones in an ethical manner.

Fear of contagion has sometimes led to discrimination against HIV-infected people not only by the public but also by medical personnel, and some related legal struggles from the early days of the epidemic continue. This summer the U.S. Supreme Court should rule on an appeal from a dentist who refused to treat an HIV-infected woman in his office. She successfully sued, alleging discrimination under the Americans with Disabilities Act. The case could change the legal interpretation of disability.

Yet discrimination is only one of the ethical issues that HIV has raised for medical professionals and society. The best available therapy, which usually combines three drugs that suppress viral replication, must be continued indefinitely and costs more than $10,000 a year. Although states have instituted assistance programs to provide anti-HIV drugs to those who would otherwise be unable to get them, the funds are limited, and so the poor have less access than those with good insurance policies.

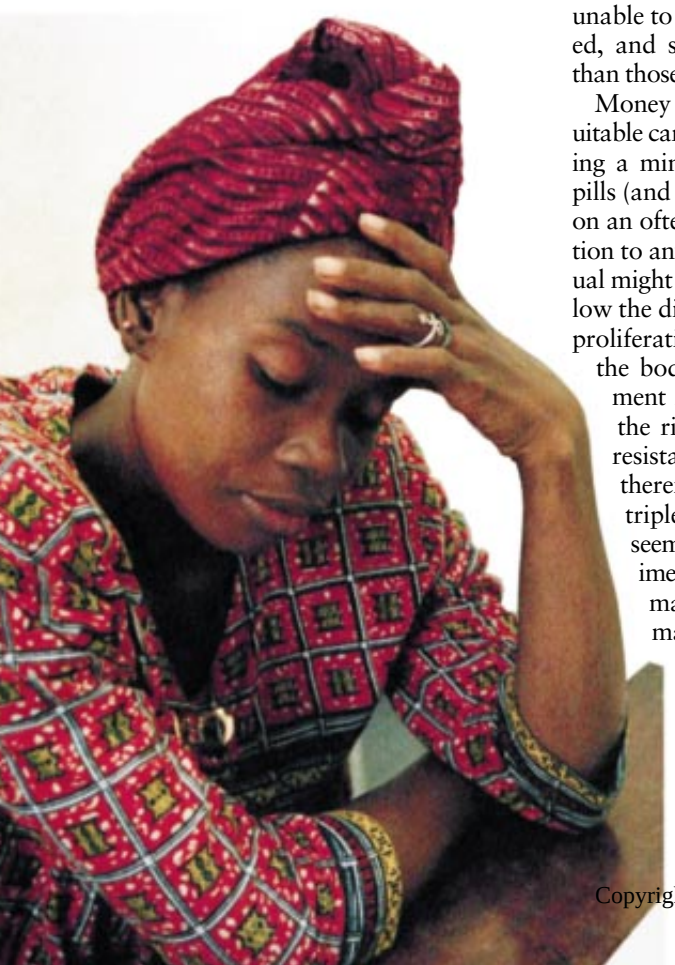
Money is not the only obstacle to equitable care. Triple therapy involves taking a minimum of eight HIV-fighting pills (and frequently many more) a day on an often complex schedule, in addition to any other medicines the individual might need. Patients who fail to follow the directions risk encouraging the proliferation of drug-resistant virus in the body, making subsequent treatment more difficult and increasing the risk that the patient will pass resistant virus to others. Physicians therefore sometimes advise against triple therapy for patients who seem unlikely to adhere to the regimen dutifully. "Compromises are made," says Douglas D. Richman of the University of California at San Diego School of Medicine, an expert on drug resistance in HIV.

Advances in therapy in the past two years have also raised ethical issues that are now affecting the conduct of clinical trials. Physicians have learned that no anti-HIV drug works for long by itself and that solo use of a drug promotes resistance to it. Therefore, biomedical ethicists agree it is no longer appropriate to conduct long-term trials of any single drug in HIV-infected patients in the U.S. The need to maintain high ethical standards has thus greatly increased the complexity—and cost—of new trials.

Ethical agreement has been impossible to find in a bitter argument over the propriety of rich countries sponsoring in poor countries studies that would be unethical at home. The controversy erupted in 1997, when Peter Lurie and Sidney M. Wolfe of Public Citizen's Health Research Group insisted that certain research being sponsored in developing countries by the Centers for Disease Control and Prevention and the National Institutes of Health, as well as the Joint United Nations Program on HIV/AIDS, was unethical and exploitative. The studies were aimed at learning whether a short course of the drug zidovudine (AZT) or various other inexpensive therapies could prevent the transmission of HIV from infected women to their babies in several countries in Africa and Asia.

Lurie and Wolfe objected to the studies because scientists knew even before they started that zidovudine reduces transmission of HIV by about two thirds if the drug is administered in a regimen known as ACTG 076 (so called because a trial with this designation proved the regimen's value). In this protocol the drug is administered during the last 26 weeks of pregnancy, during delivery (by intravenous drip) and to the baby for six weeks after birth. But the regimen, which costs about $800, is too expensive for wide use in developing countries.

The disputed studies sought to determine whether a much less expensive course of zidovudine or other medication, given only during the last three to four weeks of pregnancy and costing no more than a tenth as much as the ACTG 076 protocol, could still achieve a worthwhile reduction in HIV transmission. Lurie and Wolfe objected to the decision to place some mothers in a control group in which they received only fake medicine that was of no value to their babies. They argued that instead of



DAVID GUTTENFELDER *Associated Press*

PREGNANT WOMAN in Ivory Coast who carries the AIDS virus considers enrolling in a study of treatment that might prevent transmission of the virus to her baby.

comparing a short course of medication with a placebo, the investigators should have compared the short course with the full ACTG 076 protocol to see how closely the two were equivalent.

Defenders of the trials, notably Harold E. Varmus, director of the NIH, and U.S. Surgeon General David Satcher, argued that by including placebos investigators could learn more—and more quickly—about the value of inexpensive treatments in reducing transmission. The placebos were necessary to learn the baseline transmission rates without treatment, because these rates vary in different parts of the world, for reasons that are largely obscure.

Moreover, the defenders asserted, placebos would help identify any important toxicity of the drugs studied—not an insignificant concern, as anemia, one of the known side effects of zidovudine, is much more common in undernourished populations than in the U.S. The trials did no harm, Varmus and others pointed out, because the placebo—no treatment—was the standard of care in the countries under study. "Research does not create an entitlement," concurs Robert J. Levine of Yale University.

Lurie and Wolfe and their supporters were not satisfied. No foreseeable toxicity from zidovudine or other therapy could counterbalance the babies' lives that might have been saved by using the ACTG 076 protocol in the comparisons, they maintained. Other critics believe certain participants in the trials may not have given true informed consent.

Wendy K. Mariner of Boston University holds that trials such as the disputed ones can be ethical only if there is a plan at the outset to make the treatment available to the local population if it proves to be effective. In fact, there was no plan to provide it at the time the studies got under way, Mariner says. Yohana J. S. Mashalla, vice president of the Medical Association of Tanzania, faults the studies as exploitative, for setting out to demonstrate in developing countries the value of a therapy that might then be adopted in developed ones.

Right or wrong, one of the studies sponsored by the CDC, which took place in Thailand, did show in February that a three- to four-week course of zidovudine reduced transmission significantly—though not as much as the ACTG 076 protocol. Over the following weeks, agencies started supplying the short course of the drug to women in studies around the world who had previously received only a placebo. The manufacturer of zidovudine, Glaxo Wellcome, then announced that it would cut prices of the drug for developing countries. But as of May, it was unclear how many women who needed the medicine would gain access to it.

The question of when to use placebos represents a "genuine conflict" between getting the fastest answers to critical questions and the interests of today's patients, says bioethicist Arthur L. Caplan of the University of Pennsylvania. But in his opinion, as a researcher "you owe something to your subjects—you owe them the best options."

One "underused tool" that could provide patients with better options is wider application of innovative statistical designs, according to Christopher R. Palmer of the University of Cambridge. Mathematically rigorous data-dependent designs for studies can, in some cases, provide firm information on the effectiveness of a therapy while exposing fewer patients to suboptimal treatment than a conventional trial would. During a data-dependent study, the health of the participants is continually analyzed. The accumulating evidence about the different treatments modifies the odds used to select which therapy the next participant will receive.

