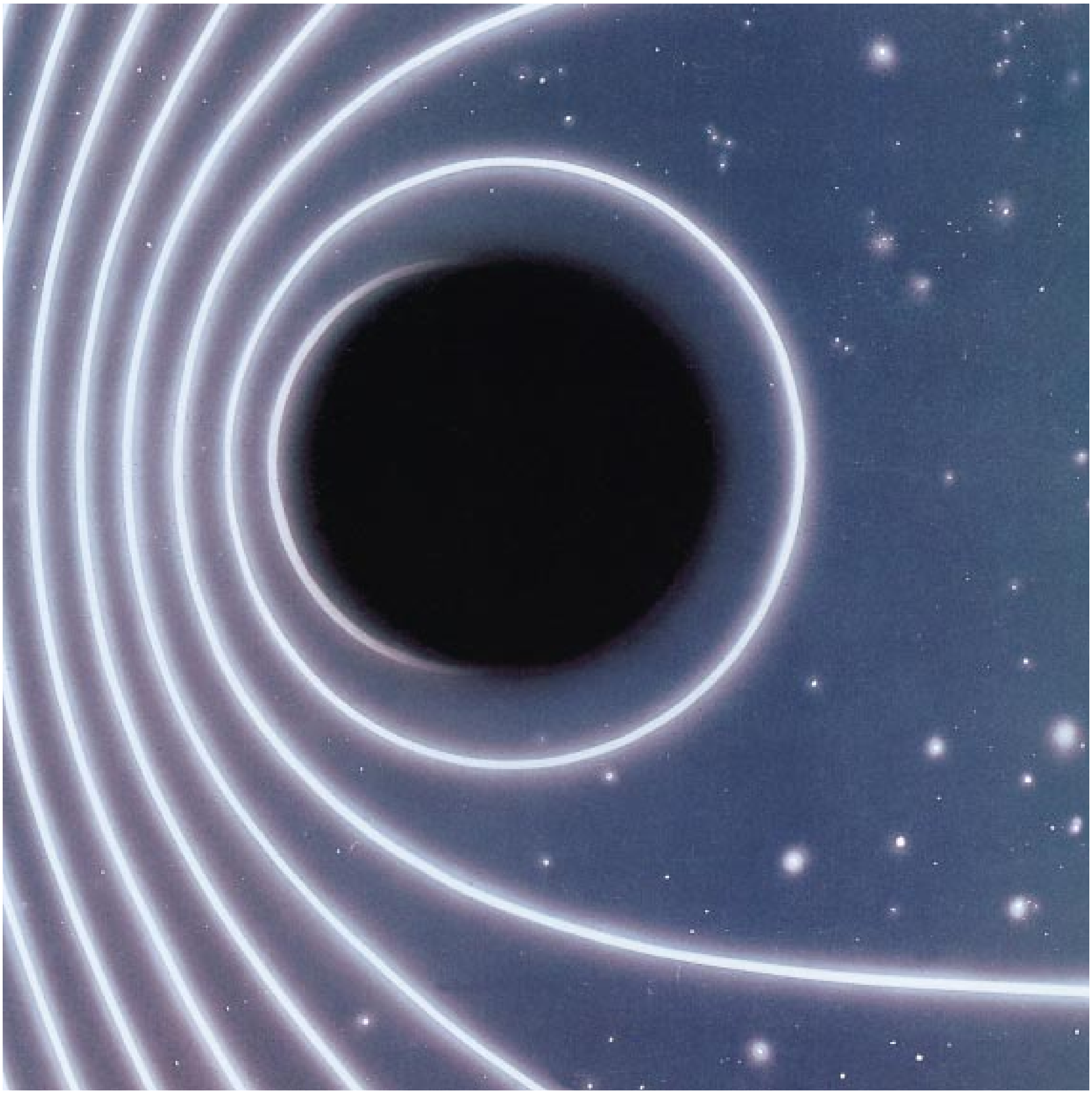
# SCIENTIFIC AMERICAN

MARCH 1993
$3.95

*Rewriting genetics with the new ABCs of DNA.*

*The technology of flat-panel displays.*

*Provoking the immune system to fight cancer.*



***Light rays*** *bent by the intense gravity near a black hole resolve a paradox in Einstein's theory of relativity.*

SCIENTIFIC AMERICAN

March 1993 Volume 266 Number 3

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 1993 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. and Canada 800-333-1199; other 515-247-7631. 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, or fax: (212) 355-0408.



# DEPARTMENTS

THE COVER illustration depicts light rays curved by the gravitational field of a black hole. The bending of light is the key to understanding many of the paradoxical effects predicted to occur near a black hole. In a region of space free of gravitational fields, light rays travel in perfectly straight lines. Near a black hole, according to Einstein's general theory of relativity, light rays are curved by varying amounts and can even be circular (see "Black Holes and the Centrifugal Force Paradox," by Marek Artur Abramowicz, page 74).

## THE ILLUSTRATIONS

Cover illustration by Alfred T. Kamajian

| Page | Source |
|---|---|
| 67 | Hank Morgan, Rainbow, Inc. |
| 68–69 | Jana Brenning (*top*), Sidney M. Brown Photography (*bottom*) |
| 70–72 | Laurie Grace |
| 75 | Jack Harris/Visual Logic |
| 78–81 | Jared Schneidman/JSD |
| 83 | Bernard Sordat, Swiss Institute for Experimental Cancer Research, Lausanne |
| 85–89 | Ian Worpole |
| 90–91 | John S. Foster, Jane Frommer and Jacquelin K. Spong, IBM Thomas J. Watson Research Center (*left*), Michael Goodman (*right*) |
| 92 | Photonics (*top left*), Planar Systems (*top right*), Michael Goodman (*bottom*) |
| 94 | Michael Goodman |
| 96 | Laurie Grace |
| 97 | Carolco |
| 101 | Waina Cheng/Bruce Coleman, Inc. |
| 102 | Patricia J. Wynne |
| 103–105 | Guilbert Gates/JSD |
| 106 | Patricia J. Wynne |
| 108–109 | Sophie de Beaune; courtesy of Musée des Antiquités Nationales, Saint-Germain, France |
| 110 | Jim Wagner (*left*), courtesy of Sophie de Beaune (*right*) |
| 111 | Courtesy of Sophie de Beaune (*left*), Johnny Johnson (*right*) |
| 112 | Johnny Johnson |
| 113 | Tom McHugh/Photo Researchers, Inc. |
| 115 | Michael Goulding |
| 116 | Joe LeMonnier |
| 117 | Michael Goulding |
| 118–119 | Patricia J. Wynne |
| 120 | Michael Goulding |
| 122–123 | Marilyn A. Houck, Texas Tech University (courtesy of *Science*) |
| 124–125 | Michael Goodman (*top*), Culver Pictures, Inc. (*bottom left*), Nik Kleinberg (*bottom right*) |
| 126–127 | Michael Goodman |
| 130 | Davis Freeman |
| 131 | Forest McMullin/Black Star |
| 145 | Andrew Christie |


# SCIENTIFIC AMERICAN®

*Established 1845*

**EDITOR:** Jonathan Piel

**BOARD OF EDITORS:** Alan Hall, *Executive Editor;* Michelle Press, *Managing Editor;* Timothy M. Beardsley; W. Wayt Gibbs; Marguerite Holloway; John Horgan, *Senior Writer;* Philip Morrison, *Book Editor;* Corey S. Powell; John Rennie; Philip E. Ross; Ricki L. Rusting; Russell Ruthen; Gary Stix; Paul Wallich; Philip M. Yam

**ART:** Joan Starwood, *Art Director;* Edward Bell, *Art Director, Graphics Systems;* Jessie Nathans, *Associate Art Director;* Nisa Geller, *Photography Editor;* Johnny Johnson

**COPY:** Maria-Christina Keller, *Copy Chief;* Nancy L. Freireich; Molly K. Frances; Daniel C. Schlenoff

**PRODUCTION:** Richard Sasso, *Vice President, Production;* William Sherman, *Production Manager; Managers:* Carol Albert, *Print Production;* Tanya DeSilva, *Prepress;* Carol Hansen, *Composition;* Madelyn Keyes, *Systems;* Leo J. Petruzzi, *Manufacturing & Makeup;* Carl Cherebin

**CIRCULATION:** Lorraine Leib Terlecki, *Circulation Director;* Cary Zel, *Circulation Manager;* Rosa Davis, *Fulfillment Manager;* Katherine Robold, *Newsstand Manager*

**ADVERTISING:** Robert F. Gregory, *Advertising Director.* OFFICES: NEW YORK: Meryle Lowenthal, *New York Advertising Manager;* William Buchanan, *Manager, Corporate Advertising;* Peter Fisch, Elizabeth Ryan. Michelle Larsen, *Director, New Business Development.* CHICAGO: 333 N. Michigan Avenue, Chicago, IL 60601; Patrick Bachler, *Advertising Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; Edward A. Bartley, *Detroit Manager;* William F. Moore. WEST COAST: 1554 S. Sepulveda Blvd., Suite 212, Los Angeles, CA 90025; Kate Dobson, *Advertising Manager;* Lisa K. Carden, Lianne Bloomer, San Francisco. CANADA: Fenn Company, Inc. DALLAS: Griffith Group

**MARKETING SERVICES:** Laura Salant, *Marketing Director;* Diane Schube, *Promotion Manager;* Mary Sadlier, *Research Manager;* Ethel D. Little, *Advertising Coordinator*

**INTERNATIONAL:** EUROPE: GWP International, Düsseldorf; Roy Edwards, GWP International, London; Kurt Bühlmann, GWP International, Zürich; Vivienne Davidson, Linda Kaufman, Intermedia Ltd., Paris. SEOUL: Biscom, Inc. TOKYO: Nikkei International Ltd. SPECIAL PROJECTS: Barth David Schwartz, *Director*

**ADMINISTRATION:** John J. Moeling, Jr., *Publisher;* Marie M. Beaumonte, *Business Manager*

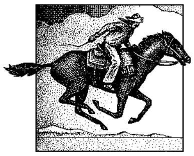
**SCIENTIFIC AMERICAN, INC.**
415 Madison Avenue
New York, NY 10017
(212) 754-0550

**PRESIDENT AND CHIEF EXECUTIVE OFFICER:** John J. Hanley

**CORPORATE OFFICERS:** *Executive Vice President and Chief Financial Officer,* R. Vincent Barger; *Vice Presidents:* Jonathan Piel, John J. Moeling, Jr.

**CHAIRMAN OF THE BOARD:** Dr. Pierre Gerckens

**CHAIRMAN EMERITUS:** Gerard Piel



# LETTERS TO THE EDITORS

## Reforming Health Care

As one who left the U.S. Department of Health and Human Services in the mid-1980s, disappointed by Washington's failure to address health care reform, I read Rashi Fein's article ["Health Care Reform," SCIENTIFIC AMERICAN, November 1992] with considerable interest. Although he covers a number of issues very well, others are omitted, underplayed or misrepresented, making it difficult to accept or reject his recommendations. Of particular concern is his treatment of cost.

His chart of per capita spending over the past decade shows the U.S. a clear cost runaway. But annual per capita increases in the U.S., in local currency adjusted for local inflation, have been about 5 percent—less than the average of 5.5 percent for the G-7 countries.

Unfortunately, one is left with the uneasy feeling that Fein has looked abroad and selected what he liked, based on conclusions and data that are at best suspect and at worst wrong. No industrialized country has cost escalation under control. Global budgeting, single-payer systems, "play or else" systems and "health care planning" haven't worked. Indeed, it is at least arguable that the only thing that hasn't been tried and seen to fail is serious competition, "managed" or otherwise.

CHARLES D. BAKER
College of Business Administration
General Management Department
Northeastern University

The answer is not more paternalism but less, getting people to face the necessary decisions about what medical coverage is really worth the cost. The proper role of government is to do what Oregon tried to do: use medical benefit per dollar spent as a criterion for choosing what services will be provided to all, regardless of ability to pay. With appropriate minimum standards in place, there is no reason not to leave the rest up to individual choice.

ALEXANDER RAWLS
Palo Alto, Calif.

Fein's conclusion that a Medicare-type system would be best is perplexing. Medicare is perhaps the single greatest cause of failure in the present system. It is a prime cause of the cost shifting that has resulted in millions of uninsured persons, most notably among the self-employed and the employees of small businesses.

A single-payer system covering all Americans and similar to the present Medicare plan would contain none of the incentives to allocate resources properly that are necessary in a free market economy. Consider how many Cadillacs or Mercedes would be on the road if one could choose those vehicles without paying for them. Further still, consider the costs of automobile insurance if every oil change or lubrication required submission for reimbursement.

MARK O. DIETRICH
Framingham, Mass.

PETER GORLIN
Saints Memorial Medical Center
Lowell, Mass.

Fein offers a practical plan for universal health insurance with a single carrier that should cut the paperwork and provide better medical care at lower cost. Many high-tech procedures are done because they pay the doctor much more than he or she gets for careful observation of the patient. In his preface to *The Doctor's Dilemma,* George Bernard Shaw remarked that if a doctor were paid to cut off a man's leg, he might reason that he needed the money more than the man needed the leg. That is a strong reason to pay doctors generously for their time and skills but not for the high technology of the operations or tests that they perform.

SAM I. LERMAN
Canton, Mich.

## Racing to Bad Health

The Mount Evans Hill Climb bicycle race starts in Idaho Springs, Colo. (elevation 7,542 feet), and continues for 28 miles to the summit (14,264 feet). It has earned some great nicknames, like "The Only Road Race in North America Where the Officials Need Oxygen."

"Mountain Sickness," by Charles S. Houston [SCIENTIFIC AMERICAN, October 1992], makes this event sound almost impossible. It requires riders to do almost exactly what should induce mountain sickness: make a rapid ascent, suffer dehydration and achieve an elevated heart rate and very high respiration. Strangely, I have not heard of anyone having serious complications; I have seen people collapse at the finish, but that is not much different from any other intense bike race.

ERIC BURT
Alamosa, Colo.

***Houston replies:***
Dozens of exhausting races are run at altitudes where pulmonary edema would seem likely. The explanation is that runners get up and down again too fast for overt edema to appear. Only rarely is high-altitude pulmonary edema clinically evident until 24 to 36 hours after reaching altitude. That fact also often protects the speed climber.

## Rumors of Its Death

I was amused to read about "the final theory of physics" and the end of science in "The New Challenges" [SCIENTIFIC AMERICAN, December 1992]. Forty years ago when I was starting my career, it also seemed that little was left to discover. Yukawa's meson had been established as the explanation for nuclear forces. A "few loose ends" like the antiproton would have to wait (until 1956) to be discovered. But that would be it.

Today when I teach modern physics, most of what I talk about was discovered in the past 40 years. Of the 17 fundamental particles I discuss, only four were known before then.

Some of my distinguished friends are determined to find the theory of everything before they are too old to understand it. I am very content in my belief that there will be much to be discovered by my young students and even perhaps by my newly born grandson.

LINCOLN WOLFENSTEIN
Department of Physics
Carnegie Mellon University

ERRATUM

The drawing of the oligosaccharide on page 84 of the January issue erroneously shows extra hydroxyl groups on the carbons linking the glucose subunits.

# 50 AND 100 YEARS AGO

MARCH 1943

"One of the greatest questions of the present war is whether modern science is capable of preventing the recurrence of epidemics which in all past wars cost more lives than were lost in battle, according to Dr. Bernhard J. Stern, in a paper presented for a Cooper Union symposium on 'Medicine in Wartime.' The influenza epidemic that followed World War I killed more victims in a few months than all the armies in four years. In the United States alone perhaps half a million died; the worldwide mortality is estimated at from ten to twenty-one million. Yet there are elements of hope in the global conflict now raging, Dr. Stern believes. There have been prodigious advances in epidemiology since the last war, and the developments in the field of sulfa drugs mark one of the most brilliant chapters in the history of medicine."

"Forty-six years after building America's first successful, full-sized submarine, Simon Lake is as full of ideas as ever. There is a common impression that the submarine has reached its peak, but Lake shakes his gray shock and declares that the boat is in its infancy. He still preaches commercial submarines, and will not consider his life work complete until he has proved their value to the world. Recently he proposed a fleet of cargo submarines as a means of solving the shipping shortage. He says they can be built as cheaply as tankers, will cost less to operate and can easily escape raiders by submerging."

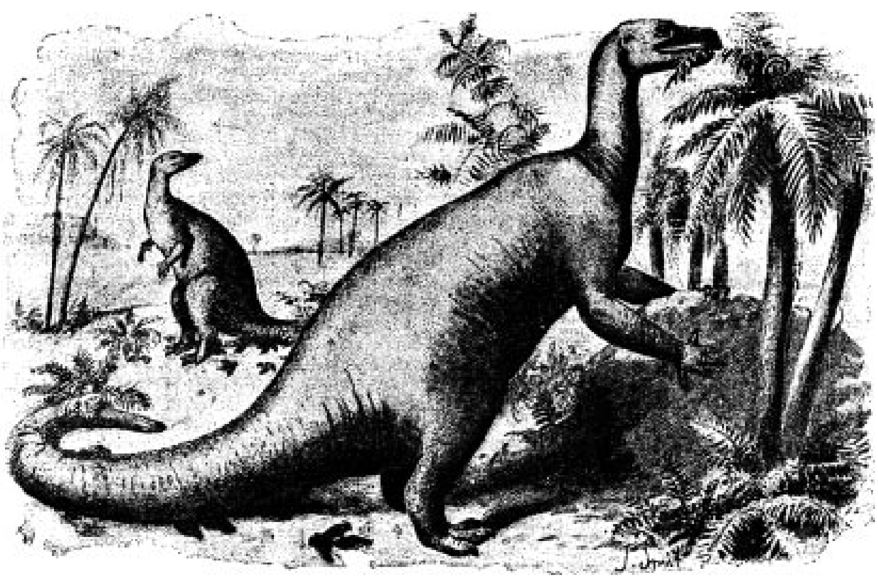
"Specimens of pseudo-fossil men can be classified into two general types: first, the 'normal' individual who represents, in one or another feature, a more primitive appearance than the average for his group; and, second, the individual who, through a glandular disorder, has suffered a marked thickening of the bony structure. The writer [Loren C. Eiseley] can testify that he long coveted the skull of an unsuspecting colleague who approached close to the Neanderthal type in one or two characteristics of the skull. I say one or two advisedly. Viewed in its entirety, my good friend's cranium would have deceived no competent anatomist into imagining him to be one of our early forerunners. If, however, the right fragment of his skull—the 'primitive' part—had been recovered from an archeological deposit of some antiquity, discussion might have arisen."