Data-dependent designs are as yet little used, partly because many medical statisticians are unfamiliar with the specialized math involved. But they are often ethically attractive for late-stage studies on serious diseases, Palmer notes.

The dilemmas of research on HIV will become especially acute when candidate HIV vaccines become available for large-scale trials. Because triple therapy is the standard of care in the industrial nations, some, including Lawrence O. Gostin, law editor of the *Journal of the American Medical Association,* say researchers have an obligation to provide it to all participants in vaccine trials who become infected. Vaccine manufacturers, though, may well blanch at the cost of supplying triple therapy even temporarily for thousands of enrollees.

HIV will continue to tax society's conscience until inexpensive, highly effective therapies and vaccines are available for all. And that day can come only through more research. Meanwhile the desperate souls who sign up for a trial of a new drug or vaccine must do so knowing they are taking a risk—but perhaps a reasonable one that may benefit humanity's collective future. SA

Further Readings on HIV and AIDS

The Global Picture

Report on the Global HIV/AIDS Epidemic. UNAIDS and WHO, December 1997. Available at http://www.unaids.org/unaids/graphics/1997/report97/index.html on the World Wide Web.

Prospects for Therapy

Guidelines for treating adults and adolescents, for treating children, and for preventing mother-child transmission are available from the HIV/AIDS Treatment Information Service (800-458-5231) and through http://www.hivatis.org/trtgdlns.html or http://www.cdcnac.org on the World Wide Web.

Plasma Viral Load and CD4+ Lymphocytes as Prognostic Markers of HIV-1 Infection. J. W. Mellors, et al. in *Annals of Internal Medicine,* Vol. 126, No. 12, pages 946–954; 1997.

Antiretroviral Drug Resistance. Edited by Douglas D. Richman. John Wiley & Sons, 1996.

Children with HIV

Antiretroviral Therapy: Evaluating the New Era in HIV Treatment. R. E. McKinney in *Advances in Pediatric Infectious Diseases,* Vol. 12, pages 297–323; 1996.

Postexposure Protection

Postexposure Treatment of People Exposed to the Human Immunodeficiency Virus through Sexual Contact or Injection-Drug Use. Mitchell H. Katz and Julie Louise Gerberding in *New England Journal of Medicine,* Vol. 336, No. 15, pages 1097–1100; April 10, 1997.

Vaccine Strategies

HIV Vaccines ...Where Are We Going? C. A. Heilman and D. Baltimore in *Nature Medicine,* Vol. 4, No. 5, pages 532–534; May 1998.

Changing Behavior

Prevention of HIV Infection: Looking Back, Looking Ahead. J. Stryker, T. J. Coates, P. DeCarlo, K. Haynes-Sanstad, M. Shriver and H. J. Makadon in *Journal of the American Medical Association,* Vol. 273, No. 14, pages 1143–1148; April 12, 1995.

Information on many HIV prevention–related topics is available from the Center for AIDS Prevention Studies at http://www.caps.ucsf.edu on the World Wide Web.

Ethical Challenges

Ethical Issues in Studies in Thailand of the Vertical Transmission of HIV. P. Phanuphak in *New England Journal of Medicine,* Vol. 338, No. 12, pages 834–835; March 19, 1998. Correspondence, ibid., pages 836–844.

THE AMATEUR SCIENTIST

by Shawn Carlson

A Year for the Oceans

Scuba diving is unlike any other experience in the world. I earned my certification when I was 14 and shortly thereafter went on my first diving adventure off the coast of Santa Catalina Island, about 65 kilometers (40 miles) south of Los Angeles. My first dive that day provided ample evidence that the ocean is a marvelously strange place. But even more impressive was diving in the Pacific later that night, an experience that blended awe for the wonders of nature with psychological stress from being engulfed in darkness. After all, in the ocean, we humans are not the top predator.

The first time I rolled off a boat at night and sank into the dark sea, I had to fight back panic. The beam of my flashlight penetrated only a few meters ahead in the water, producing an eerie and disturbing awareness of the darkness around me. But soon, being totally enveloped in the cool, velvety blanket of seawater became oddly comforting, even inviting to me. So I broke the first rule of scuba safety and separated myself from my fellow divers by about 30 meters. Then, summoning all my courage, I turned off my flashlight.

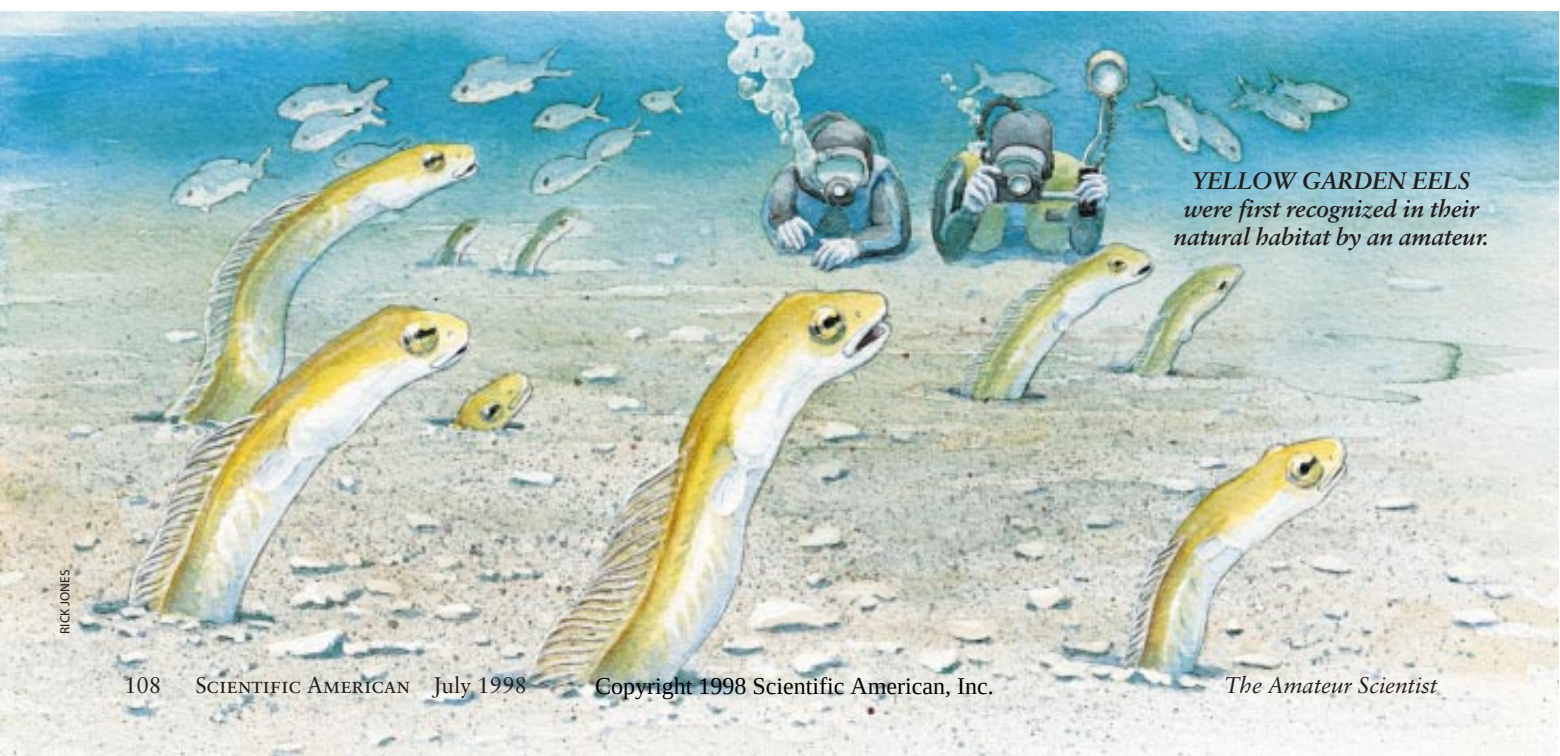
Suddenly, I felt as though I had been cast to the dark side of the moon. Fighting to steady my breath, I could sense only the sounds of my regulator and the gentle undulating surge of the current as waves crested above me. Once I calmed down, my body seemed to melt into the inky water. A few minutes later my eyes adapted, and I realized that the blackness was not absolute after all. No longer blinded by the glare of my flashlight, I could see that the waves high overhead were faintly afire with the glow of the moon. Then something large swam by, stimulating a colony of bioluminescent bacteria and kicking up a dim trail of light in its wake. Although I had no idea what kind of creature it was, I was oddly confident that it wasn't interested in eating me. Instead it ensnared me with a luminous lasso as it circled once and swam off.