MARCH 1893

"'In all the projects for signaling Mars proposed by learned Thebans, I have seen no reference to what seems to the unlearned layman the most self-evident difficulty. It is that the bright side of Mars is always toward us. If signals were sent at night from the dark side of our globe by artificial light, the flashes would have to be of such intensity that they could be seen through sunlight of that planet. The planet Venus, however, can at rare intervals be seen by day. Flashes from mirrors might at such times be sent to it. Such flashes would fall on its dark side and would be seen, if at all, by its inhabitants in their night time.'—T. M. Anderson, Col. U.S.A., Vancouver Barracks"

"The death of a centenarian Italian in a Norfolk town the other day, whose checkered life-history included service in Napoleon's 'Grande Armée' during the disastrous Russian campaign of 1812, recalls attention to the fact that of all that host the Neapolitan contingent, 10,000 strong, withstood the cold and privation much better than the other divisions, recruited as these were mainly from Northwestern and Central Europe. The view taken of the fact was this: That the Italians, born and reared in the sunny South, retained so much 'caloric' in their systems that their supply of it continued long after their fellow soldiers from less favored climes had used up theirs. In support of this the experience of other Italians was invoked who, as teachers or artists, had settled in English or Scottish educational centers, and whose power of weathering the first northern winter was much greater than during the second and third, by which time, it was contended, their supply of 'caloric' was exhausted."

"Cuvier it was whose fine imaginative reasons invented the great science of comparative anatomy and palaeontology. His splendid knowledge of existing beasts and birds enabled him to reconstruct from a fossil skull or a vertebra, sometimes from but a single tooth, the long-extinct creature in its true semblance as it had lived—to clothe it with flesh and skin, and show it in imagination, in the haunts in which it lived and moved. This, which Baron Cuvier did in graphic description of great scientific and literary beauty, Mr. Hutchinson, in his work on Extinct Monsters, published by Messrs. Chapman & Hall, has now done popularly [*see illustration at left*]."



***Iguanodon***

## Antarctic Meltdown

*The frozen continent's ice cap is not as permanent as it looks*

Flying toward the South Pole in a C-130 military transport, one might not think of the vast ice sheet below as an ephemeral phenomenon. The ice smothers virtually the entire Antarctic, an area as large as the U.S. and Mexico combined. Toward the center of the continent, ranges of towering mountains diminish and finally disappear below ice several kilometers thick. The ice cap is so massive that it compresses the underlying rock; if some Titan pried the ice away, the earth would spring up more than 100 meters.

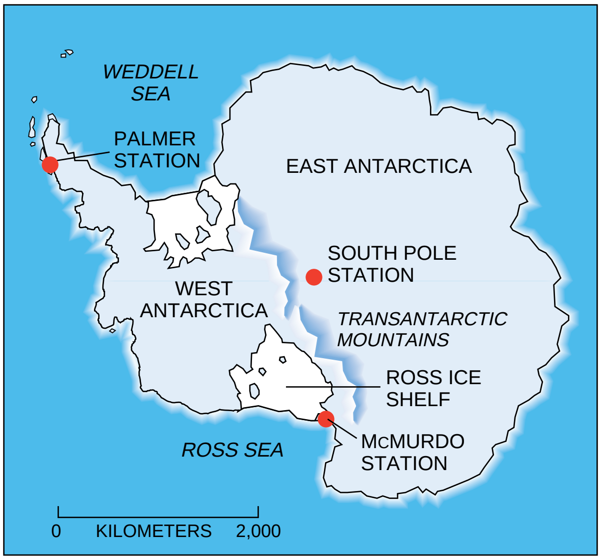
Yet signs of flux are visible, especially on the perimeter of the continent. Standing on a peak of Ross Island, home to McMurdo Station, Antarctica's largest research base, one can see mighty ice streams and glaciers descending to the sea, where they shed mass by calving icebergs and by melting. (An ice stream flows through stationary or slower-moving ice, whereas a glacier is bounded on each side by rock.) Occasional flurries of snow provide a reminder that precipitation is the ultimate source of all Antarctica's ice.

An even more dynamic picture of the ice sheet emerges from conversations with some of the scores of scientists who journey to the coldest, most hostile environment on the earth each austral summer to study the ice cap. They cite a growing body of evidence that the ice has fluctuated dramatically in the past few million years, vanishing outright from the entire continent once and from its western third perhaps several times. Collapses might be triggered not only by climatic change, such as global warming, but also by factors that are far less predictable, such as volcanic eruptions occurring underneath the ice.

"We have had a very simple view of the ice sheet's history," says Gary S. Wilson, a geologist at Victoria University of Wellington in New Zealand, "and we're only beginning to learn that it's very complex." The significance of these findings is enormous: if the Antarctic ice cap disintegrates, sea levels could surge by as much as 60 meters. "New York is

***BEARDMORE GLACIER flows from the Antarctic plateau into the Ross ice shelf. The glacier is about 25 kilometers wide.***

GALEN ROWELL *Mountain Light*




LAURIE GRACE

*WEST AND EAST ANTARCTICA, divided by the Transantarctic Mountains, may react quite differently to climatic change.*

going to be underwater," Wilson adds with a grin.

Until recently, most researchers believed the Antarctic ice cap formed during a cool era about 14 million years ago and has persisted with relatively minor shrinking and swelling since then. A simple mechanism was thought to keep the ice near equilibrium in spite of climate changes: as temperatures rose, calving and melting would increase, but so would evaporation of seawater and precipitation over the continent.

The boldest challenge to this view has come from workers led by David M. Harwood of the University of Nebraska and Peter N. Webb of Ohio State University, who contend that only three million years ago the Antarctic ice cap was virtually nonexistent. Harwood describes himself as a "garbage-pile geologist" who rummages through heaps of debris left behind by glaciers. In the mid-1980s he and Webb found some unusual glacial refuse in the Transantarctic Mountains, a rocky spine that transects the continent. The deposits contained the fossil remnants of minute marine organisms called diatoms and of a species of beech tree common to the Southern Hemisphere. The diatoms were similar to ones found previously in ocean-floor sediments three million years old.

The group concluded that three million years ago the ice sheet had collapsed, transforming the continent into a cluster of islands divided by open sea. The beech trees lived on islands that were to become the Transantarctic Mountains, and the diatoms lived in marine basins to the east of those islands. As temperatures fell and the continent froze once again, the expanding East Antarctic ice sheet shoved the diatoms up into the Transantarctic Mountains, where Harwood and Webb found them along with the beech fossils three million years later.

## Are Scientists Too Messy for Antarctica?

Scientists come to Antarctica not just to study potential catastrophes such as the ozone hole and the unstable ice cap. Biologists foray onto and under the sea ice to study seals, penguins and fish with antifreeze in their blood. Geologists tramp through mountains searching for fossils and other clues to the continent's past. On the 3,000-meter-high ice plain of the South Pole, astronomers peer through the clearest atmosphere on the earth at galaxies and other cosmic mysteries.

Each austral summer the major sponsor of these projects, the National Science Foundation, brings several journalists here to see these experiments firsthand. The reporters are housed, fed and flown to permanent stations and field sites for interviews. The red-carpet treatment has a purpose: ideally, the reporters will write stories that the NSF can use to justify the tax dollars it spends on Antarctic research, which will total $221 million in 1993.

But one of the biggest stories over the past few years has been an embarrassing one: the degradation of this delicate, frozen continent caused by human interlopers. The problem is most pronounced at McMurdo Station on Ross Island, the largest of the three permanent sites the U.S. maintains in Antarctica. (The other two are Amundson-Scott Station, at the geographic South Pole, and Palmer Station, on a peninsula south of Tierra del Fuego.)

McMurdo's population fluctuates from a low of about 250 in the sunless austral winter to about 1,200 in the perpetual daylight of summer. For every scientist working here, there are roughly four civilian and military personnel who provide support, running the cafeteria and power plant, flying the planes and helicopters—and, increasingly, managing waste.

McMurdo's muddy streets, warehouse-style architecture and volcanic-slag terrain give it the no-frills look of a mining town. When a reporter remarks on the contrast between the town and its setting, David M. Bresnahan, the senior NSF official at McMurdo, bristles. "If you think McMurdo is ugly now, talk to someone who was here three, or four, or 10 years ago," he says.

Beginning in the late 1950s, when the U.S. military founded McMurdo, crews dumped on the land and into the sea everything from food waste and junked machinery to oil, PCBs and radioactive waste. Over the past few years, complaints from Greenpeace and other environmental groups have led to a massive cleanup effort.

The dump has been swept clean and the garbage either burned or packed in containers for shipping to the U.S. McMurdo officials now claim their recycling program is the most thorough in the world. But problems still exist. A small, man-made harbor abutting McMurdo remains "as contaminated by hydrocarbons as any temperate harbor on the planet," says John S. Oliver of Moss Landing Marine Laboratories in California. Oliver has recommended pumping oxygen or other nutrients into the sediments to encourage the growth of bacteria that might break down the hydrocarbons.

McMurdo's raw sewage still spews directly into the sound. In response to international regulations, the NSF recently began macerating the sewage before discharging it. "Instead of seeing big chunks, you see a lot of little chunks," says Gordon A. McFeters, a microbiologist from Montana State University.

Last year McFeters found that human coliform bacteria from the sewage are being sucked into the water intake pipe for the base's desalination plant, which provides drinking water. He is worried that infectious viruses such as

That conclusion has been vigorously disputed. George H. Denton, a geologist at the University of Maine who has worked in Antarctica almost every summer since 1968, says his research indicates that the valleys of the Transantarctic Mountains have been frigid and relatively lifeless for at least 14 million years. James P. Kennett of the University of California at Santa Barbara, another veteran Antarctic geologist, suggests that the diatoms found by Harwood and Webb in the Transantarctic Mountains might have been blown there from some region outside of Antarctica. "Diatoms can end up anywhere," he says.

But Harwood and Webb's theory has gained some support from a team that includes Wilson and another geologist from Victoria University, Peter J. Barrett. The group collected cores from the floor of a fjord abutting the Transantarctic Mountains and discovered a layer of volcanic ash containing diatoms similar to those uncovered by Harwood and Webb. By measuring the radioactive decay of argon isotopes in the ash, the investigators concluded that the ash and the diatoms were three million years old. These findings, the researchers declared in *Nature* last October, "confirm" that deglaciation had occurred.

Wilson acknowledges that the issue remains unsettled. This season, he joined a team led by Harwood that is searching for more information on conditions during the Pliocene, a period extending from two to five million years ago. In an interview at McMurdo, just before heading out for two months of fieldwork, Wilson pointed out that the Pliocene's climate may have been only slightly warmer than today's, "so it's not only important but essential to know what was going on."

Most workers agree that at least since the Pliocene period, the East Antarctic ice sheet, defined as the region east of the Transantarctic Mountains, has remained relatively stable. The West Antarctic ice sheet is another matter. Whereas the East Antarctic consists of a single tectonic plate, the West Antarctic landmass is a jumble of small plates and geologically active rifts. Its average elevation is quite low; in fact, most of the West Antarctic ice rests on land below sea level.

The West Antarctic is also dominated by two seas, the Ross and the Weddell, whose landward regions are covered by thick, floating shelves of ice. These shelves act both as catchments for and impediments to the glaciers and ice streams feeding into them. Some researchers have speculated that if warmer, rising oceans were to melt this ice, the entire western sheet might quickly disintegrate, pushing global sea levels up by five or six meters.

In fact, Reed P. Scherer, a geologist at Ohio State, has asserted that a scenario like this occurred no more than two million years ago and probably much more recently. Scherer based his proposal on diatom-bearing cores from underneath a West Antarctic ice stream. A group led by W. Barclay Kamb of the California Institute of Technology obtained the cores in 1989 by boring through 1,000 meters of ice with heated, pressurized water.

The species contained in the sediments were known to have existed from two million to 100,000 years ago. Scherer's best guess is that the diatoms collected by Kamb's group were deposited

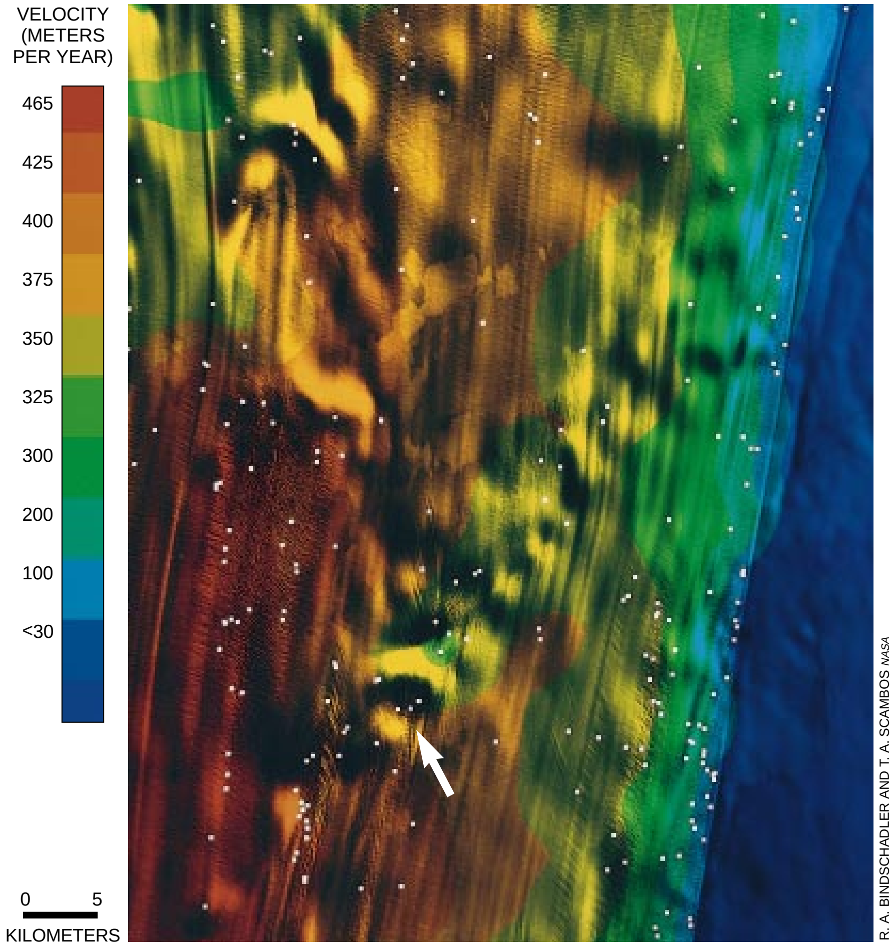
hepatitis might survive the distillation process, which heats the water to only about 80 degrees Celsius, and trigger an epidemic.

The ultimate solution to the environmental problems would be to replace messy human scientists with robots, which require no food and generate no waste. That is one possible outcome of research being funded in Antarctica by the National Aeronautics and Space Administration. The technologies NASA is testing, which could also be used for exploring Mars, include solar-powered telecommunications systems and a remote-controlled robot that swims below the ice of Antarctica's few lakes. The showpiece is an eight-legged robot named Dante, built at Carnegie Mellon University. Its mission was to crawl down into the crater of Mount Erebus, a spectacular, 3,794-meter active volcano that dominates the landscape of Ross Island.

Dante stumbled even before it reached the Antarctic. Last October, during a test run on an artificial slag heap in Pittsburgh, half of Dante's legs broke. Workers quickly rewelded the legs and shipped the robot to McMurdo anyway. Dante was crippled again in January, minutes after it had begun descending into the smoldering crater, when a fiber-optic control cable snapped. "An unqualified success," a NASA spokesperson insisted to a reporter. But for now, it seems, research in Antarctica will continue to be done by food-consuming, waste-producing scientists. *—John Horgan*

JOHN HORGAN

***McMURDO STATION boasts a bowling alley, a fitness center, a chapel, a state-of-the-art $23-million laboratory and three bars.***




***VELOCITIES VARY in a single West Antarctic ice stream, as shown in this Landsat image. Velocities are color-coded; white marks indicate points of measurement. The arrow shows a depression that could lie above a volcanic hot spot.***

in sediments during a warm interglacial period about 400,000 years ago. "The deep-sea and climate records all indicate this was a time of unusual warmth," he notes.

According to Scherer, his data do not limit the West Antarctic ice sheet to just a single collapse in the past two million years. Indeed, a computer model done by Douglas R. MacAyeal of the University of Chicago lends weight to the possibility of multiple collapses even within the past million years. The model also suggests that collapses may have been relatively fast and unpredictable.

MacAyeal's model, which he presented in *Nature* last September, was based on data about the climate's behavior during the past million years and on current information about the dynamics of ice streams, which are known to require liquid water for lubrication at their base in order to move. MacAyeal found that the ice behaved erratically during the entire time span and collapsed outright three times—750,000, 330,000 and 190,000 years ago.

These collapses did not coincide with warm periods. One important reason, MacAyeal maintains, is that fluctuations in surface temperatures can take millennia to propagate down through ice sheets. Occasionally, a wave of relative warmth would provide just enough heat to melt the underside of a previously frozen, static ice stream. "You get a phase transition at the base of the ice," MacAyeal explains, "and—poof!—you get acceleration of the ice."

Satellite data reveal that ice streams in the West Antarctic do indeed behave erratically. For six years, Robert A. Bindschadler, a glaciologist at the National Aeronautics and Space Administration Goddard Space Flight Center, has been analyzing Landsat photographs of the ice streams, marking their progress by measuring the movement of crevasses and other surface features. He has found that some ice streams hurtle along at more than two kilometers a year and are losing much more mass than they are gaining through precipitation; others show no discernible movement. Velocities can vary widely even within the same ice stream.

"Things are wildly deviant from a steady-state system" in the West Antarctic, agrees Donald D. Blankenship, a geologist at the University of Texas at Austin. The implication of this finding, he notes, is that the near-term behavior of the West Antarctic ice might depend less on external, climatic factors than on internal ones—such as conditions at the base of the ice.

Until recently, Blankenship elaborates, glaciologists believed ice streams glide on a thin film of pressurized water. In the mid-1980s he found evidence that the lubricating layer usually consists of a thick slurry of water and sedimentary rock. An ice stream might accelerate or stop, Blankenship says, as it moves from one type of rock to another or if conditions at its base change in some other way. "It might even be something as odd as the changing of an aquifer."

Over the past two years, Blankenship has found evidence of another mechanism that could trigger acceleration of the ice stream: volcanism below the ice. Signs of volcanism are common in Antarctica. The Ross Island area in particular is littered with cinder cones, and on a clear day at McMurdo one can see a banner of smoke trailing from the crest of Mount Erebus, an active volcano that rises 3,794 meters above Ross Island. The possibility that volcanoes might be smoldering under ice streams in the West Antarctic first occurred to Blankenship six years ago while he was flying over the West Antarctic and noticed circular depressions several kilometers across in some ice streams. "I remember writing 'volcanism' in my notebook," Blankenship says. Later he noticed similar depressions in satellite images. Unlike crevasses and other superficial features, these depressions did not move with the ice but remained fixed.

Blankenship was able to test his hypothesis a year ago with the Airborne Lithosphere and Ice-Cover Experiment (ALICE). It consists of a Twin Otter airplane outfitted with magnetometer, gravimeter, radar and laser altimeter, which together can determine the thickness of the ice and the nature of the underlying rock. With Robin E. Bell of the Lamont-Doherty Geological Observatory, the co-leader of ALICE, Blankenship focused the instruments on a large depression in the West Antarctic ice sheet. Sure enough, the sensors detected a conical structure with the unique magnetic signature of volcanic rock slightly upstream of the depression, underneath about 1,200 meters of ice. In addition to this evidence of "active volcanism," Blankenship says, he and Bell have found extensive volcanic deposits underlying ice sheets.

Even disregarding what Blankenship calls the "wild card" of volcanism, researchers may be hard-pressed to forecast Antarctica's future. One common belief is that increased precipitation induced by global warming over the past century is causing the ice sheet to grow. That claim is unsubstantiated, according to Stanley S. Jacobs of Lamont-Doherty. In a recent *Nature* article entitled "Is the Antarctic Ice Sheet Growing?," Jacobs concluded that there are not enough data for a definitive answer: "It is too early to say how the Antarctic ice sheet will behave in a warmer world."

Ellen S. Mosley-Thompson of Ohio State agrees. She notes that ice cores provide only rough estimates of the amount of precipitation in a given year, and real-time measurements of precipitation around the continent have been spotty. This season, Mosley-Thompson flew to the South Pole to set up equipment for long-term measurements of precipitation, but she says it could take many years to provide data good enough to constrain models predicting Antarctica's future. "We can't even predict the weather for Columbus, Ohio," she says, "and this is an entire continent."

In the meantime, scientists can only speculate. David R. Marchant, a geologist at the University of Edinburgh, is asked to do just that by journalists visiting him at his camp on Ross Island, where he is spending the austral summer. Marchant, his skin reddened by the wind and 24-hour sunlight, says he favors the stablist position in what has been called the stablists-versus-dynamicists debate. Yet even he acknowledges that, if the past is any guide, great changes could be in store for Antarctica.

During the peak of the last ice age 18,000 years ago, he points out, the shelf of ice covering the Ross Sea slumped to the seafloor, causing the ice streams feeding it to back up. The ice shelf burst its seams, flowing upward into surrounding valleys and burying this sooty spit where Marchant and his two students have pitched their canary-yellow tents. On the other hand, ice cores and other evidence indicate that the East Antarctic ice sheet shrank during that period, starved of precipitation.

If global warming persists—a possibility Marchant actually doubts—this scenario could happen in reverse. Within a century or two, the West Antarctic ice sheet could disintegrate, triggering a surge in sea levels. Eventually, over thousands of years, sea levels might drop again as increased snowfall caused by the warmer weather builds up the East Antarctic ice. "This is just arm waving," Marchant says. "But it's one possibility that should be out there."—*John Horgan*

## Lethal Cascade

*A model for the neurologic damage found in AIDS*

The neurologic manifestations of AIDS are a particularly terrifying aspect of the disease. As many as 60 percent of those who have AIDS suffer memory loss, motor and behavioral difficulties and cognitive impairment. In some cases, nearly half of the neurons in areas of the cortex die. Yet the mechanisms of human immunodeficiency virus (HIV) infection in the brain have remained, in large part, elusive.

Now, however, scientists say a picture is emerging from disparate findings, many of them previously considered to be at odds. "Even last summer there was a fair amount of contentiousness between this person's factor and that person's factor," comments Richard T. Johnson, chief neurologist at the Johns Hopkins University School of Medicine. "It was chaos until a couple of months ago. Now it may all fit together."

Despite the optimism, the story of how HIV damages the brain remains complicated and incomplete. Recent evidence suggests that certain strains of HIV preferentially affect the organ. But no one knows how the virus initially gets in. One theory posits that infected white blood cells serve as Trojan horses, covertly ferrying the virus across the blood-brain barrier. Another maintains the blood-brain barrier is somehow disrupted, permitting the entry of viral particles or infected cells.

Once there, the virus appears to infect and replicate in scavenger cells: brain macrophages and macrophage-like cells, called microglia. Although neurons die in the course of the disease, they do not, in general, seem to be infected—but "it is still an open question," says Janice E. Clements, a neuroscientist at Johns Hopkins.

Researchers have proposed many routes by which the infected cells could bring about neuronal death, explains Richard W. Price, a neurologist at the University of Minnesota. One possibility is that HIV-laden macrophages and microglia release cytokines and other cellular compounds that can be toxic.

A second theory suggests that products made by the virus itself cause neurotoxicity. For example, HIV and HIV-infected cells can shed a protein called gp120, which is found on the surface of the virus. This protein binds with a molecule on a form of immune system cell—*T4* cells—that allows the virus to enter them. Gp120 has also been shown to bring about the production of cytokines and to alter calcium channels. The latter change, either independently or in conjunction with the neurotransmitter glutamate, can damage neurons.

The promise of synthesis was prompted by ongoing discoveries about cytokines and astrocytes, a type of glial cell that sustains neurons. Although proliferation of astrocytes is a characteristic of AIDS, the reason for their unnatural growth was not clear. But Howard E. Gendelman, a virologist at the University of Nebraska Medical Center, and his colleagues may have an answer. They suggest that interplay between infected macrophages and astrocytes causes the macrophages to make tumor necrosis factor and interleukin-1. These cytokines, in turn, spur the astrocytes to proliferate. The researchers found that without this interaction cultured, HIV-infected macrophages were unable to do their deadly work. This fact "is important because macrophages require astrocytes," says Leon G. Epstein, a neurovirologist at the University of Rochester.

The team also discovered what appears to be a positive feedback loop. The cytokines are regulated by arachidonic acid metabolites and platelet-activating factor, which have been implicated in studies of neurologic AIDS and which may be released by macrophages and astrocytes. These compounds promote the production of more cytokines. "Astrocytes may serve as an amplifier," Epstein notes. He adds that these findings may help explain one commonly observed phenomenon of AIDS infection: very few HIV-infected cells bring about extensive damage.

The astrocyte-macrophage interaction model appears to dovetail with a general theory of neuronal impairment. In many forms of brain injury—including stroke—damage or death is brought about by glutamate. This messenger, operating through a receptor called NMDA, can cause neurons to become too excited: like overloaded fuses, they burn out. Some of the cytokines and the compounds produced by infected macrophages and by astrocytes may sensitize the neurons to the deleterious effects of glutamate. "There may be a final common pathway with many initiating factors," Gendelman says. "The NMDA receptor could be that pathway."

And gp120 has a role in this activity as well. Not only does the viral protein stimulate the synthesis of toxic cytokines, it appears that it, too, interacts with astrocytes. Dale J. Benos of the University of Alabama at Birmingham and others have found that gp120 can alter astrocytes, interfering with their normal function of glutamate uptake. The glutamate stimulates additional NMDA receptors, increasing the potential for neurotoxicity, Benos explains.

Taken together, these findings "bring up some really nice possibilities for therapeutics," Johnson notes. "If you could figure out what the cascade is, you could treat aspects of it." To this end, researchers, including Stuart A. Lipton, a neurologist at Harvard University, have been studying NMDA antagonists, compounds that prevent glutamate from binding to the receptor, and calcium channel blockers—many of which are approved for other uses. Lipton has been able to prevent cell damage in cultures, and one NMDA antagonist is already in clinical trials. "The chance of affecting this disease is much more positive than I thought it was a year ago," Epstein says. *—Marguerite Holloway*

## Noah's Freezer

When Gregory Benford heard biologists discussing the rates at which species are disappearing, he was struck by the resignation in their voices. "They were uniformly gloom and doom," he recalls. "They all believe we are going to lose a big piece of biodiversity."

So Benford, a physicist at the University of California at Irvine and a popular author of science fiction, decided some action was in order. "It occurred to me that if we think we're going to lose it, we have a moral obligation to try to save some samples," he says. Last November in the *Proceedings of the National Academy of Sciences,* Benford proposed doing just that—not only in zoos, gardens and refuges but also flash-frozen in liquid nitrogen.

"Admitting your ignorance about the number and dispersal of species, you should just sample randomly in a threatened region," Benford says. Small plants and insects might be frozen whole; other animals and trees might be represented by embryos or tissue samples. "A plausible estimate is that you could get representatives of a few percent of the total species."

A few percent may not sound like much, which is why Benford calls the effort "an emergency salvaging operation" rather than a species inventory. Labels on the samples would state only their place of origin. No effort would be made to identify or describe the specimens. "The main thing," he says, "is to get the data and to process them as little as possible."

Indeed, any attempt to identify the species en route to the freezer would probably make the project impossible. Benford argues that there are not enough taxonomists to catalogue a broad sample from the endangered regions. Yet nonspecialists could easily be trained to collect and freeze specimens. The project could therefore be conducted fairly inexpensively with local labor anywhere in the world. Liquid nitrogen would be the refrigerant of choice because it offers the best combination of low cost (about 25 cents per liter) and high reliability.

By Benford's estimate, it would cost less than $2 billion for biologists to collect samples from all the tropical rain forests and store them for a century. "You can do something on the cheap," he says. "It's not like the Superconducting Super Collider." Independent groups could each tackle a habitat. Benford has suggested, for example, that the Sierra Club might consider sampling and freezing the species from the endangered redwood forests.

But what will anyone ever do with the frozen compendium? "That's dependent on future technology," Benford says. Some simple organisms might still be alive after they were defrosted, but most would not. The DNA inside the cells, however, would be largely intact. Benford believes the genetic information in the DNA could be analyzed for its secrets. The DNA might even be inserted into living cells to re-create an extinct species, à la *Jurassic Park.*

According to Benford, one reviewer of his paper had asked why he did not propose storing just the DNA. "The reason is that it's more expensive to pull the DNA out of a beetle than it is to put the beetle in a bag," Benford explains. "And you get much more information out of a whole beetle."

At the urging of the National Science Foundation, Benford says, he plans to organize a small conference later this year to discuss the idea. Critics will undoubtedly find weaknesses in it. The sampling of any habitat will almost certainly be biased in some way: soil microorganisms, for example, may be sampled less thoroughly than larger animals or plants. It is hard to guess how much information about an entire species could be deduced from a single frozen individual. But even if Benford's freezing plan is imperfect, the question remains: What are the alternatives? *—John Rennie*

## Young Suns

*Telescope technology pulls the veil from infant stars*

A meeting of the American Astronomical Society may seem an unlikely place to be confronted with a proud couple's baby pictures. But that did not discourage Stephen E. and Karen M. Strom of the University of Massachusetts at Amherst. And the interest expressed was more than polite, perhaps because the pictures portrayed newborn stars, still swaddled in thick clouds of gas and dust.

In collaboration with K. Michael Merrill of Kitt Peak National Observatory, the Stroms have produced images that reveal aspects of star formation never before seen. The observations also provide information about the disks of matter that seem to surround young stars. Such disks, which are thought to form the raw material of planetary systems such as the solar system, are the subject of provocative new observations by the *Hubble Space Telescope.*

One of the primary obstacles to watching stars being born is that the births take place deep within dense nebulae. Enshrouding dust scatters visible light, obscuring the earliest stages of stellar formation. Infrared rays have a longer wavelength than does light and so are able to penetrate the thick clouds and to supply information about what is happening within. Only for the past few years, however, have detectors existed that can generate high-resolution images of the infrared sky.

Using these detectors, which Stephen Strom says "have revolutionized protostellar astronomy," Merrill and the Stroms inspected Lynds 1641, an interstellar cloud lying on the outskirts of the Orion Nebula. (The Orion Nebula is visible to the naked eye as a fuzzy "star" in Orion's sword.) They captured unprecedentedly clear views of star-forming regions that include some of the youngest stars ever seen—about 500,000 years old; the sun is 4.6 billion years old, for comparison.

As the researchers peered deep into the stellar nursery, they observed stars clustered into eight small gatherings, each about one light-year wide and containing anywhere from a few dozen to 150 members. Previously, astronomers had observed stars forming either alone or in vast congregations, which made it impossible to see the details. The new results "show that the most common path of stellar formation may be in small groups," Stephen Strom says. The huge, spectacular associations of youthful stars seen elsewhere in the Orion Nebula probably consist of many small, overlapping stellar aggregates of slightly varying ages.

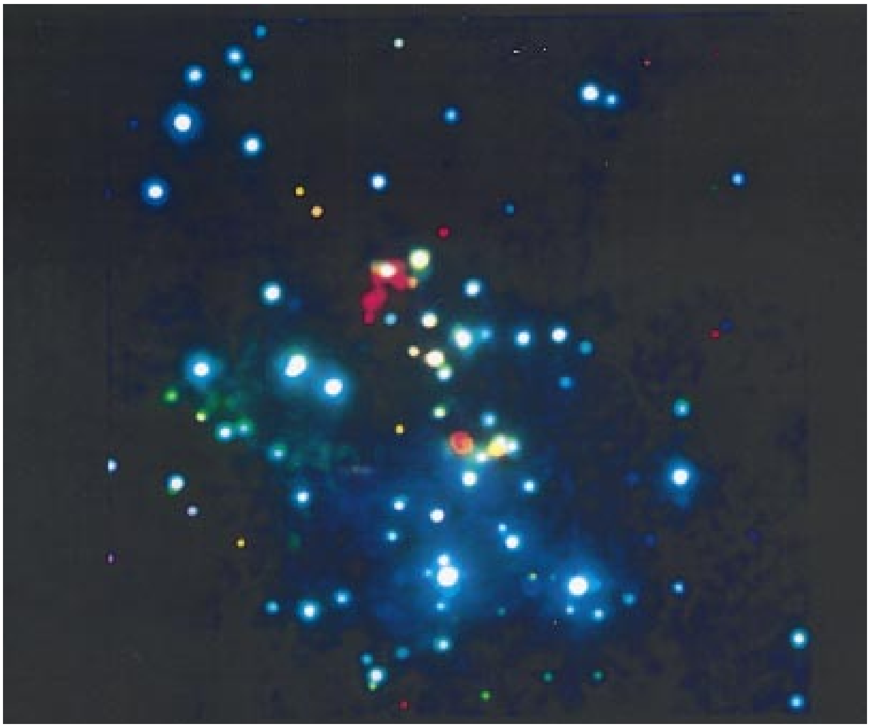
Although they usually begin their lives in groups, most stars end up traveling alone through the Milky Way. The infrared portraits of Lynds 1641 offer a telling view of how stars come to live the single life. The youngest, most thoroughly enshrouded stars lie in the dense centers of the star-forming regions. Gravitational interactions with their neighbors eventually expel stars from the aggregation and send them along their solitary paths. Images produced by Merrill and the Stroms clearly show a population of older, less heavily obscured stars that appear to have made a recent exodus from the aggregation.