Eventually I rejoined my diving companions and later learned that they had never missed me. They had been too busy shining their flashlights on the bottom and coaxing curious lobsters from their dens to take much notice of my temporary absence. I probably shouldn't have risked drifting so far from them, even for a short while. Still, the magic of those few minutes has stuck with me for nearly a quarter of a century.

Every diver has had similarly moving experiences. In fact, I am sure that many of the people who read this column share my passion for both science and the underwater world. That is why, in this International Year of the Ocean, I'm delighted to report that marine biologists all over the world have finally recognized that sport divers can be powerful allies. Exciting new investigations are coming to life that could not be carried out without the help of amateurs.

Reef Check is one of the most ambitious of this new brand of endeavor. In a single year this program has become one of the largest international collaborations of marine researchers ever organized. Run from the Hong Kong University of Science and Technology, Reef Check coordinates 100 teams of volunteers who last year surveyed more than 300 coral reefs in the territorial waters of 31 countries scattered throughout the world. These teams are led by professional marine biologists who teach volunteer divers the techniques they need to make a thorough assessment of the health of coral reefs—hot spots of biodiversity that are second only to tropical rain forests.

Before Reef Check began operations last year, no one had mounted system-



YELLOW GARDEN EELS were first recognized in their natural habitat by an amateur.

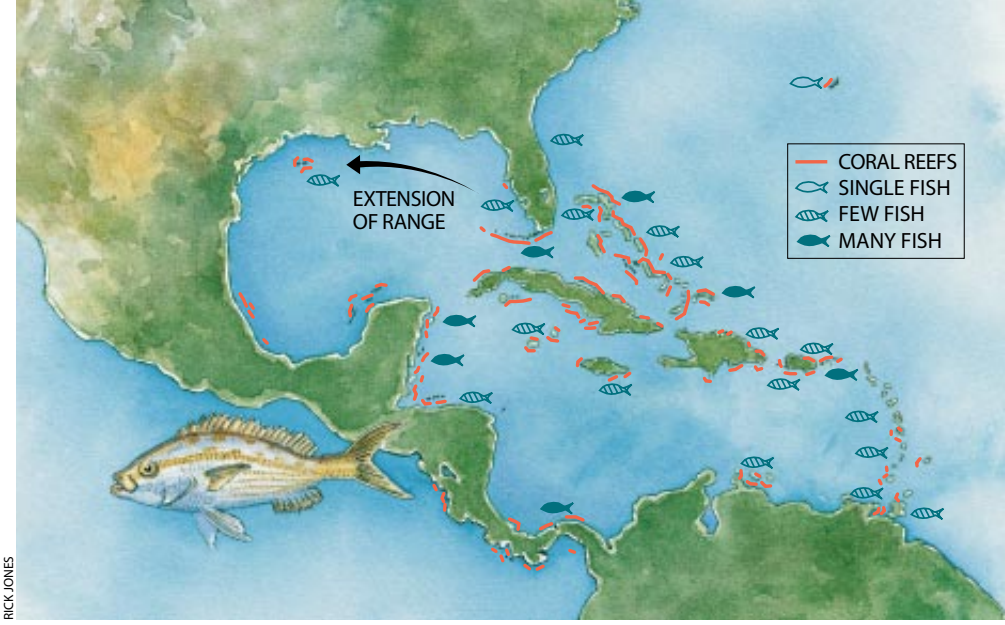
RICK JONES

atic surveys of these vital ecosystems on such a broad scale. So the news was disturbing when Reef Check's first overview turned up plenty of evidence of damage being done to these fragile structures. Continuing work should provide detailed pictures of how these habitats are changing and whether efforts to protect them are working. To get involved, consult Reef Check's site at www.ust.hk/~webrc/ReefCheck/reef.html on the World Wide Web.

A second organization that now routinely enlists amateurs is the Reef Environmental Education Foundation (REEF). REEF, which has more than 10,000 members, is headquartered in Key Largo, Fla. (telephone: 305-451-0312, or www.reef.org on the Web). But don't confuse REEF with Reef Check. Reef Check focuses on coral-based ecosystems, whereas REEF surveys fish populations in various places, whether coral is present or not.

REEF assembles teams of professional and amateur scientists and sends them to do biological surveys all over the Caribbean, Florida, the Gulf of Mexico and the California coast. Beginning this summer, REEF will include groups working near the coasts of Oregon, Washington State and British Columbia. Shortly thereafter REEF will extend its reach to include North Carolina as well. Membership is free, but participation in the surveys could set you back a bit. Scuba divers and snorkelers pay fees ranging between $600 and $1,700 to spend their vacations working with world-class marine biologists.

To learn how to identify the species they will most likely encounter, volunteers attend an informal "fish school" in tandem with their scientific missions. REEF also provides these participants with special fish identification cards to carry on their dives. All the information they record goes into an ever-growing database of fish sightings. But you don't have to join one of these organized expeditions to contribute. A small army of divers now take REEF's species identification cards with them whenever they poke about the bottom. (Contact REEF to obtain a fish ID card for a modest fee.) Largely as a result of their efforts, REEF's database—which now exceeds 13,000 entries—is one of the most comprehensive sets of observations about fish populations ever assembled.



RICK JONES

COMPILATION OF AMATEUR SIGHTINGS
from REEF revealed that the yellowtail snapper expanded its range to the corals of the northern Gulf of Mexico in 1997. This fish, which normally lives around coral reefs, may have used the numerous oil platforms in the area as stepping-stones.

REEF volunteers also help to protect fragile marine habitats in other ways. For instance, on a recent trip to Dry Tortugas, off the coast of southern Florida, REEF volunteers found an area that held a surprisingly rich array of coral down to around 40 meters (131 feet) below the surface: large star, pillar, plate and brain coral form a living carpet that covers half the seafloor there. This expanse of nooks and recesses provides an extraordinary habitat for wildlife. Its discovery by volunteers prompted the National Marine Fisheries Service to ban ships from dropping anchor in the area (which would damage the bottom) and to mount their own surveys of this previously unknown patch of deep coral.

Several REEF-trained volunteers have gone on to make other important discoveries. For instance, the Yellow Garden Eel was once known to biologists only because tattered specimens had been dredged up from the Gulf of Mexico decades ago. Then, in 1996, REEF member Ken Marks spotted the yellowish eels in a popular dive spot off Florida's Deerfield Beach. When his photographs of the curious creatures reached the Smithsonian Institution, experts there confirmed Marks's suspicion that he had recorded the first image of this elusive eel in the wild. Amateurs such as Marks have had little difficulty convincing professional marine biologists of their value.

Even the federal government has finally wised up to the potential of amateur scientists. Since 1994 the National Oceanic and Atmospheric Administration (NOAA) has promoted the Great American Fish Count (GAFC), a yearly survey of marine fish. During the first two weeks of July, snorkelers and divers record a snapshot of ocean life. Just as with the Audubon Society's annual Christmas bird count [see "The Joys of Armchair Ornithology," The Amateur Scientist, April 1997], biologists can now study how wild populations are changing from year to year.

NOAA concentrates its volunteers in the national marine sanctuaries located near the Channel Islands off the southern California coast; Monterey Bay off the coast of Monterey, Calif.; the Flower Garden National Marine Sanctuary southeast of Galveston, Tex.; and the Florida Keys National Marine Sanctuary. To get involved, check out the GAFC Web site at www.fishcount.org or call Christy Pattengill at 800-862-3260. Your efforts may help keep the ocean the dazzling natural wonderland I found it to be when I first learned to dive. SA

For more information about this and other projects for amateurs, visit the Forum section of the Society for Amateur Scientists at http://web2.thesphere.com/SAS/WebX.cgi on the World Wide Web. You may also write the society at 4735 Clairemont Square, Suite 179, San Diego, CA 92117, or call 619-239-8807.

by Ian Stewart

The Bellows Conjecture

Every amateur carpenter who has tried to build a bookcase knows that rectangles are not rigid. If you lean against the corner of a rectangular bookcase that has not been constructed well, it will tilt sideways to form a parallelogram—and in all likelihood collapse. A triangle, on the other hand, is rigid: you cannot change its shape without changing the length of at least one side. In fact, the triangle is the only rigid polygon. A bookcase built in any other polygonal shape—rectangle, pentagon, hexagon—must be braced in some manner. For example, cross-struts can be added, which has the effect of dividing the polygon into a pattern of rigid triangles.

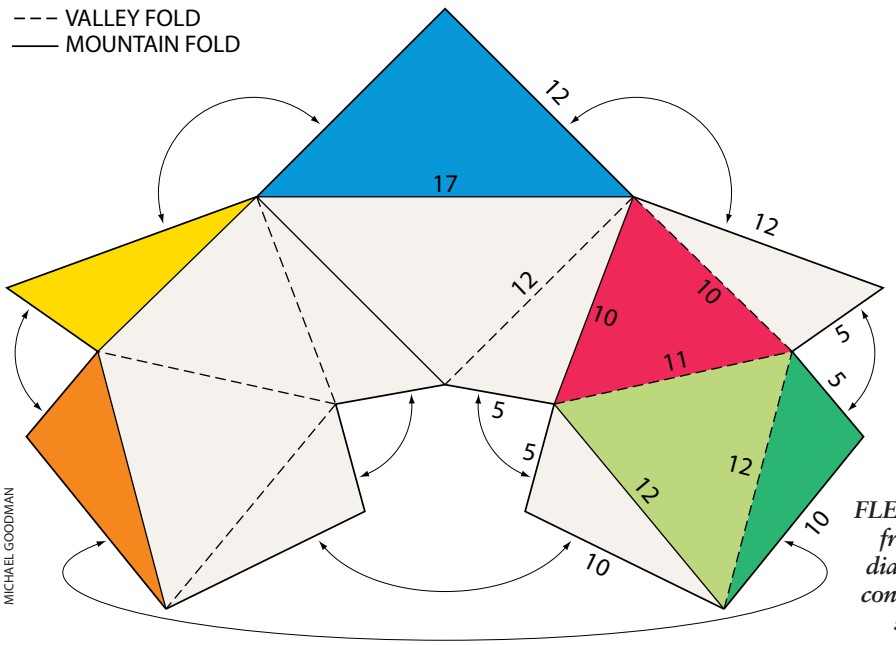
Another way to shore up a bookcase is to nail a flat back onto it. This takes the question into the third dimension, where everything becomes far more interesting. For nearly 200 years, mathematicians have been puzzled by the rigidity of polyhedra—solids with a finite number of polygonal faces that meet along straight-line edges. Until recently, it was assumed that any polyhedron with triangular faces must be rigid. But that assumption turned out to be false. There exist "flexible" polyhedra that can change shape without distorting any of their triangular faces. And just last year a trio of mathematicians proved the long-postulated Bellows Conjecture, which states that the volume of a flexible polyhedron stays constant as it changes shape. The proof, which was based on an ancient Greek formula, has opened a whole new realm of mathematical research.

Anyone who has folded origami figures knows that it is possible to make many flexible shapes: birds that flap their wings, frogs whose legs move, and so on. Are these figures examples of flexible polyhedra? The answer is no. When the paper frog moves its legs or the paper bird flaps its wings, the paper bends slightly. The same goes for an accordion: the instrument's bellows is able to expand and contract because the facets of the bellows bend and stretch. But the faces of a flexible polyhedron do not bend at all, not even by a billionth of an inch. When a polyhedron flexes, the only things that change are the angles at which the faces meet. Imagine that the faces are hinged along their edges. Everything else is perfectly rigid.

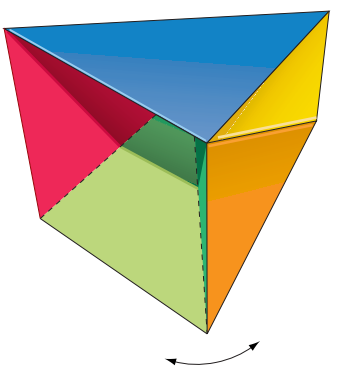
This field of study dates back to 1813, when the French mathematician Augustin Louis Cauchy proved that a convex polyhedron—one without indentations or recesses—cannot flex. But what if there are indentations? Raoul Bricard, a French engineer, discovered that nonconvex polyhedra can indeed flex if the faces of the polyhedron are allowed to pass through one another. Of course, this is impossible for a real object in the physical world. Bricard's figures can be visualized, however, if we remove the polyhedron's faces and replace the edges with a framework of rigid rods.

In the 1970s Robert Connelly, now the head of Cornell University's mathematics department, modified Bricard's nonconvex polyhedron in such a manner that it remained flexible, but its faces no longer passed through one another. The construction was simplified by Klaus Steffen of the University of Düsseldorf, who discovered a flexible polyhedron with nine vertices and 14 triangular faces [*see illustration below*]. It is amusing to make a model of this polyhedron out of thin cardboard and see how it flexes. As far as anyone knows, this is the simplest possible flexible polyhedron, although it would be very difficult to prove such a statement.

Mathematicians who investigated these polyhedra quickly noticed that as they flexed, some faces moved closer together while other faces moved farther apart. It looked as if the total volume of



MICHAEL GOODMAN



FLEXIBLE POLYHEDRON can be constructed from thin cardboard using a blowup of the diagram (left). *Make the indicated folds, then connect the edges. When completed, the model should be able to flex as shown* (above).

the polyhedra did not change during their motion. Dennis Sullivan of the City University of New York tested this hypothesis by making a small hole in a flexible polyhedron and filling the model with smoke. When he flexed the smoke-filled polyhedron, no smoke puffed out of the hole. This crude experiment suggested—but of course did not prove—that the volume of the model remained constant. The proposition became known as the Bellows Conjecture because it implied that a flexible polyhedron does *not* act like a bellows, which changes volume—blowing out and sucking in air—as it changes shape.

The first interesting feature of the Bellows Conjecture is that its two-dimensional analogue is false. When a flexible polygon, such as a rectangle, collapses into a parallelogram, its area gets smaller. Clearly, there is something unusual about three-dimensional space that makes a polyhedral bellows impossible. But what?

To solve the puzzle, Connelly and two other mathematicians—Idzhad Sabitov of Moscow State University and Anke Walz of Cornell—focused on an ancient formula for the area of a triangle. Although scholars believe the formula was derived by Archimedes, it is usually credited to Heron of Alexandria, who wrote down the proof. Heron was a Greek mathematician who lived sometime between 100 B.C. and A.D. 100. His formula, shown in the box above, relates the area (x) of a triangle to the lengths of its three sides (a, b, c). Notice that the equation is polynomial: its terms are whole-number powers of x, a, b and c.

Sabitov came up with the novel idea that there might be a similar polynomial equation for any polyhedron, relating the volume of the solid to the lengths of its edges. Such an equation would be a remarkable discovery. There are some well-known formulas for special polyhedra—such as cubes and rectangular boxes, whose volume equals the product of length, width and height—and something a bit like Heron's formula for tetrahedra, solids with four triangular faces. But nobody had ever derived a general formula that would give the volume of any polyhedron. Could the brilliant mathematicians of the past really have missed such a wonderful idea? It seemed unlikely.