The stars in Lynds 1641 reveal other changes as they age. Nearly all of the youngest stars emit more infrared radiation than one would expect from a star alone. That observation, in conjunction with other evidence, suggests disks of dust surrounding the stars absorb light and reemit it as infrared rays. "Disks are a natural part of the star formation process," Stephen Strom says. "Most of the material that makes up the star passes through a disk."

Disks play an important role in early stages of stellar evolution. Conservation of angular momentum implies that stars should rotate far faster than they actually do. Theoretical models indicate that disks exert a drag that slows down the star and transfers angular momentum outward into the disk. Those models help to explain the odd fact that 99.5 percent of the angular momentum in the solar system resides in the planets, not in the far more massive sun.

Joanne M. Attridge and William Herbst of Wesleyan University, who have measured the rotation periods of 40 stars ranging from one million to 10 million years old in the Orion Nebula, find that diskless stars tend to rotate four times as fast as their disk-endowed brethren. "Even naked stars must have gone through a disk phase," Herbst says, but they evidently shed their disks early on. The older, outlying stars in Lynds 1641 exhibit less infrared excess than do their inner neighbors, offering a tidy example of how disks dwindle as stars mature. By the time stars are about 10 million years old, nearly all spectral signatures of disks disappear.

The fate of the material in the disks has long fascinated astronomers. In our solar system at least, matter in the disk gathered together into planets. The Stroms cite "compelling evidence" that a similar process is occurring around many fledgling suns. Thin disks of dust



NATIONAL OPTICAL ASTRONOMY OBSERVATORIES

*NEWBORN STARS are revealed in this false-color infrared image. Red denotes the most heavily obscured objects, which are visible only at the longest wavelength observed.*

grains have been observed around a handful of fairly mature stars, including the bright stars Vega and Fomalhaut. Such grains "have no business surviving," Stephen Strom notes, because they should quickly spiral into their central star. The persistence of dusty disks implies the presence of parent bodies, possibly comets or asteroids, which collide with one another to produce the dust. Much to their regret, astronomers cannot yet tell whether larger, planet-size objects orbit the stars as well.

The *Hubble Space Telescope* is contributing additional information about circumstellar disks. C. Robert O'Dell of Rice University, working with two graduate students, and J. Jeffrey Hester of Arizona State University have studied stars in the bright inner regions of the Orion Nebula. There the researchers perceive stars and disks silhouetted against brilliant radiation from the hottest stars in the nebula. The radiation strips material out of the disks and blows it into a tail. Based on the rate at which the disks evaporate, O'Dell calculates that they contain about 15 times the mass of Jupiter—a healthy amount of material from which to make a planetary system.

The normally mild Stephen Strom lights up at the mention of the images. "I really wanted to see those disks," he says, slightly wistful that he was unable to participate in the discovery. O'Dell describes the disks as "a missing link in our understanding of how planets like those in the solar system form."

Practically every new finding adds another indication that planetary systems are a common consequence of the way stars are born. Baby stars, it would seem, are naturally inclined to start families of their own. —*Corey S. Powell*

## Quark Quest

*Have all six flavors finally been observed?*

It is easier to find beauty than truth. Beauty, also known as the bottom quark, was discovered almost two decades ago, but truth, the top quark, cannot be found anywhere in the cosmos, except perhaps at Fermi National Accelerator Laboratory in Batavia, Ill.

In recent months, physicists there have recorded two "interesting events" that might be the signature of the top quark, says Melvyn Shochet, spokesperson for the Collider Detector at Fermilab. "To claim discovery of top, you probably need something like five to 20 events in each of several decay modes," he adds.

Top remains the only one of the six quarks whose existence has not been confirmed. Most matter—that is, protons and neutrons—is made of quarks known as up and down. Other flavors—strange, charm and bottom—can be produced only in particle accelerators and perhaps in dense, massive stars. Top, if it exists, has probably not made an appearance since the hot, explosive birth of the universe.

"Some might say that the discovery of the top quark would be somewhat of an anticlimax because we have very, very strong reasons to believe it exists, and we know its mass within a certain range," remarks Nobel laureate Steven Weinberg of the University of Texas at Austin. "It may not sound like it is very important to know whether we know the mass of the top quark precisely. In fact, it is enormously important."

The top quark should provide an essential clue as to why all particles have the masses they do. In particular, physicists are puzzled about why every fundamental particle has two siblings that are the same in every way except for their mass. For example, the bottom quark responds to weak, strong and electromagnetic forces in very much the same way as the strange and down quarks, yet bottom is 25 times more massive than strange and 700 times heavier than down.

To explain why some particles have more mass than others, physicists have devised several theories, the simplest of which is the Higgs mechanism. Just as the electric charge of a particle says something about how strongly it interacts with electromagnetic fields, the mass of a particle is related to how strongly it couples to the so-called Higgs field, according to theory. Such a field would manifest itself in experiments as a new type of particle, the Higgs boson. "An accurate measurement of the mass of the top quark would give an important clue to the questions: Is there a Higgs particle? What is its mass? What kind of experiment do you have to do to find it?" Weinberg explains.

So why has it proved so difficult to create a top quark when conjuring up bottom quarks is a cinch these days? The first reason is that it is heavy. Experiments at Fermilab show that the top quark is at least as massive as a silver atom and more than 20,000 times heavier than an up quark. Fermilab scientists produce such massive particles by smashing protons together with their antimatter counterparts. The collision releases 1.8 trillion electron volts of energy, which may or may not turn out to be enough energy to generate top quarks.

The second reason the top quark eludes detection is that it is extremely unstable. No one expects a top quark to stick around for more than a millionth of a billionth of a billionth of a second. It disintegrates into a menagerie of secondary particles, which can then be detected.

To find one interesting event, Fermilab researchers and their computer system must sort through billions of events, some of which involve hundreds of particles. Then, if they find one that seems to look like the decay of a top quark, they must prove that it was not produced by one of a dozen processes that can mimic the top quark's signature. "When you see only one event," Shochet says, "there is no way to determine whether it is the top quark or not."

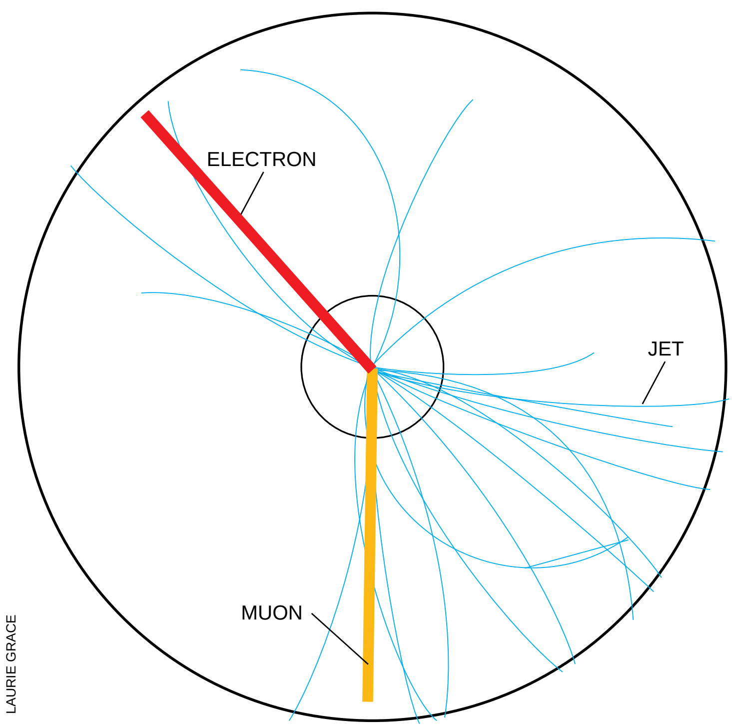
The Collider Detector at Fermilab recently recorded one possible top-quark event, and D-Zero, the newer of the two mammoth instruments at Fermilab, has observed a second. Each of the events consists of a shower of particles that could be the consequence of the disintegration of a top quark and its antimatter counterpart, an antitop quark.

Each of these two particles decays into a bottom quark and a particle called *W*, which is better known for its role in conveying the weak force. The bottom quark then disintegrates, producing jets of more mundane particles. The *W* particle decays into either an electron or its sibling, a muon. So what was actually measured at Fermilab was an energetic electron, a mercurial muon and several jets. Unfortunately, it is also possible that these particles were generated by the decay of something other than a top and antitop pair.

The Fermilab observations include another bit of evidence that supports the top-quark hypothesis. The bottom quark produced by the decay of the top can join up with another quark to form a stable particle. That particle zipping along at close to the speed of light will travel as far as a few millimeters before it breaks apart into jets of lighter particles. To identify this signature, Fermilab scientists have added a "vertex detector" to the Collider Detector at Fermilab; the apparatus distinguishes those particles that decay in the center of the detector from those that break up a short distance away. The vertex detector should enable researchers to identify bottom quarks unambiguously and thereby make it easier to recognize top-quark events.

One of the two events touted as a top-quark decay seems to show jets forming away from the center of the detector, which suggests the decay of a bottom quark. Yet Fermilab investigators have not had enough experience with the new vertex detector to be certain of their measurements. "What we really need is lots of collisions to look at so that we produce enough of these objects to see signal above background," Shochet comments.

Fermilab scientists have good reason to be cautious about their findings. In 1985 investigators at CERN, the European laboratory for particle physics near Geneva, claimed discovery of the top quark, only to be proved wrong later. Some three years ago Fermilab workers recorded a top-quark candidate, but the evidence was inconclusive. Yet maybe, just maybe, physicists will soon know the truth. —*Russell Ruthen*




***TOP QUARK?** Physicists at Fermilab detected, in 1989, an event consisting of an electron, a muon and many jets of particles. The event could represent the decay of a top quark and its antimatter partner. Fermilab recently recorded two similar signatures.*

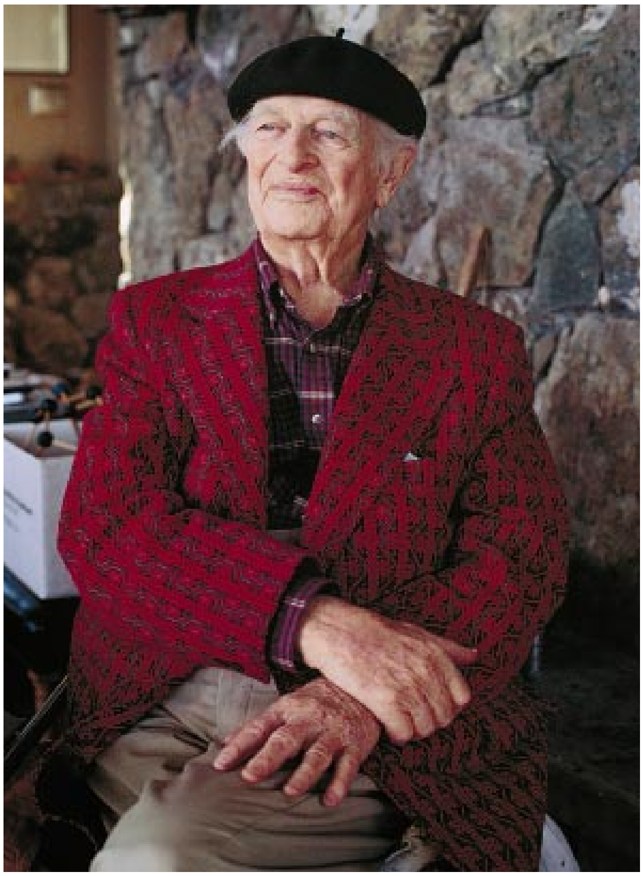
# Stubbornly Ahead of His Time

Linus C. Pauling does not look like a juggernaut. With his crinkly blue eyes and ruddy cheeks, he could easily play the role of wise, kindly grandfather—and in fact Pauling, who turned 92 in February, has 15 grandchildren and an equal number of great-grandchildren. His black beret, pulled down over a fringe of snowy hair, adds a jaunty, continental touch.

It is only when he speaks that Pauling reveals the implacable intensity that has characterized his extraordinarily long and productive career. Whether elucidating his theory of chemical bonding or extolling the benefits of vitamin C, Pauling marshals names, dates and other facts with the fierce precision of a trial lawyer.

He exhibits a bracing self-regard. Luck, he remarks, rarely played a role in his scientific discoveries. “My success as a scientist has been largely the result of having broader knowledge than most scientists, in particular having a remarkably extensive knowledge of empirical chemistry, and also knowing mathematics and physics.” When I express the hope that we can touch on all the important aspects of his career during our interview, he looks at me skeptically and replies, “How many days have you got?”

Fair answer. Pauling not only helped to lay the foundation of modern chemistry, biochemistry and molecular biology, he also erected much of the edifice. A supreme theorist and experimentalist, he recast chemistry in the mold of quantum mechanics and pioneered techniques such as x-ray and electron diffraction for deciphering the structure of molecules. Pauling has won many honors, including the Nobel Prize for Chemistry. The British journal *New Scientist* has called him one of the 20 greatest scientists of all time, on a par with Newton, Darwin and Einstein.



JASON GOLTZ

***LINUS PAULING is the only scientist to have won two unshared Nobel Prizes—for chemistry in 1954 and for peace in 1962.***

Yet this quintessential scientific authority is best known today as a maverick. His protests against the U.S. development of nuclear weapons during the chilliest years of the cold war led him to be assailed as a communist sympathizer. For almost three decades, moreover, Pauling has been battling the biomedical establishment over his claims about vitamin C and other nutrients.

This struggle continues. A number of recent studies have shown that a high intake of vitamins—and vitamin C in particular—is indeed associated with lower susceptibility to disease and longer life expectancy. The findings triggered a flurry of attention for Pauling, including an adulatory profile in *People* magazine last fall. Skeptics pointed out that the studies do not demonstrate a cause-and-effect relation between vitamins and resistance to disease, nor do they prove the value of the doses advocated by Pauling, which are hundreds of times higher than the recommended daily allowances established by the Food and Drug Administration.

On a more personal level, Pauling has had to endure cutbacks at the nonprofit institute he founded in 1973 to investigate his vitamin theories. The Linus Pauling Institute of Science and Medicine in Palo Alto, Calif., has been in financial straits for years. Pauling also discovered a year ago that he has cancer of the prostate gland. He insists, of course, that the cancer was “put off for 20, 25 years because of my high intake of vitamin C and other vitamins.” (Pauling takes 18 grams of vitamin C a day, 300 times the FDA’s recommended daily allowance.) If he does not achieve his goal of living to be 100, he says, the reason will be that he started taking megadoses only 27 years ago.

Asked if it bothers him that he still has to fight so hard for recognition and respect, Pauling shrugs. “I’m accustomed to having my ideas received with skepticism,” he replies. The problem, he suggests, recalling a remark by the eminent biologist René Dubos, may be that he is always 20 years ahead of his time. Pauling then offers another quote, which is as close to self-criticism as he comes: his wife, Ava, who died in 1981, used to observe that “I am just too stubborn to change my mind about anything under pressure.”

Pauling’s appetite for scientific minutiae—and his enormous self-assurance—was manifest early. Growing up in Oregon, he devoured books on mineralogy, chemistry and physics. “I mulled over the properties of materials: why are some substances colored and others not, why are some minerals or inorganic compounds hard and others soft,” he says. “I was building up this tremendous background of empirical knowledge and at the same time asking a great number of questions."

After graduating from Oregon Agricultural College (now Oregon State University), Pauling entered the California Institute of Technology. In three years he had gained a doctorate in physical chemistry and "a feeling of confidence in my own thinking." Heading to Europe in 1926, he immediately plunged into quantum mechanics, which was still in its infancy. "In 1926 I published the first paper that applied quantum mechanics to systems with more than one electron," he says. By the late 1920s, he contends, he was "the only person in the world who had a good understanding of quantum mechanics and an extensive knowledge about chemistry."

After returning in 1927 to Caltech, where he remained until the 1960s, Pauling devised a quantum theory of chemical bonding, the phenomenon whereby atoms and molecules become affixed to one another by sharing electrons. One of his key concepts was resonance, in which a molecule fluctuates between two different states and gives rise to a new, intermediate state.

Feeling by the end of 1930 that inorganic chemistry "was pretty well taken care of," Pauling focused on the skeinlike molecules from which living things are knit. His investigations of the blood protein hemoglobin led to a theory of native and denatured proteins, which explains, for example, how egg white gels when cooked.

In 1939 Pauling poured his knowledge into one of the most influential science texts ever written, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals.* His theories were not universally accepted. Some Soviet scientists proclaimed resonance to be incompatible with dialectical materialism, while some Western chemists complained that the theory was based on molecular structures whose existence remained unproved.

Yet scientists ignored Pauling's work at their peril. This moral emerges from Pauling's recollection of how he discovered the helical shape of proteins. It was 1948, and Pauling was puzzling over the three-dimensional structure of a common protein called keratin. Bedridden with a cold, he sketched the molecule on a piece of paper and began bending and twisting the paper, trying to find a structure that might reproduce published x-ray diffraction data. Pauling finally came up with a helical model that would account for most of, but not all, the data. He decided not to publish his results until he could resolve the discrepancy.

Meanwhile a group led by the distinguished physicist W. Lawrence Bragg had published a paper proposing a different—and incorrect—helical structure for proteins. The physicists had ignored findings on polypeptides that Pauling had published years earlier. "They hadn't read my book!" Pauling exclaims, still astonished after all these years. Pauling published his correct version of the protein helix two years later.