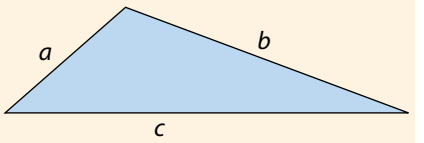
Nevertheless, suppose such a formula exists. Then the Bellows Conjecture must be true, because the volume of a polyhedron would depend solely on the lengths of its edges. As the polyhedron flexes, the lengths of its edges do not change—so the volume of the polyhedron must not change either. There is one problem with this line of reasoning: a polynomial equation can have several distinct solutions, so in theory the volume of the polyhedron might jump from one solution to another. But such a leap cannot happen in the physical world—if the flexing is done gradually, the volume cannot suddenly change to a new value. So it must remain constant.

The mathematicians began their proof by modifying the formula for the volume of a tetrahedron, which is similar to Heron's formula but more complex. Just as any polygon can be divided into triangles, any polyhedron can be divided into tetrahedra. The volume of the polyhedron is the sum of the volumes of its tetrahedral components. This method alone will not solve the problem: it leads to a formula that involves the edges of all the tetrahedral pieces, and many of these lines are not edges of the original polyhedron. Instead they are diagonal lines that run from one corner of the polyhedron to another, and their lengths may change when the polyhedron flexes. The mathematicians had to "massage" the formula algebraically to get rid of the variable components.

It was a messy process. For an octahedron—a solid with eight triangular faces—the massaging procedure resulted in a polynomial equation with the volume raised to the 16th power. More complex polyhedra required higher powers still. By 1996, however, Sabitov had devised an algorithm for finding a volume formula for any polyhedron. In 1997 the team of Connelly, Sabitov and Walz greatly simplified the algorithm.

The reason why such equations exist for all polyhedra is not fully understood. In two dimensions, there is no equivalent besides Heron's formula, which applies only to triangles. Connelly and Walz think they know how to prove a four-dimensional Bellows Conjecture, but for five or more dimensions, the problem is wide open. It is fascinating to see how a simple experiment with some folding cardboard led to a marvelous and totally unexpected mathematical discovery. SA

Heron's Formula



Suppose that a triangle has sides a, b, c and area x. Let s be the semiperimeter.

Then $$s = (a+b+c)/2$$
$$x = \sqrt{s(s-a)(s-b)(s-c)}$$

If we square this equation, the result is
$$16x^2 + a^4 + b^4 + c^4 - 2a^2b^2 - 2a^2c^2 - 2b^2c^2 = 0$$

JOHNNY JOHNSON

FEEDBACK

In response to "The Lore and Lure of Dice" [November 1997], many readers suggested variations on the set of "nontransitive dice" mentioned in the column. Those dice had faces as follows: A (3, 3, 4, 4, 8, 8); B (1, 1, 5, 5, 9, 9); and C (2, 2, 6, 6, 7, 7). In the long run, B beats A with probability $^5/_9$, C beats B with probability $^5/_9$ and A beats C with probability $^5/_9$. George Trepal of Sebring, Fla., pointed out that these numbers, suitably arranged, form the columns of a magic square (*below*), in which all the rows and columns add up to the same sum. Moreover, if the rows of this square are used for the die faces, the resulting set of dice—A (8, 8, 1, 1, 6, 6); B (3, 3, 5, 5, 7, 7); and C (4, 4, 9, 9, 2, 2)—is also nontransitive: A beats B with probability $^5/_9$, B beats C with probability $^5/_9$ and C beats A with probability $^5/_9$. Similar sets of dice made from larger magic squares seem not to have the nontransitive property. Can anyone prove or disprove this?

Trepal's best set of nontransitive dice—that is, the set using the lowest numbers—had faces as follows: A (1,1, 4, 4, 4, 4); B (3, 3, 3, 3, 3, 3); and C (2, 2, 2, 2, 5, 5). In this case, A beats B with probability $^6/_9$, B beats C with probability $^6/_9$ and C beats A with probability $^5/_9$. —*I.S.*

8	1	6
3	5	7
4	9	2

JOHNNY JOHNSON

THINKING ABOUT THINKING

Review by Jack Cohen

Brainchildren: Essays on Designing Minds

BY DANIEL C. DENNETT

MIT Press, Cambridge, Mass., 1998 ($40; paperbound, $20)

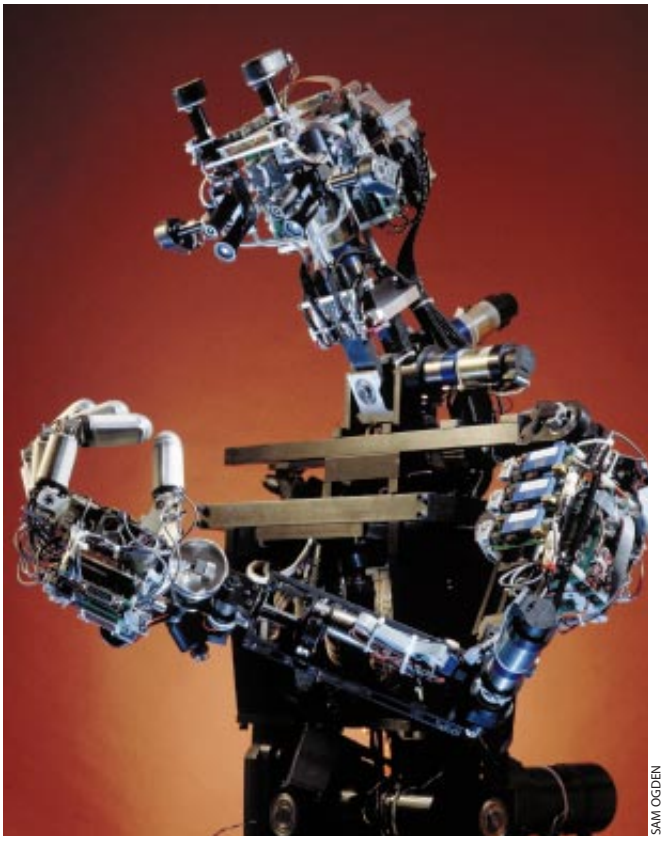
Dan Dennett is a big man. He takes up a lot of space in the physical, intellectual and academic worlds. I have never heard anyone wonder that he is the Distinguished Professor of Arts and Sciences at Tufts University; he lives up to it. He produces books like *Kinds of Minds* and *Consciousness Explained*. He has already had a special issue of *Philosophical Topics* devoted to his work and a collection called *Dennett and His Critics* from Oxford University Press. And there are many critics; he is in no majority that I can detect, yet everyone professionally concerned with minds takes notice of what he says. He is in fact a cult figure, which is strange because the song he sings is not a pop song but an extended aria. It requires attention, but it rewards the reader with that most valued commodity, interest: he changes people's minds—about minds.

I first met his thinking attached to that of Douglas Hofstadter in *The Mind's I*. Hofstadter I knew as the author of the wonderful *Gödel, Escher, Bach: An Eternal Golden Braid*; this chap Dennett I hadn't heard of (bear in mind that I am an English reproductive biologist). But I then bought his essays, *Brainstorms*. Essayists I enjoy. And here was an addition to my list, with the persuasiveness of Peter Medawar, the style of Lewis Thomas, the personal touch of Isaac Asimov.

Now he has produced this new collection, as diverse as ever but with two major themes: consciousness in people, machines and animals; and thinking about consciousness in people, machines and animals—metaconsciousness? (He quotes Hofstadter: "Anything you can do, I can do meta.") He claims the book is to make essays that originally appeared in, for example, the journal *Raritan* more accessible. One, from *Poetics Today*, is a nice piece on his experience of field ethology with vervet monkeys, which should be read by all who think that some animals, especially primates, must have minds like ours.

In addition to these difficult-to-find essays, Dennett included some of his better-known position statements. For example, "Real Patterns" (from the *Journal of Philosophy*) asks about the transactions and transformations that begin with regularities in visual fields, or sounds or odors, and that result in beliefs and opinions—changes in the structure of a mind. He believes, as most of us now do, that the "mind" is a diverse set of tools for perceiving and thinking. (In this regard, he devised the image of pandemonium, many demons—computer-speak demons—offering items that compete for attention.) Thus, the changes that occur are disparate, diverse, very rarely affecting all of the "me." Indeed, the essay "Real Consciousness" picks up on critics of his disbelief in such a "unitary" consciousness.