Pauling declined to work on the Manhattan Project during World War II. But temporarily setting aside his pacifist inclinations ("Hitler had to be stopped"), he supervised the development of other military technology, including armor-piercing shells and a new class of explosives. He points out that in 1948 President Harry S. Truman awarded him the Presidential Medal for Merit for his wartime service.

*At night, Pauling spends several hours scouring scientific journals for "things I don't understand."*

After the war, however, Pauling decided that "averting a nuclear catastrophe is so important that I'd better do my part." He began speaking out against nuclear weapons and arguing that they had made war obsolete. His severest critic early on was his wife. Dissatisfied with one of his lectures, she warned him that if he could not address the issue of peace with the same authority that he displayed on scientific subjects, he should not even try. Pauling subsequently immersed himself in studies of international affairs.

In the 1950s, during communist witchhunts by Senator Joseph McCarthy, Pauling was harshly attacked for his views, and the U.S. State Department revoked his passport. Only at the last minute did the government allow Pauling to travel to Sweden to receive the 1954 Nobel Prize. Pauling was hardly cowed. He wrote a book called *No More War!* that was published in 1958, and that same year he organized a petition of scientists opposed to nuclear testing. He won the Nobel Peace Prize in 1962 and much of the credit for a ban on atmospheric nuclear tests signed the following year by the U.S. and the U.S.S.R.

Pauling remains distrustful of authority. For that reason, he does not advocate the concept of world government, as many pacifists do. "If we had a world government, Hitler reincarnated might gain control over it," he explains. "And in any case, the power elite would no doubt strive to get control just as they have control over the United States."

In the 1960s Pauling transformed himself yet again—into a prophet for vitamins. His belief in the value of consuming large quantities was based on earlier work he had done on optimal doses of drugs. Whereas most drugs become toxic at high doses, "I realized that vitamins are essentially nontoxic even in very large amounts. Perhaps one or two people over a period of decades have died from an overdose of vitamins."

Pauling's studies convinced him that the optimal dosage of vitamins was much higher than the intake from a normal diet. He has emphasized the ability of vitamin C to ward off specific maladies, including the common cold, cancer and, most recently, heart disease. But he maintains that vitamin C provides protection "from essentially all diseases."

Pauling recalls that a nutritionist who reviewed his 1971 book *Vitamin C and the Common Cold* complained that he had never had a course in nutrition and "probably would flunk the course we give to our first-year students." Although such skepticism persists in the biomedical community, Pauling is confident his views will eventually be validated. "Of course," he adds, "I would say they were validated long ago."

Somehow Pauling finds the time for pure science. He does his best work in Big Sur, Calif., where he owns a 160-acre ranch on a wild stretch of coast overlooking the ocean. On a typical day there, Pauling rises before dawn and works through the afternoon writing papers and letters and making calculations. After watching the evening news on television, he spends several hours reading science journals "looking for things I don't understand."

A mystery that caught Pauling's attention almost a decade ago is an odd form of matter called quasicrystals. Unlike most physicists, Pauling concluded that the unusual fivefold symmetry of the materials derives from a conventional crystallographic phenomenon known as twinning. The broad acceptance of more exotic explanations for quasicrystals, he suggests, is symptomatic of a more general decline in crystallography. "The young crystallographer doesn't think. He puts the crystal into the automatic diffractometer, which is coupled to the computer, which then works out the structure—and maybe couples to a computer system that writes a paper, I don't know!"

In spite of—or because of—Pauling's complaints, "the quasicrystal people don't take what I say really seriously." He smiles as he speaks. History, he seems sure, is on his side.—*John Horgan*

SCIENTIFIC AMERICAN

# Why America's Bridges Are Crumbling

*Inadequate maintenance has piled up a repair bill that will take decades to pay off. Indeed, the scope of the problem is only now becoming clear*

by Kenneth F. Dunker and Basile G. Rabbat

America is in desperate need of new bridge work. Of the roughly half a million highway bridges in the U.S., more than 200,000 are deficient. Some are merely obsolete—built in a time of smaller vehicles and narrower roads—but the rest, for one reason or another, are incapable of sustaining the loads that current design standards demand. More than 130,000 bridges carry markings that restrict the weight of trucks passing over them, and about 5,000 have been closed. Every year, on average, between 150 and 200 spans suffer partial or complete collapse. Sometimes the collapse creates a spectacular disaster, such as the 1983 fall of the Mianus River Bridge on Interstate 95 in Connecticut.

KENNETH F. DUNKER and BASILE G. RABBAT have collaborated on analyses of the National Bridge Inventory since 1988. Dunker is an associate professor of civil and construction engineering at Iowa State University, where he studies methods for strengthening structurally deficient bridges. Rabbat manages structural codes for the Portland Cement Association. He received his B.S. in civil engineering from Alexandria University in Egypt and his doctorate from the University of Toronto. Rabbat also serves on the American Concrete Institute committee that develops the building code for reinforced concrete.

Current estimates of the cost for remedying all deficient bridges start at about $90 billion. The problem is a result of more than half a century of construction and subsequent inadequate maintenance. Indeed, the past 25 years are replete with federal programs intended to repair decaying bridges.

If the problem is to be solved in a definitive manner and so eliminate the need for major emergency outlays in the future, then its roots must be clearly understood. Among the questions: What kinds of bridges are most likely to be deficient? How do bridges fall into disrepair? How much danger do decaying bridges pose? To answer such questions, we analyzed information from the National Bridge Inventory, a data base maintained by the Federal Highway Administration (FHWA). Our work shows that some common perceptions about the issue are wrong.

Large urban bridges are commonly perceived as being in the worst condition. Deficiencies are most common, however, among short spans that should be simple to maintain in good repair. Similarly, bridge work on major highways is highly visible and gives the impression that the problem is one of heavily traveled routes, yet a vast number of deficient bridges lie concealed along lightly traveled back roads.

The state of disrepair into which the nation's bridges has sunk surprises no one in government or the highway construction industry. Ever since the Silver Bridge across the Ohio River collapsed in December 1967, killing 46 people, states have kept extensive records of bridge safety and adequacy. The Silver Bridge disaster happened in part because of poor inspection by local authorities; consequently, the Federal-Aid Highway Act of 1968 mandated both national bridge inspection standards and training for bridge inspectors. Today most U.S. highway bridges are inspected every two years.

Each state forwards the results of its inspection to the FHWA for inclusion in the National Bridge Inventory. The federal government relies on these data to determine the scope of national highway bridge needs and to administer federal funding programs. Eligibility of a bridge for such funding at present requires that it be classified as "structurally deficient" (unable to carry standard loads) or "functionally obsolete" (too narrow or lacking sufficient clearance). About 39 percent of the entire stock of highway bridges in the U.S. is classified as deficient according to one of these two definitions. Furthermore, roughly

**MIANUS RIVER BRIDGE collapse killed three people and disrupted traffic on Interstate 95 in Connecticut for months. The failure was traced to a combination of risky design and poor maintenance.**

## How Bridges Fail

In physical terms, bridges deteriorate because of weather and traffic. Water corrodes steel and can scour away bridge foundations. Meanwhile every car and truck that passes over a bridge causes it to flex. Excessive loads can cause cracks that destroy the structure's integrity.

Although engineers agree on the general mechanisms of bridge failure, the details are not well understood. Bridge structures are too complex for complete computer analysis, and so simulations require a host of simplifying assumptions. Consequently, the results seldom match behavior in the field.

Once a bridge has begun to deteriorate, the process of decay accelerates. The portions of metal beams that are under the most stress corrode more rapidly, and stress concentrations increase as the thickness of sound metal decreases. Similarly, damaged structural members have reduced load-bearing capacity and are thus more vulnerable to the effects of heavy traffic.

Simple problems, if allowed to progress unchecked, can lead to severe damage. Debris on a bridge deck, for example, may block drains, causing water to accumulate. In cold regions, water may freeze inside the deck, cracking it. Deicing compounds, however, form salt solutions that rapidly corrode reinforcing bars and other structural members. When the salt content reaches a critical level, the concrete must be replaced even though the cement and aggregate are still sound.

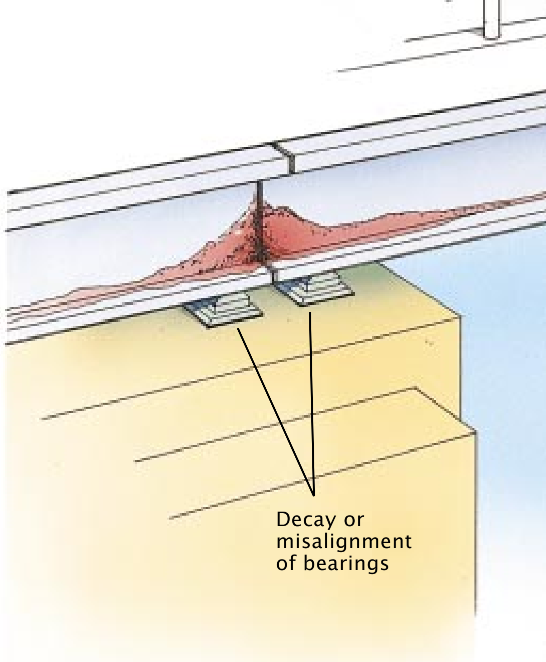
Draining the water poses its own problems. Engineers have learned at great expense to direct water away from structural elements and bearing surfaces, where the combination of salt and stress can destabilize a bridge in a few years. Yet complex drainage systems are expensive and require periodic maintenance to keep them from clogging.

Bridge substructures face problems similar to those of the superstructure and deck. In addition, they are often much more difficult to inspect and to repair. The Schoharie Creek Bridge in upstate New York, for example, collapsed in 1987 because flowing water had scoured away its foundation (*see photographs below*). Since then, states have begun employing scuba divers to inspect waterways—often in murky water where they must work entirely by feel. If a diver finds damage, workers must erect a cofferdam to make repairs.

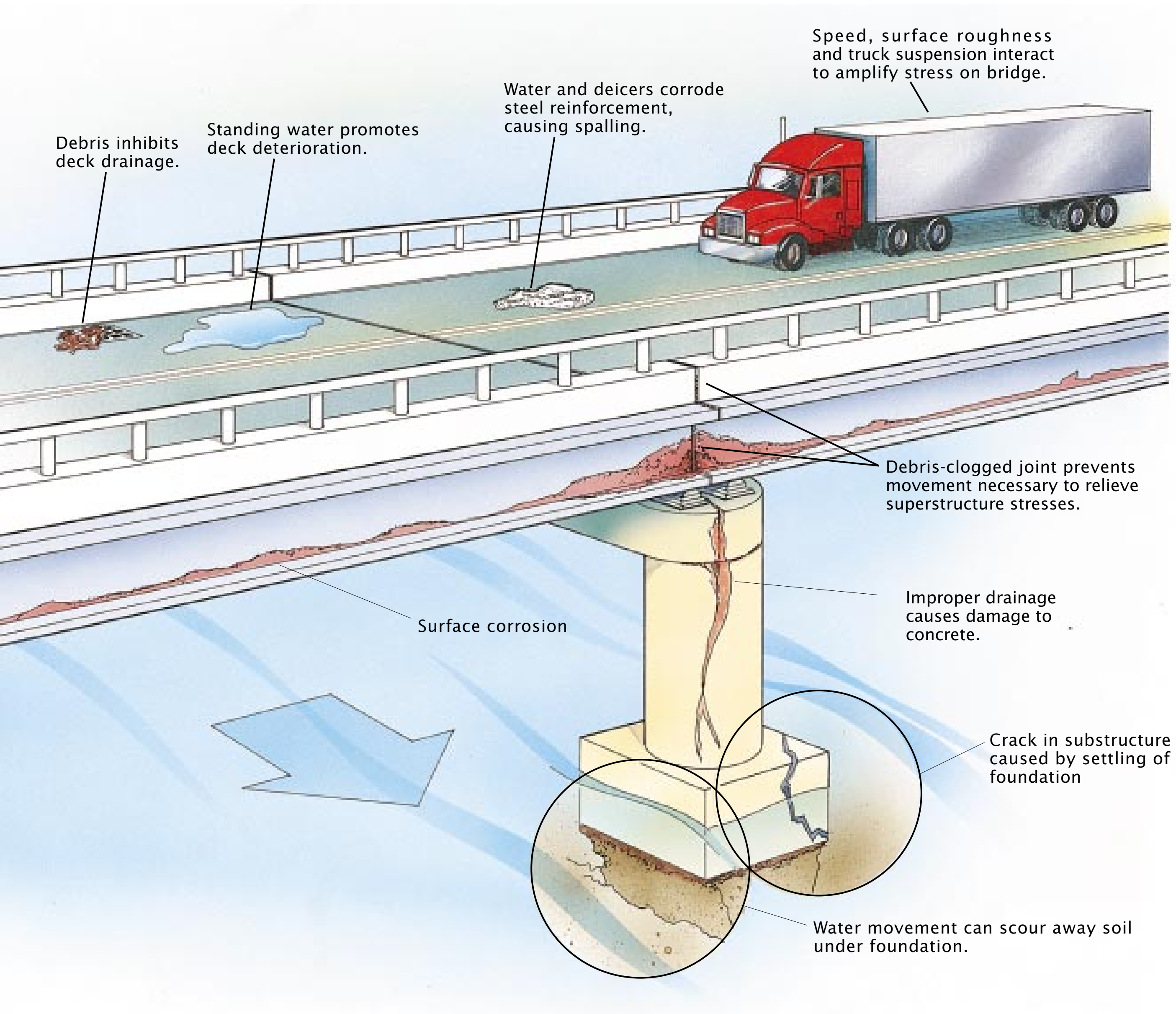
One failure mechanism over which engineers have little control is accident. According to Issam E. Harik and his colleagues at the University of Kentucky, collisions caused 42 of 79 nationally reported bridge failures between 1951 and 1988. Almost half of the collisions involved ships that rammed bridge supports; the Huey P. Long Bridge near New Orleans suffered two strikes in five years, as did the Sunshine Skyway Bridge near Tampa.

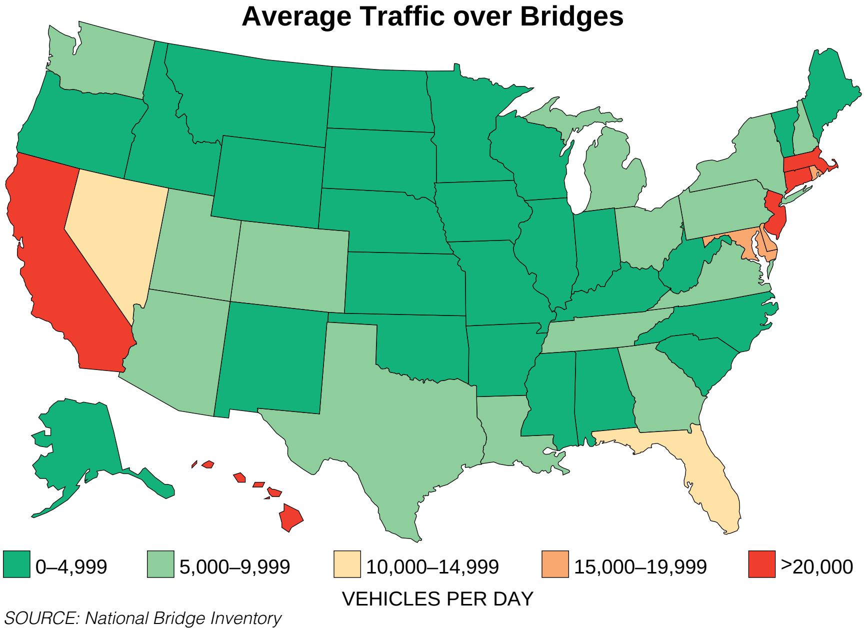
Finally, there are the collapses that defy statistical analysis—such as the failure of the Buckman Bridge near Jacksonville, Fla., in the spring of 1970. Engineers filled the bridge's hollow pilings with river water before sealing them. Anaerobic bacteria feasted on the cardboard lining of the pilings and generated enough methane to rupture the pilings and bring down part of the bridge.

Bridge superstructure is susceptible to corrosion, water damage, metal fatigue and stress caused by vibration.




**The Collapse of the Schoharie Creek Bridge, April 1987**


Speed, surface roughness and truck suspension interact to amplify stress on bridge.
Water and deicers corrode steel reinforcement, causing spalling.
Debris inhibits deck drainage.
Standing water promotes deck deterioration.
Debris-clogged joint prevents movement necessary to relieve superstructure stresses.
Improper drainage causes damage to concrete.
Surface corrosion
Crack in substructure caused by settling of foundation
Water movement can scour away soil under foundation.




**STRESS ON BRIDGES caused by traffic can be estimated from the number of vehicles traveling over them each day. These loads are one factor in bridge deterioration; weather and maintenance may also determine whether a bridge becomes unsound.**

25 percent of all bridges require posting for reduced loads. We have made a more thorough analysis of a subset of these bridges: those that span more than 20 feet (and are not classified as culverts) and were built between 1950 and 1989. Because our sample excludes older bridges, the deficiency percentages are not quite as high as for the entire population, but we believe our conclusions can still be generalized.

Our work shows a surprising pattern—or rather lack thereof—in the National Bridge Inventory. The conditions of bridges vary widely from state to state: less than 5 percent of bridges in Hawaii, California, Nevada, Arizona and Florida need repair or replacement, whereas more than 40 percent of those in Mississippi and New York State are structurally deficient. (Some of these variations may be related to differing inspection practices, but the import of the numbers is clear.)