Too many philosophers, it seems to Dennett, still want "definitions"; they want to play with counters that have defined values, even in studies of the mind. When such philosophers play in science gardens like physics, this works. Definitions abound: an electron is a well-defined concept. Dennett shows in many of these writings, especially in "Do-It-Yourself Understanding" and "Self-Portrait," that biology isn't like that. "All mammals have mammal mothers," for example, but it isn't an infinite regress, as these philosophers might claim: there is a finite but *indefinite* history of mammals. Nearly all biological concepts have such blurred edges, nearly all biological processes are effective but inefficient, and we biologists work with them without difficulty and



SAM OGDEN

HUMANOID ROBOT, COG, is the long-term project of a team at the Massachusetts Institute of Technology. Dennett, the group's resident philosopher, considers it unlikely that anyone will ever make a robot that is conscious in the way humans are. But, he concludes, if Cog can eventually carry on conversations in something like a natural language, the robot will rival its own monitors—and the humans who interpret them—as a source for what it is doing and feeling, and why.

without persnickety definitions. Biology is like "discovered" engineering, and reverse-engineering thinking is needed to understand the whys and wherefores of animal and human behavior.

And too many philosophers believe in their own introspections. Dennett uses the "frame" problem, so difficult to access by introspection, as his demonstration of introspection's failure. When we deal with *anything*, we unerringly isolate the problem, the issue, what we are doing, from the irrelevant universe of possibilities that surrounds it. Our robots can't do this yet, perhaps because we can't see how we do it. Dennett's example is beautifully down-to-earth: the problem of getting a midnight snack from the fridge. All the givens, all the unknowns, all the assumptions—about the stickiness of the slices of turkey for the bread, say—are very difficult to locate by introspection; nevertheless, one "just goes and gets a sandwich."

"When we deal with anything, *we unerringly isolate the issue from the irrelevant universe of possibilities that surrounds it. Our robots can't do this yet."*

Sifting Wheat from Chaff

Only in another piece, about programming such detail into a robot, does this miraculous ability of our brains get exposed. Nobody knows how we do this. All animals do it. They don't get caught up, as stupid robots do, listing all the unproblematic things ("That tree probably won't move, that rock probably won't move") and what might happen to the problematic ones ("That bird might fall on my head"). Dennett looks at R1-D1, R2-D1 and their kin—clever R2-D2, of course, has solved this.

Like R2-D2, animals just get on with their business and exclude the irrelevant. A simple basis for this could be that nerve cells habituate, so that even uncomplicated creatures don't "see" trees and rocks, and most animals have tuned-down senses, so that what excites them is specially tuned-*for* and therefore probably important. Dennett calls this decoration of the important sense impressions "salience." I think he could usefully relate this emphasis of significant items to his arguments about qualia. Qualia, such as "redness" and "the smell of bacon," are decorations of the perceptions in the mind, which were transduced from light-frequency and molecules in the air. Dennett takes on qualia in two essays. One, about Hard Questions of consciousness, invokes zombies, and it is a failure for me. Zombies are imaginary people who are identical down to molecules with the rest of us but who don't have qualia; they react to red light, but they don't see "red," as we do, inside their heads. This imaginary exercise somehow shows some philosophers, but not Dennett or me, that consciousness (as qualia) is "not materially rooted."

Dennett gets very highbrow about this nonsense and invents zimboes, a kind of improved-model zombie, and makes a lot of comparisons with other people's models to show how silly the concept is. I didn't, and don't, understand how zimboes help disprove anything. Zombies can exist only in dualist universes; imagining them begs the very question they are set up to test. (In *Figments of Reality* Ian Stewart and I used the imagined "zombike" to debunk zombies: this is exactly like a regular bicycle, down to molecules, but the back wheel doesn't go round when you turn the pedals. A comparable argument would show, just as stupidly, that locomotion, like consciousness, is "not materially rooted," an ineffable hyperproperty of the system.)

Another failed essay, for me, is the final one in the book, "Information, Technology and the Virtues of Ignorance." In it, Dennett contends that our moral obligations to act must become overwhelming with the increased information we all have. So, according to him, these obligations must be limited again—as they were in the past—by increasing our relative ignorance. This kind of thinking seems to me to be another example of the "frame" problem: we will construct little "relevant," framed moralities for ourselves by failure to consider most of the world's problems, even if we do know about them.

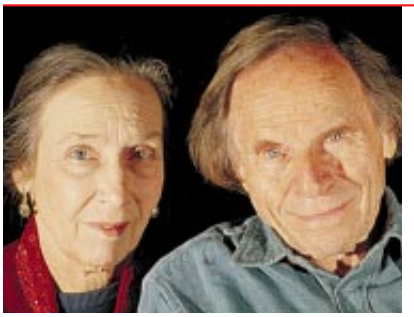
There are some bonuses—including a wonderfully critical but just-persuaded account of multiple-personality psychoses, written with Nicholas Humphrey; a very positive response to the critics of Julian Jaynes's *The Origins of Consciousness in the Breakdown of the Bicameral Mind*; comments on artificial life (AL) as well as artificial intelligence (AI)—all good-tempered, often witty.

Several of the essays tackle (and tackle very well) the issues of animal minds and our ability to think about them in terms of our own mind. Quotes from *Moby Dick* about whale vision being left-right separate softened me up for Thomas Nagel's question about what it's like to be a bat. Then I learn that rabbits' brains don't transfer learning across the midline—train them with the right eye, the left shows no response! Next I'm challenged by experiments with snakes, which use, apparently must use, three different sensory modalities to catch and eat prey: they hunt and strike only visually; they find the prey only by smell (even if it is squeezed in their own coils); and they find the head, to swallow the prey, by touch.

I am not naive in these animal-behavior matters. I've kept many kinds of mammals: llamas, Chinese hamsters, polecats, bush babies and other creatures, like mantis shrimps and octopuses, that show considerable intelligence. And many snakes, big spiders, odd insects and fish, including the intelligent cichlids. Yet Dennett tells me a story that convinces me, about the "minds" of these creatures, which is different from what I believed 10 years ago. He takes my attention to places—experiments with rabbit-brain sidedness, for example—that decorate this new view with believability. He shows me how *different* minds are, even vervet monkey minds. Again he has got in among my prejudices and changed my mind; this new book has opened new areas of contention for me.

This man improves my universe and yours. Enjoy him. SA

JACK COHEN, a reproductive biologist, has positions in both the Ecosystems Unit of the biology department and the Mathematics Institute at Warwick University in Coventry, England.



WONDERS

by Philip and Phylis Morrison

The Sum of Human Knowledge?

Taking stock often arises out of concern over loss. An early instance is the Museum at Alexandria, whose library, a gem of classical learning, was torched amid the decline of Hellenistic culture about 1,500 years ago. Not one papyrus remains of all its treasures, long reasonably judged as some 600,000 scrolls. Following the customary chapter divisions of the Homeric epics (preserved as widely held best-sellers), we can reckon one scroll to be the equivalent of 25 pages of print. By this measure, Alexandria shelved 50,000 average books, of which we have only a few vivid texts from other sources.

China, India, Iraq and Iran also had rich collections back then, although the oldest clay "manuscripts," like almost all carvings in stone, never stored reams of text. We roughly estimate the entire literature of that time—its quarter of a billion people matching the U.S. of today, but far fewer of them readers—at about 100,000 books.

As World War II ended and many cities lay in ashes, Harvard University president James B. Conant expressed profound misgivings over the future of our store of human learning, newly opened to nuclear fire. Full partner to M.I.T. president Vannevar Bush in the leadership of U.S. civilian wartime research, Conant continued personal oversight of the huge army-led work on fission and fusion after the war. No civilian at war's end was closer to the heart of the Manhattan Project. Back on campus late in 1945, he privately charged the librarian of Harvard to consider the burdensome task of saving in multiple all the valuable texts in print anywhere.