Most of the explanations advanced for these differences do not hold up in light of the evidence. Many motorists, for example, associate auto-body rust with bridge corrosion. They reason that the same deicing compounds that eat away automobiles will also destroy steel bridge members and disintegrate reinforcing bars in concrete beams and decking. For the most part, however, auto rust is uncorrelated with the proportion of deficient bridges. Some northern states have high percentages of deficient bridges, but others do not.

Another school of thought ascribes structural deterioration to heavy truck traffic, apparently with good reason. Trucks place more than 10 times the load of an automobile on a bridge, and irregularities in the road surface can cause truck cargoes to bounce, amplifying the stress even further. Moreover, overweight trucks are a leading cause of bridge collapses. Nevertheless, the National Bridge Inventory shows an inverse correlation between average daily traffic and unsound bridges. Spans on the roads less traveled are more likely to be unsound. Thus, the largest block of problem states is not in the “rust belt” of the North, nor in the high-traffic regions of the Northeast and Southwest, but rather in the southeastern U.S. [*see illustrations above and on opposite page*].

Some analysts have suggested that structural deterioration is simply a matter of age and that the states that built many bridges early in the postwar period should have the highest proportion of deficient ones. In fact, however, there is no relation between the average age of a state’s bridges and its percentage of deficient spans. There is some correlation between the total miles of bridges in a state and the percentage of deficient spans, which suggests that the issue is not so much one of age but of maintenance.

Along with the issue of maintenance go correlations between bridge materials and deterioration—not surprisingly, timber bridges are generally in the worst condition. Only a handful of interstate highway bridges have been built from timber since the late 1950s, but of those that remain more than half are in need of repair or replacement. Furthermore, wood continues in use for state and county roads. About one quarter of the timber bridges built between 1985 and 1989 are deficient, as are the majority of those built before 1975.

The situation for steel and concrete is more complex. The percentage of defective concrete spans is relatively low and constant from state to state. The condition of steel bridges, in contrast, varies significantly from state to state; in some jurisdictions, a steel bridge is no more likely to be deficient than a concrete one, but in others the ratio is greater than three to one.

This difference is attributable almost entirely to the poor condition of bridges on state and county roads. Most bridges on interstate and U.S. highways are eligible for federal aid and thus are subject to a consistent set of design, inspection and maintenance policies. Local spans, however, depend on local monies for maintenance, and standards are far from uniform.

The National Bridge Inventory reveals that short-span steel bridges on little-used roads (averaging less than 1,000 vehicles per day, or about one every 90 seconds) are more likely to be deficient than those on more heavily traveled routes. Because steel bridges, like timber ones, tend to deteriorate unless they receive regular maintenance, this statistic implies that local authorities are handicapped by insufficient funds. They therefore allocate what money they do have to address problems that affect the largest number of drivers.

Indeed, more than 800 new bridges (built between 1985 and 1989) on state, county and city roads are classified as structurally deficient. This number accounts for between 5 and 15 percent of the bridges built in the states in question, most of them in the Southeast. It is apparent that some jurisdictions have responded to funding shortfalls by taking the extreme measure of designing and building substandard spans.

What are the effects of such ill treatment of the nation’s infrastructure? Although fewer than a dozen people typically die every year in bridge collapses, another 1,000 are killed in accidents involving bridges that are deficient, obsolete or have inadequate traffic-control provisions. Bridge closings divert drivers and disrupt traffic—among the most famous recent closings was that of the Williamsburg Bridge, between Manhattan and Brooklyn. From April until August 1988, more than 100,000 drivers a day had to find alternative routes. Even closings of small bridges can cause major dislocations, especially if they serve as the only route into or out of a region.

Even if a bridge is merely posted for reduced loads, the consequences can be significant. Transportation planners must compute additional costs for detour of trucks hauling freight and, for very low allowable loads, the costs for detour of school buses and fire trucks. In a few cases, there is no detour: fuel oil, school bus service and fire service are unavailable to areas accessible only by the posted bridge.

Restricting traffic over a bridge to less than the 40 tons usually allowed is intended to protect it from additional damage resulting from structural overloads—as well as to guard drivers from the disastrous consequences of having a bridge collapse under them. The standard signs employed for this purpose show three silhouettes: one of a straight truck, one of a semitrailer and one of a double trailer. The three different load limits take into account the fact that longer trucks with more axles reduce the load on any single part of the bridge—indeed, many long trailers are longer than a significant percentage of highway bridges.

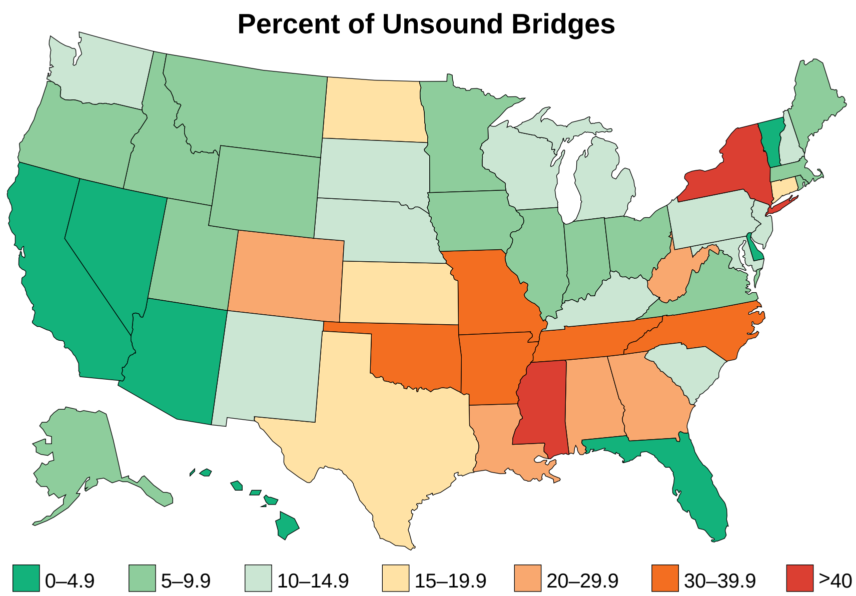
Bridge posting can also have perverse effects. An Iowa State University graduate student preparing a posted bridge for strengthening research asked a local trucker for his typical response to a load restriction sign. The trucker replied that he increased speed and drove down the center stripe on the bridge.

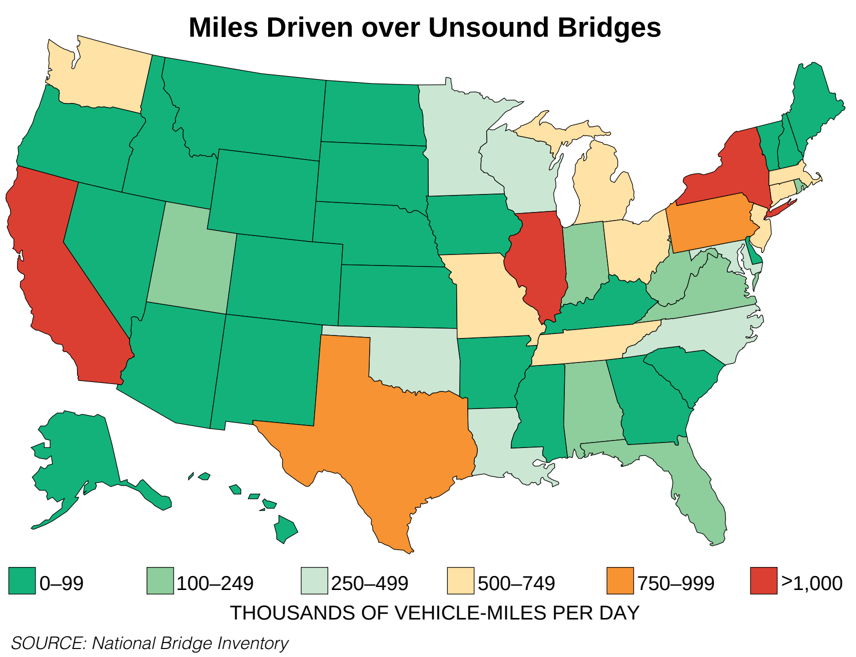
The trucker's reaction to the sign is troubling because bridge engineers believe that increasing the number and severity of overload cycles—such as those caused by an overweight vehicle—accelerate deterioration. Driving down the center stripe, though dangerous from a traffic-safety point of view, limits the overload to one truck on a two-lane bridge. Increasing speed, however, magnifies the effect known as dynamic amplification, which significantly increases the truck's deleterious effect on a bridge. When a truck is moving quickly, potholes or other irregularities in the road surface cause the load to bounce up and down.

U.S. bridge engineers typically estimate that this bouncing imposes an extra load of up to 30 percent of the truck's weight. Tests on one bridge in Australia demonstrated dynamic amplification that doubled the stress on a bridge. As a result, engineers may also post speed limits for heavy vehicles on bridges whose carrying capacity has been reduced. Because such limits are not generally enforced, relying on them to control bridge stresses is risky.

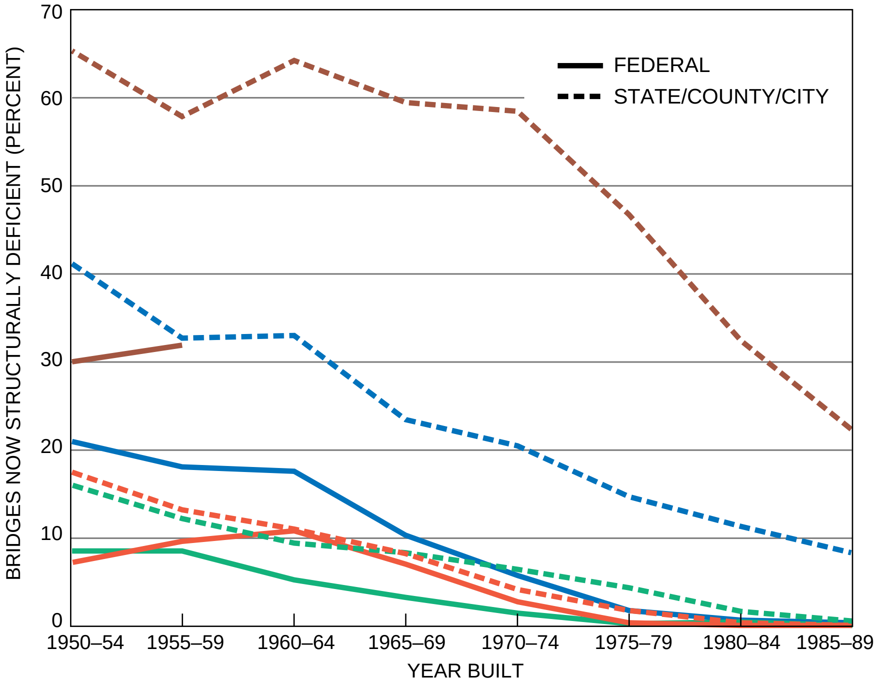
Even assuming that posted load limits could be enforced, simply inspecting bridges and posting them is obviously not a long-term solution to the problem of structural deficiency. Nor is there apparently enough money available in state, city or federal budgets to rebuild every deficient bridge according to current standards.

Some engineers have suggested relaxing bridge design criteria on lightly traveled routes, in effect institutionalizing what is apparently already the practice in some areas. For example, on routes that carry fewer than 10 cars




**PROPORTION OF DEFICIENT BRIDGES shows wide variation from state to state (*top*). Although decaying infrastructure is generally associated with the industrial Northeast and Midwest, the highest structural deficiency rates appear in the southeastern U.S. These percentages, however, do not necessarily correlate with what most drivers see; the bottom map shows the number of vehicle-miles driven each day over unsound bridges.**




**DETERIORATION of bridges depends on construction materials but even more on maintenance policies. Solid lines show the current condition of bridges on interstate and federal routes, where uniform standards are enforced nationwide. Bridges are grouped by year of construction; the line for timber bridges ends at 1959 because very few were built on interstate routes after that date. Dotted lines show the current condition of bridges on state, county and city roads, which have few consistent sources of funding.**

each way in an hour, one-lane bridges would cause negligible delays or hazards, at perhaps two thirds the cost of the two-lane spans now required. Designing bridges to be flooded periodically instead of remaining above high water in all but the worst storms could also reduce costs.

Change in construction methods offers another means of achieving economy. Some states have already begun to make wide use of prefabricated bridge sections—mostly of reinforced or prestressed concrete—that reduce from months to weeks the time required to build bridges. Such bridges tend to have very low structural deficiency percentages, in part because the quality of construction is easier to control in the factory than in the field. In addition, concrete bridges appear to have lower maintenance requirements than steel or timber ones. According to the National Bridge Inventory, the structural deficiency rates for concrete bridges are about the same for bridges on interstate and federal routes (where maintenance standards are high) as for those on state, city and county routes (where standards are at best variable).

In some parts of the country, prefabricated timber bridges are also being built. Laminated beams, pressure-treated with preservatives, form the bridge structure, and laminated panels carry the road surface. Some bridge designers assert that such wood bridges can be built at a cost competitive with that of concrete or steel bridges. Bridges built thus far under the federal Timber Bridge Initiative have failed to live up to the assertion.

Given the magnitude of the funds that could eventually be spent on repairing bridges, it is no surprise that competing industries are jockeying for position. Consequently, it is particularly important that state and federal planners have access to the best possible information about the performance of various bridge types and the condition of bridges that must be strengthened or replaced.

In 1988 the FHWA improved the information-gathering process by revising bridge inspection procedures to make them more uniform from state to state. (Before the revision, the U.S. General Accounting Office had estimated that as many as 15 percent of the bridges in some states were improperly classified.) Instead of applying subjective criteria to structural conditions, inspectors now work with a guide that specifies in detail how bridge problems should be recorded.

Now that the information being collected is uniform among the states, the FHWA is pushing the development and implementation of bridge management programs. These formal methods, backed by computer software, help officials to track bridge conditions, including the progress of scheduled maintenance. Planners can thus analyze the precise nature of deficiencies and spot trends that could presage emergencies. One of the tragicomic aspects of current ad hoc ways of caring for bridges is the neglect of simple, cheap maintenance measures that could slash overall costs. In New York City, for example, observers have estimated that a few tens of thousands of dollars spent on painting and cleaning might have forestalled millions of dollars' worth of structural repairs.

Several state departments of transportation already have such programs, and others are working to put them in place. Once the information-gathering process is complete, planners will know more exactly what is wrong with the nation's physical infrastructure and so have the chance to remedy its troubles in the most cost-effective manner.

A few months after the Silver Bridge collapse, an editorialist at the *Engineering News-Record,* the weekly magazine of the construction industry, warned that "the time, effort and money spent on bridge inspection [should] not grow out of proportion to the problem." Although bridge inspections are far more extensive than anything that could have been imagined in 1968, that warning has proved sadly irrelevant.


FURTHER READING

METHODS OF STRENGTHENING EXISTING HIGHWAY BRIDGES: NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM REPORT #293. F. W. Klaiber, K. F. Dunker, T. J. Wipf and W. W. Sanders, Jr. Transportation Research Board, National Research Council, September 1987.

RURAL ROADS AND BRIDGES: FEDERAL AND STATE FINANCING and RURAL ROADS AND BRIDGES: A DILEMMA FOR LOCAL OFFICIALS. Norman Walzer and David L. Chicoine. U.S. Department of Agriculture, Office of Transportation, April 1989.

THE 1991 STATUS OF THE NATION'S HIGHWAYS AND BRIDGES: CONDITIONS, PERFORMANCE, AND CAPITAL INVESTMENT REQUIREMENTS. U.S. Department of Transportation, Federal Highway Administration, July 2, 1991.

# Black Holes and the Centrifugal Force Paradox

*An object orbiting close to a black hole feels a centrifugal force pushing inward rather than outward. This paradoxical effect has important implications for astrophysics*

by Marek Artur Abramowicz

If you have ever traveled in a car, bus or train as it sped around a bend, you have experienced the centrifugal force: the outward push, away from the center of the curve that grows stronger as the vehicle's speed increases. You can therefore imagine how surprised my colleague A. R. Prasanna of the Physical Research Laboratory in Ahmedabad, India, and I were when we realized recently that Einstein's general theory of relativity predicts that in certain circumstances the centrifugal force may be directed toward, not away from, the center of a circular motion. We demonstrated that if an astronaut manages to steer a spacecraft sufficiently close to some extremely massive and compact object, such as a black hole, the astronaut would feel a centrifugal force pushing inward, not outward. Contrary to everyday experience, an increase in the orbital speed of the rocket strengthens the inward push of the centrifugal force.

According to our calculations, in the region close to a black hole not only does the centrifugal force reverse direction but all dynamic effects that depend on the sense of inward and outward are also reversed. This realization is important for understanding some aspects of the physics of black holes, which are believed to be a crucial part of the mysterious central engines that power the brightest galaxies in the cosmos. Investigations of the centrifugal force paradox have provided some tantalizing insights into the behavior of these galactic energy sources.

The reason for the centrifugal force paradox is the fantastically strong gravitational field produced by a black hole. As Albert Einstein predicted in 1915, a gravitational field warps space and bends light rays. In 1919 Sir Arthur Stanley Eddington confirmed this prediction by measuring the minute deflection of rays passing close to the sun. The gravitational field of the sun will bend a light ray less than one thousandth of a degree if the ray grazes the surface. Because a black hole generates a gravitational field far stronger than that of the sun, it can deflect light to a correspondingly greater extent.

Astronomers have not observed black holes directly, but they have gathered enough indirect evidence to convince most scientists that black holes must really exist. During the past two decades, astronomers have identified many objects that seem to contain black holes. These include several bright x-ray sources in our galaxy and many so-called active galactic nuclei, which are unusually bright cores of some distant galaxies.

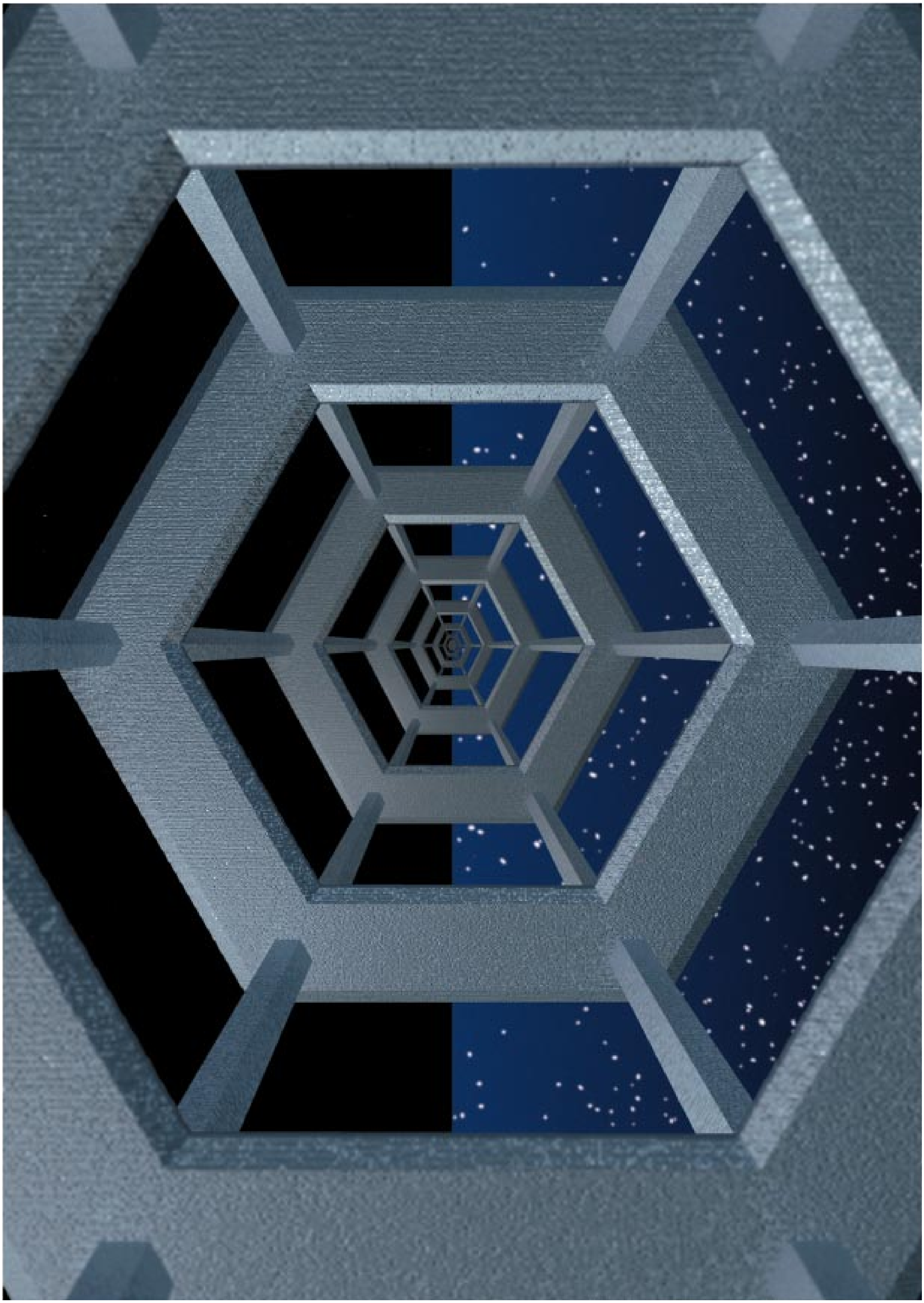
A black hole traps forever any radiation or matter that gets too close to it. This point of no return defines the size of the black hole, or its gravitational radius. A black hole that has the same mass as the sun should have a gravitational radius of about three kilometers. If a light ray travels parallel to the surface of the black hole at a distance equal to, say, three times the gravitational radius, it will be bent by about 45 degrees. Most remarkably, if a light ray passes the black hole at a distance of exactly 1.5 times the gravitational radius, it will orbit the black hole in a perfect circle. The existence of the circular light ray is a key element in the centrifugal force paradox.

Jean-Pierre Lasota (now at the Paris Observatory) and I discovered the first hint of the paradox quite by chance, almost 20 years ago. We were working at the Copernicus Astronomical Center in Warsaw on a rather technical problem in the general theory of relativity. In particular, we were struggling with a complicated formula derived by Bożena Muchotrzeb, one of our students. Something was obviously wrong. The formula yielded a prediction about what force an object would feel if it orbited around a black hole along the same path as a circular light ray. The formula implied that no matter how fast the object moved, it would always feel exactly the same total force pushing inward. In particular, a motionless object would feel exactly the same inward force as a projectile that traveled around the circle at almost the speed of light.

We thought this could be nothing

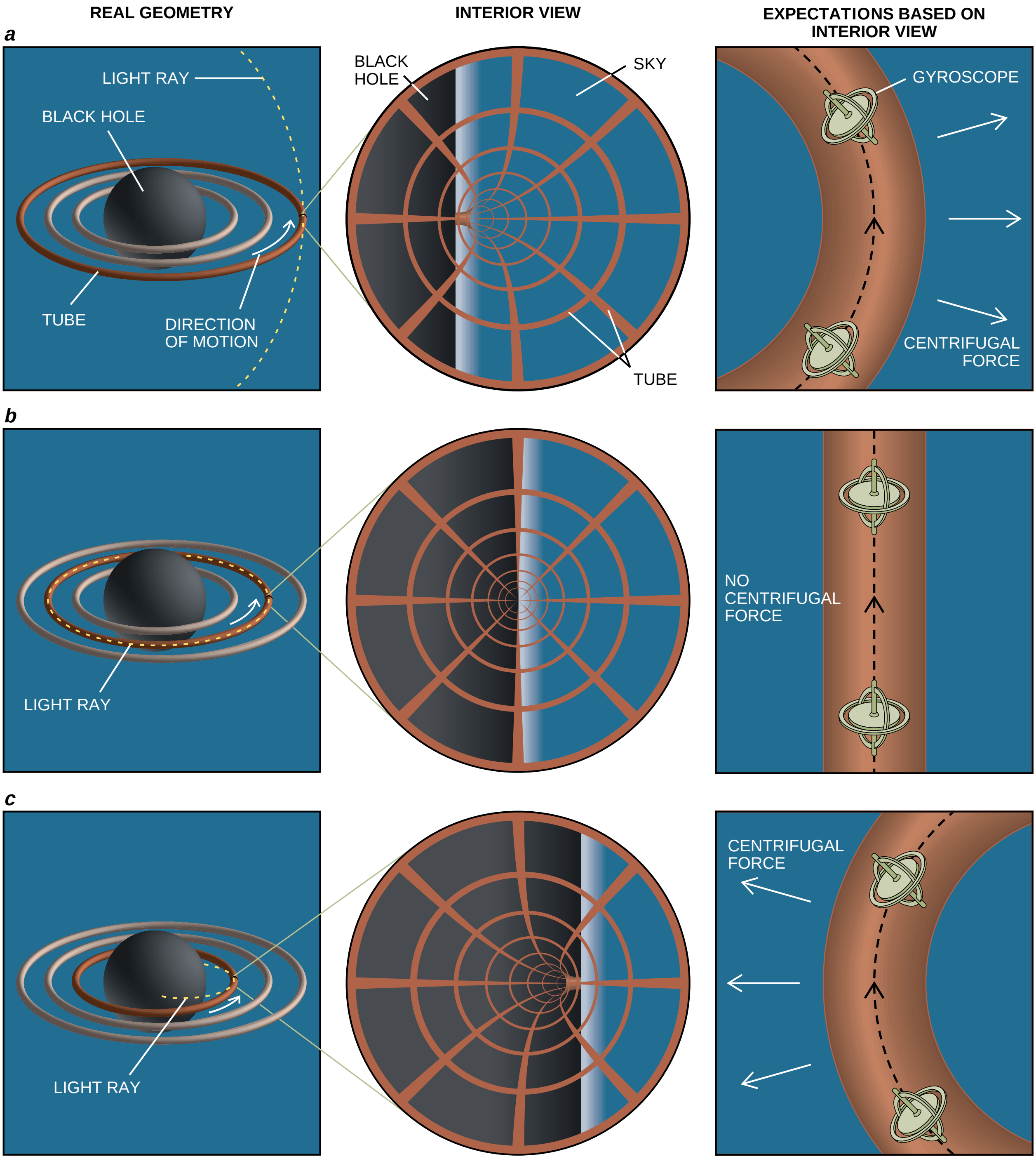
MAREK ARTUR ABRAMOWICZ is chair of the astrophysics department at the University of Göteborg in Sweden. In 1974 he earned his Ph.D. in theoretical physics from the University of Warsaw. Until recently he was assistant professor of astrophysics at NORDITA, the Nordic institute for theoretical physics, in Copenhagen. For more than a decade, he has collaborated closely with Dennis Sciama, first at the University of Oxford and then at the International School for Advanced Studies in Trieste. His interests include a wide variety of issues in astrophysics, from active galactic nuclei to neutron stars to general relativity.

**SPACE STRUCTURE made of girders and hexagonal ribs stretches around a spherical black hole at an altitude equal to 1.5 times the radius of the hole. Although the structure curves around the hole, it would actually appear straight to an observer inside. The effect occurs because at that particular altitude, the gravitational field of the hole is so strong that light rays travel in perfect circles around the hole. Furthermore, an observer traveling around the hole within this structure would feel no centrifugal force [*see box on page 78*]. The slight distortion of distant hexagonal ribs is also a consequence of the bending of light.**



## Centrifugal Forces Near a Black Hole

Three tubes built around a black hole would appear circular to a distant observer but would not necessarily appear that way to someone inside the tubes. The first tube (*a*) is fairly far away from the black hole, where light rays travel in nearly straight lines. In this case, both observers would see the tube curve around the black hole, and both would predict correctly that an object traveling within the tube would be pushed outward, away from the black hole, by the centrifugal force. A gyroscope traveling within the tube will precess as a result of the centrifugal force. The second tube (*b*) is constructed around a region of space where light rays are bent in perfect circles by the gravitational field of the black hole. Because the light bends, the observer inside the tube would see it as perfectly straight and would correctly predict that there should be no centrifugal force. The third tube (*c*) is very close to the black hole. In this case, light rays are curved so much that the tube appears to curve away from the black hole. The observer inside the tube would now predict correctly that the centrifugal force would push an object inward, toward the black hole, and would cause the gyroscope to precess.



but nonsense. According to elementary dynamics, the centrifugal force depends on the orbital speed, whereas the gravitational force does not. Therefore, the total force—which is just the sum of the centrifugal and gravitational forces—must also depend on the orbital speed. Because the formula did not give the answer we expected, we were firmly convinced that it could not possibly be right. Yet after carefully repeating all the calculations in its derivation, we could find no mistakes. As it turned out, the formula was correct, as well as its paradoxical prediction about how matter behaves when traveling along the path of a circular light ray.

There are no true paradoxes in physics. Sometimes we may find a phenomenon paradoxical because of the inertia of our minds—we hold on to an incomplete mental picture that prevents us from understanding how things actually work. Lasota and I realized that motion along the path of a circular ray appears to be so acutely paradoxical because it is difficult to accept the fact that although this light ray is really circular, it is also, in a certain sense, perfectly straight.

To develop the proper intuition about circular light rays, imagine two astronauts (say, Bob and Alice) who conduct experiments inside a space station built around a black hole. The station is a circular tube centered exactly on the path of the circular ray so that the axis of the tube and the path of the ray coincide. The astronauts know that the axis of the tube is circular because Bob has measured the curvature of the walls along the length of the tube using straight rulers. Yet because of the bending of the light rays, they see the tube as perfectly straight!

Imagine that Alice attaches a search lamp to the tube so that it rests in the center. She then starts walking along the tube, away from the search lamp. For Alice, the lamp always appears in the center, and it is never obscured from view by the bend of the tube. Wherever she is, the light of the lamp reaches her along the same circular path. If Alice looks behind her, she sees the lamp become progressively dimmer as she moves farther and farther away. If she peers forward, she sees the lamp become progressively brighter. In fact, the light from the lamp circulates around the tube many times, so Alice sees multiple images of the lamp.

Although Alice might have some difficulty explaining why the lamp appears both behind and in front of her and although she might be confused by the multiple images, she must conclude that the tube is straight because its walls never obscure the lamp. Judging from what she sees, therefore, she would not expect any centrifugal effects to act on objects moving inside the tube. She would deduce that the centrifugal force should be zero. She would also guess that the only force that acts on objects inside the tube is the gravitational force, which does not depend on orbital speed. Alice can make accurate predictions by judging a situation on the basis of what she actually sees. I call this the seeing-is-believing principle.

The true significance of the seeing-is-believing principle was not revealed to me until 1985. One day in the spring of that year I gave an informal lunch talk about the circular light-ray paradox at the Institute for Theoretical Physics in Santa Barbara, Calif. I was fortunate to be able to address several experts in relativity theory, including Brandon Carter of the Paris Observatory. The day after my talk Carter came up with a brilliant idea. He realized that if an object moves at a constant speed along the path of any light ray—circular, curved or straight—the force that keeps the object on course does not depend on how fast the object is moving. To be sure, the object follows the path of a light ray in space, but the speed of the object is, of course, less than that of light.

For example, if a rocket were to follow the path of a light ray past the sun, it would need to turn less than a hundredth of a degree gradually. To stay on course keeping a constant speed, it would need to fire boosters in a direction perpendicular to its trajectory. Yet the force that the boosters would need to exert would not depend on how fast the rocket was moving.

Carter suggested that the seeing-is-believing principle should hold everywhere, in any gravitational field. In other words, if any object traveled at constant speed along the path of a light ray that was curved by some gravitational field, the object would behave as if it were traveling in a straight line. Carter, Lasota and I proved later that this suggestion was correct so long as the associated gravitational field did not change over time. We developed the concept of optical geometry, which provides a very useful framework for understanding the dynamic behavior of objects in strong gravitational fields. Later, John C. Miller of Trieste Astronomical Observatory and Zdĕnk Stuchlík of the Silesian University in Opava discovered some basic relations between dynamics and geometry in this framework, and Norbert Wex of the Max Planck Institute for Physics and Astrophysics in Munich suggested an elegant and clever way of adapting optical geometry to the case of black holes that rotate.

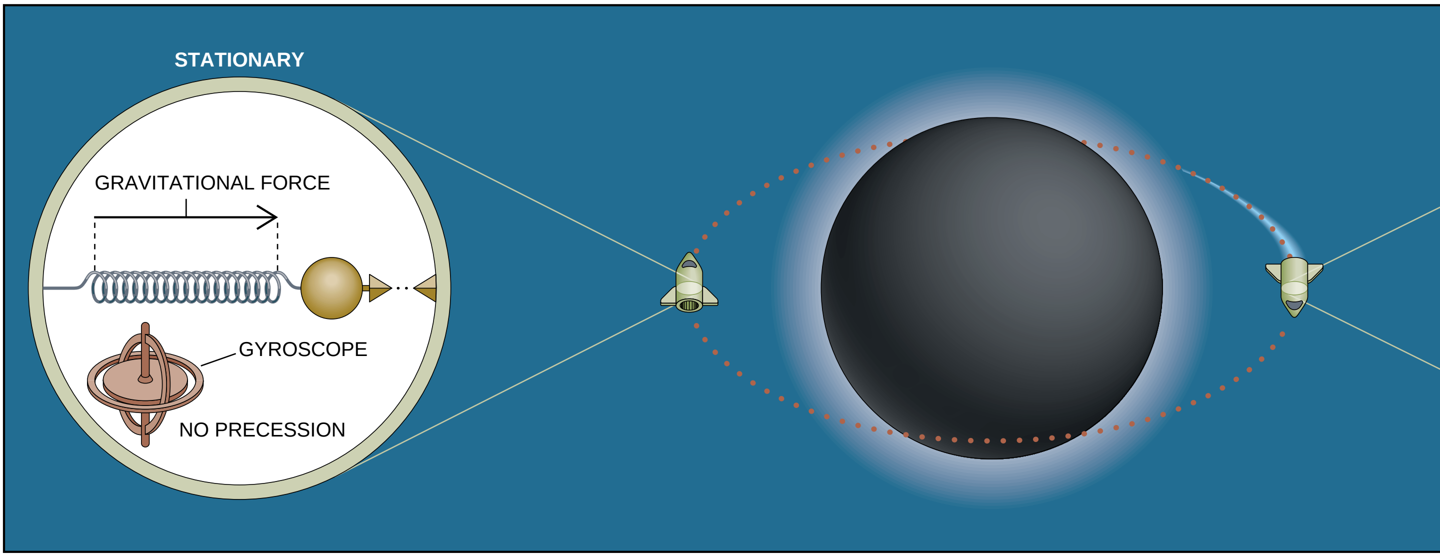
Conventional geometry of space is based on measurements made with standard straight rulers, which define the unit of length. Optical geometry, on the other hand, depends on measurements made using light signals.