The librarian's thoughtful report was soon at hand: worthwhile text amounted to 2.5 billion printed pages—that is, 10 million books. Although the postwar world population was only about 10 times larger than that of Alexandrian time, even the well-selected texts of 1945 outweighed the whole run of the classics by 100-fold. The librarian judged the task feasible but warned that it could not be carried out in secret and that the mere effort would deeply alarm the scholarly world. Conant put aside his foreboding, never again to pursue it during his lifetime.

In the second half of this century the broad theory of information developed, and the electronic digital computer has matured into a dominant industry. The grand Library of Congress will itself soon be outstripped in gross information storage by the tapes and hard disks held in myriad offices and households. Today the information unit of widest use is the byte, designed for the world of the desktop computer. It directly measures the digital memory needed to store the aggregate of our thought. A byte is in practice about a character of English text. The pulses that fly everywhere now encode not text alone but also formidable amounts of new data and images, the whole now measurable on a single scale.

The biggest byte makers are the television stations of the world.

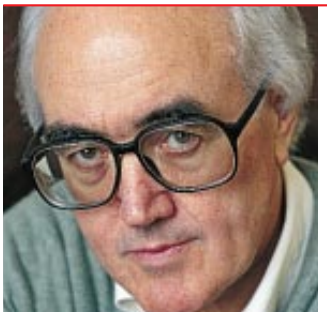
But this is new ground; the judgments needed are far from settled. We urge readers to bring their own views to the overarching topic we sketch here.

Fifty years ago Conant had sought to preserve by redundancy a library of 10 terabytes, or 10 trillion bytes of text. An ordinary book without illustrations, neither tome nor atlas, uses roughly one million bytes (one megabyte in the jargon of the day). Some 100 kilobytes of storage preserve a full-color photograph the size and quality of a modest book page; the frugal result will satisfy no artist's eye but is acceptable as compact storage. The storage needed for a book's text is thus doubled by only 10 such pictures. Speech, telephony, recorded music, vast lists, regiments of numerical data, photographs, cinema, video (much of it made interactive)—this is the diverse knowledge we fish for today in the sea of pulses.

The Library of Congress of 1998 holds only twice the postwar book count. But mainly because of its sound recordings (3.5 million of them!), the total store comes to a couple of petabytes—a million billion bytes, 1,000 times the space taken up by all its published texts. Holdings are richest when they are the professionally published work of talented historians, authors, artists, editors, composers or critics or of the theorists, labs and field stations. Yet that portion may well be outnumbered by our modern folk knowledge. Note also that publication in large editions does not add to information; only originals count. Hotel Bibles spread the gospel but do not augment it.

One plausible estimate of all text is based on the production of paper for writing and printing. Fresh text is less than 100 terabytes every year, 10,000 newspapers (by UNESCO count) included. Just about as much is held in all recorded music—the original sound, not copies, but not the mere scores either. Cinema contributes its flickering gallery of images: 5,000 films each year

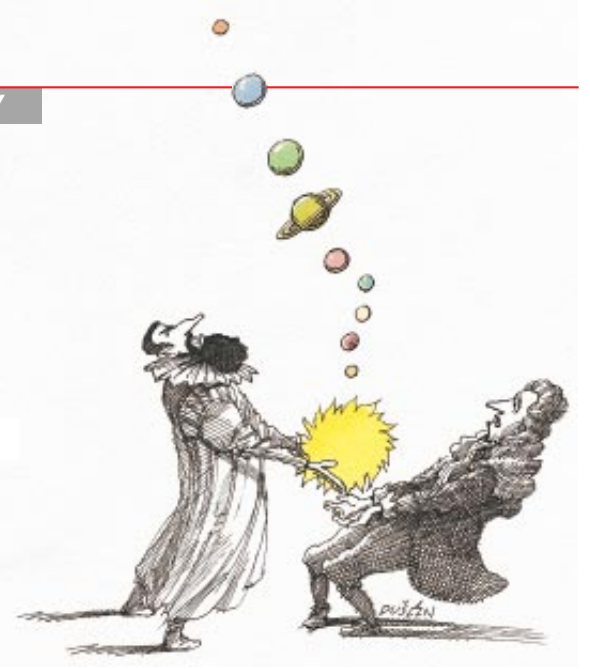
Continued on page 117



CONNECTIONS

by James Burke

Heavy Stuff



DUSAN PETRICIC

Hindsight always gives you 20/20 vision, doesn't it? So you can wonder why, in 1609, when Johannes Kepler (a German astronomer with the particularly unfortunate affliction of multiple vision) published his discovery of the mysterious attraction holding the planets in orbit around the sun, he didn't come up with a more scientific name for what was doing that to the heavenly bodies. Nope. He called it "Holy Spirit Force."

A couple of years later, when he met the English poet John Donne, Kepler gave him a copy of the book he'd written about his discovery to take back to the king of England. So says Donne's biographer and parishioner Izaak Walton, the fellow who put an entirely new cast on fishing with something he wrote in 1653 for all those out-of-work Anglican churchmen who had lost their jobs during the Puritan Commonwealth, entitled *The Compleat Angler: The Contemplative Man's Recreation.* Actually the thing wasn't entirely his. The chapter on flyfishing was written by a well-heeled literary pal and drinking buddy named Charles Cotton. Who also dabbled in a bit of translation. Cotton's 1685 version of Montaigne's *Essays* is still reckoned to be a classic.

The French writer in question was one of those types always on the edge of trouble, thanks to his vociferous skepticism regarding all forms of authority. At a time when all forms of authority still included people with the power of the thumbscrew, this was a risky game to play. Still, Montaigne (who transcribed "The only thing that's certain is that nothing is certain" on the ceiling beams of his study) managed to stay one step ahead of the theological sheriff. Traveling in 1580 to Florence, he was flabbergasted by the gardens of the Medici Villa di Pratolino, where he saw all the latest high-tech toys for boys (a.k.a. Renaissance princes). The gardens featured such marvels as movable artificial scenery, water-powered pipe organs and mechanical figures, as well as musical waterfalls—all the talented work of Bernardo Buontalenti, architect, engineer and man for all seasons. Buontalenti was also extravagantly into theater (he set one up at the Uffizi for the Medicis in 1585), stage design and boffo musical spectaculars of various kinds, whenever one of his princely patrons had some occasion to celebrate. This when he wasn't managing the Tuscany water supplies.

The man with a metal nose saw the impossible: a new star.

Another drain on Medici finances was Sir Robert Dudley, a questionable English adventurer on the run (with his mistress) from a second wife and numerous daughters back home. Where, he claimed, if it hadn't been for family infighting and accusations of illegitimacy, he would have been Earl of Warwick. In Italy he reclaimed marshland between Pisa and the sea, built canals, practically invented the city of Livorno and introduced English shipbuilding to the Medici navy. In 1647 Dudley produced *Concerning the Secret of the Sea,* the first complete collection of charts based on Mercator's projection. The hottest cartographers at the time were the Dutch—the Blaeu family in particular. Father and sons were mapmakers to the Dutch East India Company, whose main interests lay in exploring the fabled riches of the East and bringing back as much of them as they could lay hands on. Blaeu *père,* who had established the family printing business in Amsterdam around 1602, kept his maps up-to-date by nobbling captains straight off their ships so he could monopolize the latest navigational poop.

Patriarch Willem Janszoon Blaeu had started his working life (and his interest in maritime matters) as clerk to a herring company. After which he became assistant to a Danish astronomer who had a metal replacement nose, the irate Tycho Brahe. Who had been walking home from his lab one evening in 1572 and saw the impossible: a new star in heavens that were supposed to be unchanging. The first guy to whom he pointed out the nova and who didn't laugh at him was his friend the French ambassador, Charles de Dancey, venerable bon vivant and eminent good guy. A recent charitable act of his had been to go to Malmö prison and to offer to intercede, with the king no less, on behalf of an inmate. One of the period's eminent bad guys.

James Hepburn, fourth Earl of Bothwell, was a swashbuckler for whom things had badly unbuckled. Fleeing Scotland for Scandinavia by ship, he had the extraordinary ill luck to fetch up in Bergen, where the local duke didn't believe Bothwell's tale of being a Scots gent looking for work, because this particular ruler had a female relative whom Bothwell had done wrong by years before. A habit of his. Fact being, Bothwell was on the lam because he had left his wife in the lurch back in Scotland when it became clear he was about to be fingered for complicity in the murder of her previous husband. No big deal, had

Mrs. Bothwell not been otherwise known as Mary, Queen of Scots.

Who was in her own mess because she was a legitimate claimant to the English throne, and this was not good when that seat was already occupied by Elizabeth. It was inevitable that a hothead conspirator like Mary would end up losing her own head. Elizabeth, meanwhile, had other things to worry about. Such as Sir Francis Drake et al. crossing the Atlantic and discovering to their horror that on the other side their compass needles no longer pointed at the North Star. Crisis-level investigations conducted by Her Majesty's personal physician and science maven William Gilbert convinced himself the earth was a giant magnet with a North Pole that attracted needles and distinct from the celestial north pole of the earth's rotation.

Which got a legion of other famed experimenters noodling away. Namely a German mayor called Otto von Guericke, who fashioned an earth model out of sulfur in 1650, rubbed it vigorously and brought various needles close, to see which way they would point. This *was* the heyday of propeller-heads. But during a rubbing session, Guericke idly footnoted that the ball gave off a bang and a spark. Turned out to be electricity. Soon everybody was rubbing, including Francis Hauksbee.

In London, in 1706, Hauksbee demonstrated an amazing electricity generator to the Royal Society: a small, evacuated glass globe mounted on a spindle. When the globe was spun at high speed, and while a hand was held against the glass, a purple light appeared inside the sphere. Nearby threads became attracted to the glass. Moving on to finer things, Hauksbee discovered that if two slim glass tubes were inserted into a liquid, the liquid rose higher in the narrower tube. Because Hauksbee suspected something might be happening between the liquid and the glass sides of the tube, he turned to the man whose middle name was "attraction": Isaac Newton.

By 1717 Newton was able to explain capillary action as an effect of surface tension. At that time the great man was also the most famous person in the entire universe for clearing up the mystery of Kepler's "Holy Spirit Force" and giving it the name by which both it and Newton are known today: gravity.

Enough of these weighty matters. SA

Wonders, *continued from page 115*

now, but over the century since the Lumière brothers only about one petabyte, mostly black-and-white and brief. People's own images, their personal snapshots, are many. Their number worldwide can be estimated at 50 billion a year from the production of photographic emulsions (digital snapshots are still few). Even in highly compressed approximation, these images of cute kids and family gatherings amount to half a petabyte a year—perhaps 10 petabytes summed over the years.

The biggest byte makers are the television stations of the world; although it is hard to correct for innumerable repeats, our best source puts their originality at one tenth of all they send out and so allots them under 100 petabytes annually.

Telephony adds 10 trillion call-minutes a year in the U.S.; the audio sounds of world calls add up to some 1,000 petabytes worldwide, or a few exabytes. Speech face-to-face must add several exabytes of sound. (Their content transcribed from acoustics into written words of text is 1,000-fold smaller, only a few petabytes yearly.) Sum it up to reach a total at hand of 10 exabytes, mainly audio sampling of transitory speech, with not much of human discourse so far recorded or recalled.

Search, not compression, will be the main challenge. Validated professional output, from TV soccer to World Wide Web pages to weather maps and *Titanic,* is perhaps 100 petabytes, nearly all of it from TV. Book texts are tiny, and data, maps and well-packed photographs add only a few petabytes more. In 10 years the manufacture of magnetic disks and tapes may hold anything that could be retold in pixel or word, though not yet spanning the rich, unspoken skills of hand and eye and social life.

Such an understanding is dryly mechanocentric. One might foresee a new tide of creativity. Biography could include not only the famous and the rascals but also illustrated records, exciting beyond any terse genealogy: Everyperson and her cousins on digital file. Then family chronicles would appear by the 100 million. No estimate of the eventual human store seems quite credible as yet.

A fine piece recently done for the Getty Information Institute by Michael Lesk of Lucent Technologies was our major impetus for this month's column. SA

WORKING KNOWLEDGE

TOUCHPAD POINTING DEVICE

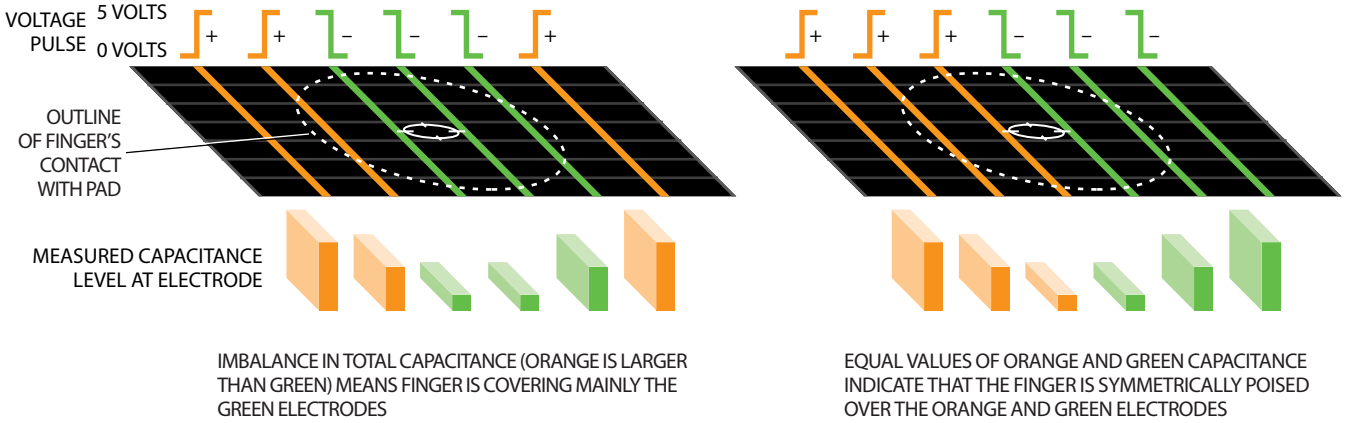
by George Gerpheide
CEO and Founder, Cirque Corporation

The most common pointing device on new laptop computers is the touchpad, a small black or gray rectangle typically located below the keyboard. Moving a finger across the pad causes the cursor to move similarly on the screen.

Touchpads began appearing on laptops only about four years ago, but they have already displaced built-in trackballs as the standard pointing device on portable computers. Today more than two thirds of all laptops sold come equipped with touchpads. (The remainder, mostly IBM and Toshiba models, use the small, red joystick-like pointer, which resembles a pencil eraser and is located between the "G," "H" and "B" keys on the keyboard.) Touchpads offer better ergonomics for many people, such as those afflicted with arthritis. And because the devices are completely sealed and have no moving parts, dirt or grit cannot get inside them, making them more suited to dusty or corrosive environments, such as workshops, factory floors and garages.

The most popular type of touchpad is capacitance-based. It measures changes in capacitance as the user's finger alters the imperceptibly minute electric fields above the pad.

ILLUSTRATIONS BY SLIM FILMS



IMBALANCE IN TOTAL CAPACITANCE (ORANGE IS LARGER THAN GREEN) MEANS FINGER IS COVERING MAINLY THE GREEN ELECTRODES

EQUAL VALUES OF ORANGE AND GREEN CAPACITANCE INDICATE THAT THE FINGER IS SYMMETRICALLY POISED OVER THE ORANGE AND GREEN ELECTRODES

FINDING THE FINGER involves shifting two patterns of voltage pulses. To check the capacitance at every single point at which the electrodes cross would take too much time, resulting in sluggish response to finger movement. Instead two patterns (*left and right diagrams*) of positive (*orange*) and negative (*green*) pulses are applied to the electrodes and the electric charge from the mutual capacitance is measured. Calculations involving the ratio of the total charge measured for each pattern locate the finger in relation to the boundary between regions of positive and negative pulses. The two patterns must shift as the finger moves to keep this boundary near the center of the finger. In this illustration the patterns are shown for only the set of vertical parallel electrodes; in operation, the same shifting patterns are executed on the other (orthogonal) set of electrodes, enabling two-dimensional tracking of finger movements at speeds up to 100 centimeters (40 inches) a second.