In conventional geometry, one can measure the length of a curve by counting how many rulers fit along the curve. The distance between two points in space can then be defined as the length of the shortest curve between them. This shortest curve is known as a geodesic. If one makes measurements in a flat space or, alternatively, in a space free from gravitational fields, the shortest curve, or geodesic, between two points is just a straight line.

In optical geometry the distance between two points in space is defined as half of the time it takes for light to travel from one point to the other and back. The time is measured by a clock located at the first point. In a space free from gravitational fields, optical geometry is exactly the same as conventional geometry because both the light rays and the geodesics are straight. Thus, in this case, the geometry of space is traced by the light rays.

According to Einstein's general theory of relativity, the three dimensions of space and the one dimension of time together form a four-dimensional space-time. In any space-time, with or without a gravitational field, light always moves along geodesics, and therefore it always traces the geometry of space-time. In a space warped by a gravitational field, however, the light rays are curved and in general do not coincide with geodesics. Therefore, in the general case, the geometry of space is not traced by light rays.

Optical geometry restores the connection between the geometry of space and the paths of light rays by rescaling all of the "true" distances (that is, measurements made by straight rulers). The application of optical geometry is similar, in many respects, to the procedure followed when making a flat map from a round globe. Optical geometry is a way of making convenient maps of a curved space, but it has some of the same difficulties that conventional cartography does; namely, a globe cannot be represented on a flat map without some kind of distortion. In both conventional cartography and optical geometry, a particular representation may minimize the aberration of some features while distorting others beyond recognition. The choice




**TWO SPACECRAFT in the same orbit around a black hole can be used to measure the centrifugal force. Each spacecraft has a gyroscope and a weight hung from a spring. As each spacecraft orbits the hole, it is maneuvered so that the weight points to a mark on the hull. One spacecraft adjusts its orbital speed to zero so the gyroscope does not precess; the centrifugal force on the spring must therefore be zero, and the total force equals the gravitational force. The other craft travels at whatever**

of representation is dictated by the purpose for which the map is being made. For example, the well-known Mercator projection exaggerates polar regions but is invaluable to navigators because it shows all lines of constant direction as straight lines. Similarly, optical geometry distorts true distances but is very useful in studying light propagation and dynamics because light rays are geodesics in the map provided by optical geometry. (To be sure, light rays are geodesics at least whenever the gravitational field does not change in time and its material source does not rotate.) Thus, although light propagation and dynamics are not connected in the conventional geometry of space, they are connected in optical geometry.

The rescaling used by optical geometry is an example of a mathematical procedure often used in the theory of relativity and technically known as a conformal transformation. The rescaling straightens the curved light rays, and so they appear as geodesics in optical geometry.

By applying optical geometry, physicists can isolate certain complicated technicalities imposed by the curvature of space and concentrate on the basic physical issues. This type of conformal transformation allows us to understand dynamics in curved spaces intuitively. The dynamics always agrees with what is seen. Optical geometry fully explains the seemingly paradoxical behavior of objects moving along the path of the circular light ray.

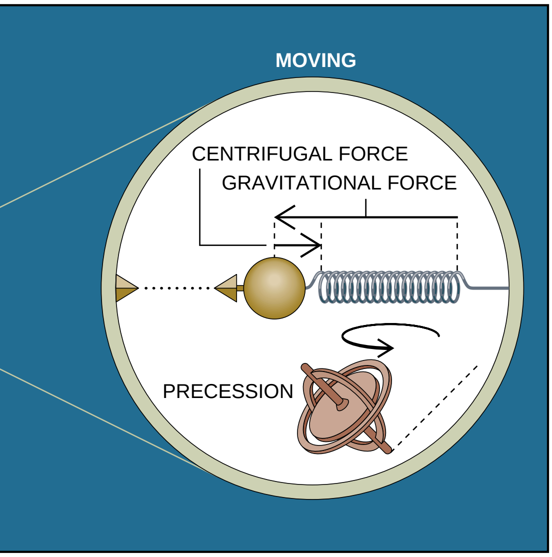
Perhaps the most important general result obtained with the help of optical geometry is that in certain situations space appears to be turned inside out. I realized this when reading a rather technical paper written by Malcolm Anderson and José P. S. Lemos, two research students of Donald Lynden-Bell of the University of Cambridge. Anderson and Lemos demonstrated that if a cloud of gas travels in orbit very close to a black hole, the viscous stresses in the cloud transfer angular momentum inward. This finding was strange because viscous stresses ordinarily transfer angular momentum outward.

Indeed, the outward transfer of angular momentum through viscous stress is a principle of fundamental importance to astrophysicists. It helps to explain how a cloud of gas (known as an accretion disk) orbiting a central black hole supplies the energy that powers the active nuclei of some galaxies. The viscous stress tends to make the rotation of the accretion disk more rigid, thereby slowing down the rapidly rotating inner part of the disk and speeding up the slowly rotating outer part. In this way, the angular momentum is carried outward.

Anderson and Lemos discovered that viscous stress could convey angular momentum inward, but they did not convincingly explain why. After reading their paper, I suddenly realized that optical geometry suggests a powerful explanation of the effect and several similar, surprising results. I found that the space close to the black hole is turned inside out; the outward direction as defined by straight rulers is directly opposite to the outward direction as defined by light rays. In the situation described by Anderson and Lemos, the angular momentum is indeed transported outward, as it should be, but "outward" must always be understood in the sense of optical geometry. In the somewhat familiar situation faraway from a black hole, the outward direction of conventional geometry agrees with that defined by optical geometry. Yet close to the hole these two directions are opposite, and thus the angular momentum is transported inward with respect to conventional geometry, which seems to be paradoxical.

To understand why this is so, you should again envision a circular space station around a black hole where Bob and Alice conduct experiments. In this case, however, the station is not built around the circular light ray but instead is constructed around a smaller circle centered on the black hole. Bob measures the true distances using a standard straight ruler; Alice uses light signals to make her measurements. For convenience, assume Bob and Alice always look down the length of the tube with the black hole on their left. Using a standard ruler, Bob finds that the tube bends to the left. And indeed, his measurements agree with real geometry; if he were simply to touch the tube with his hands, he would feel the walls bend to the left. He concludes that the outward direction is to the right.

Bob knows from everyday experience that the centrifugal force pushes in the outward direction. He would therefore predict that it should push objects to




**speed it chooses. The centrifugal force on its spring can be deduced by measuring the tension and comparing the results with those from the other craft.**

the right. Similarly, he would guess that viscous stresses transport angular momentum to the right. The truth is exactly the opposite.

Alice makes a different set of measurements, based on what she actually sees, and ultimately reaches the right conclusion. She asks Bob to walk away from her, holding a search lamp so it moves along the axis of the tube. Now if somehow the light rays were not bent by the gravitational field of the black hole (that is, the rays were straight), the lamp would disappear behind the left part of the tube, and Alice would conclude that the tube was bent to the left. If the light rays were circular, the lamp would not disappear at all; the tube would seem straight. Yet the tube is so close to the black hole that the light rays bend even more than circular rays. Alice therefore sees the lamp disappear on the right and concludes that the tube bends to the right. Thus, she predicts that the centrifugal force pushes to the left and that the viscous stress transports angular momentum to the left. Her predictions are correct—as guaranteed by the seeing-is-believing principle. Note that in terms of conventional geometry inside the tube, the centrifugal force attracts toward the center of the circular motion.

During the past few years, optical geometry has also been successfully applied to several astrophysical problems involving the behavior of rotating matter in very strong gravitational fields. The two most important problems of this type are the gravitational collapse of rotating stars and the coalescence of two extremely dense objects known as neutron stars. John Miller and I have demonstrated that optical geometry can be very useful in tackling these problems. We have provided a simple explanation for the strange behavior of the shape of a rotating star undergoing a contraction. Using nonrelativistic theory, one expects that if a rotating body of gas shrinks while conserving its mass and angular momentum, it must become progressively flatter. In 1974, however, Subrahmanyan Chandrasekhar of the University of Chicago and Miller, then at the University of Oxford, found that, according to Einstein's theory, in the last stages of the contraction, when the gravitational field is very strong, the increase in flattening ceases and the rotating star becomes more spherical. Miller and I found the correct explanation for this effect by using optical geometry and by considering the unusual behavior of the centrifugal force in the strong gravitational field.

It took quite some effort to convince my colleagues that the reversal of the centrifugal force is a real physical effect. The issue was how one could possibly define and measure the centrifugal force in a strongly curved space. This question raises several rather subtle points that my critics and I examined with great care by performing many lengthy calculations. I made most of my progress by answering several challenging questions raised by Fernando de Felice of Torino University. As a result of my friendly battles with de Felice, I have adopted a particular definition of centrifugal force. My definition is not unique, but I find alternative ones less useful and convincing.

For the purpose of measuring centrifugal force, I imagine two spaceships that travel in the same orbit around a black hole. Bob now pilots one craft; Alice commands the other. Each spacecraft carries two pieces of equipment: a gyroscope and a weight that hangs on a spring. By measuring its length, Bob and Alice can determine the tension in the spring. The tension, in turn, equals the sum of the two forces acting on the weight: the gravitational force and the centrifugal force.

To measure either one of these forces alone, Bob and Alice must change the orientation of their spaceships as they orbit the black hole; both pilots must rotate their spacecraft so that the stretched spring always points toward a mark on the hull. The direction of the spring is therefore fixed in the ship but not in space. On the other hand, the gyroscope in each ship always points to a fixed direction in space, and therefore it will precess relative to the direction of the spring as the ship moves along its orbit.

To measure the gravitational force, Bob brings his ship to a halt; he knows when he has stopped because his gyroscope does not precess. He can then conclude that the force stretching the spring is the gravitational force alone. Bob communicates his result to Alice, who continues to speed around the black hole on the same orbit. Alice measures the total force that stretches her spring and finds the centrifugal force by subtracting the gravitational force that Bob measured. Although this method for measuring centrifugal force seems elaborate, it has the advantage of being exactly the same in both weak and strong gravitational fields.

The practical value of optical geometry is that it provides a convenient way of handling difficult problems in general relativity. It is also useful pedagogically because it leads to an intuitive understanding of several effects of relativity that are important to modern astrophysics. With the help of optical geometry, these phenomena no longer seem to be paradoxical or confusing.

On a more basic level, optical geometry shows that "inward" and "outward" are not absolute concepts; they are relative in spaces warped by strong gravitational fields. Today we have no problem understanding that left and right, as well as up and down, are relative. Yet that was not always the case. In biblical times some people considered left and right to be absolute—a belief supported by some passages in the Old Testament and other ancient scriptures. A few centuries ago some thought that up and down were absolute; they could not imagine the inhabitants of the opposite side of the earth walking upside down. Perhaps by the end of the next century, no one will be surprised that inward and outward are relative.

FURTHER READING

ON TRAVELLING ROUND WITHOUT FEELING IT AND UNCURVING CURVES. M. A. Abramowicz and J.-P. Lasota in *American Journal of Physics,* Vol. 54, pages 936–939; October 1986.

REVERSING CENTRIFUGAL FORCES. Bruce Allen in *Nature,* Vol. 347, No. 6294, pages 615–616; October 18, 1990.

RELATIVITY OF INWARDS AND OUTWARDS: AN EXAMPLE. M. A. Abramowicz in *Monthly Notices of the Royal Astronomical Society,* Vol. 256, No. 4, pages 710–718; June 15, 1992.

THE WALL OF DEATH. M. A. Abramowicz and E. Szuszkiewicz in *American Journal of Physics* (in press).

# Teaching the Immune System To Fight Cancer

*Certain molecules on tumors can serve as targets for attack by cells of the immune system. These tumor-rejection antigens may provide a basis for precisely targeted anticancer therapy*

by Thierry Boon

At its best, the immune system is the ideal weapon against infectious disease. It eliminates viruses and bacteria that invade the body and kills infected cells, yet it leaves healthy tissue intact. The system is so precise because it responds only to specific targets called antigens: molecules or fragments of molecules that belong to the foreign invaders. In general, antibody molecules inactivate pathogens and toxins that circulate in body fluids, whereas white blood cells called cytolytic *T* lymphocytes destroy ("lyse") cells that have been penetrated by viruses.

The specificity and power of the immune system have not escaped notice of cancer researchers. Assuming that *T* lymphocytes might be able to eradicate cancer cells as effectively as they lyse virus-infected cells, investigators have long hoped to identify tumor-rejection antigens: structures that *T* lymphocytes can recognize on tumor cells in the body. These workers reasoned that antigens appearing exclusively (or almost exclusively) on cancer cells could be manipulated in ways that would trigger or amplify a patient's insufficient immune reaction to those targets.

Definitive evidence that tumor-rejection antigens exist on human tumors has been elusive. Yet in the past few years, my colleagues and I at the Ludwig Institute for Cancer Research in Brussels have gathered unequivocal proof that many, perhaps most, tumors do indeed display such antigens. Equally important, we have developed ways to isolate genes that specify the structure of these antigens. Moreover, we and others have seen indications that *T* lymphocytes that normally ignore existing tumor-rejection antigens can be prodded to respond to them. Hence, the design of therapies to generate such *T* cell responses to well-defined tumor-rejection antigens has finally become feasible.

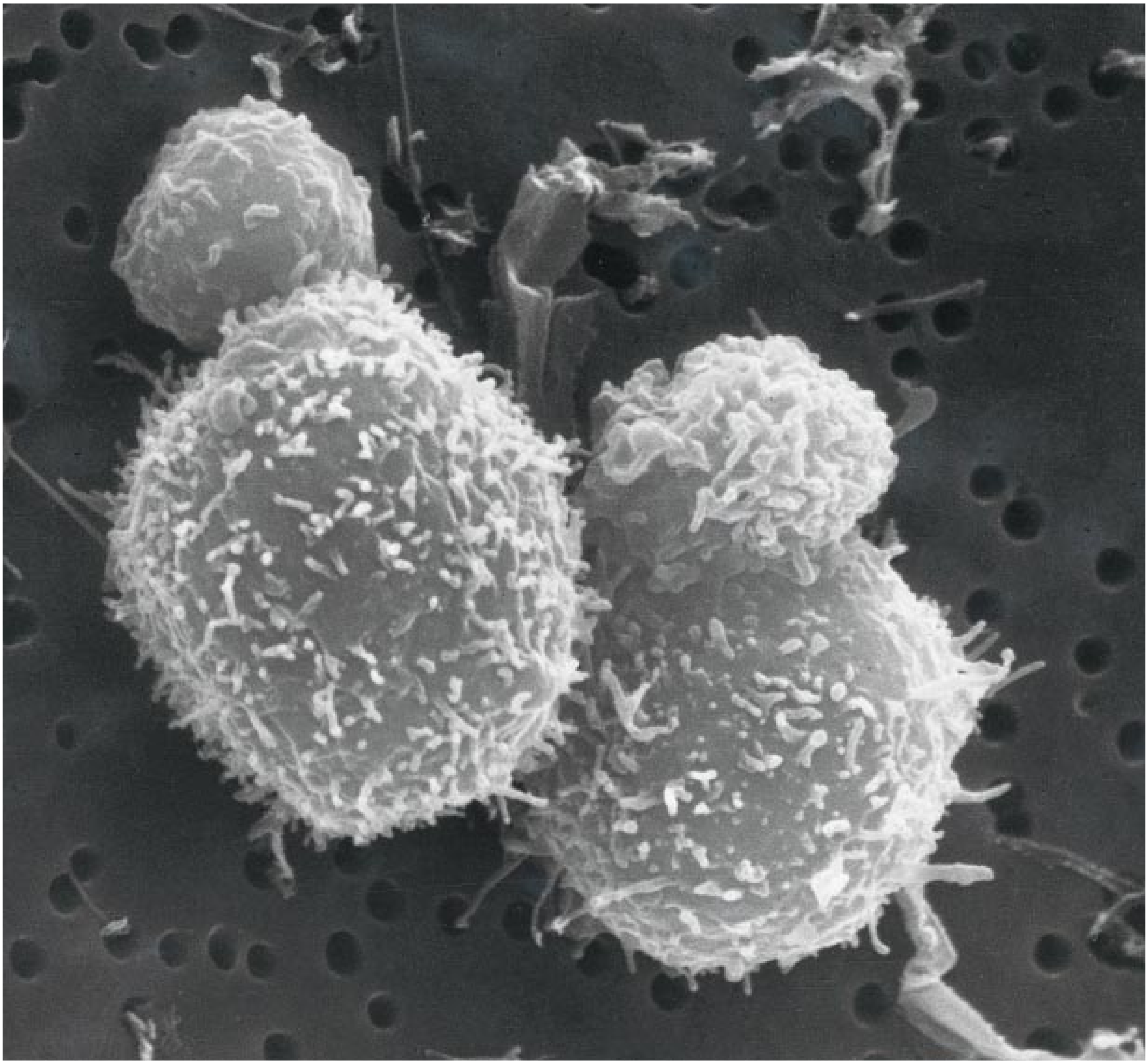
The first clues that tumor-rejection antigens sometimes arise on tumors were uncovered in the 1950s, before the distinct roles of antibodies and *T* cells were elucidated. Several researchers—notably E. J. Foley of the Schering Corporation in Bloomfield, N.J., Richmond T. Prehn and Joan M. Main of the National Cancer Institute and George Klein of the Karolinska Institute Medical School in Stockholm—had generated cancers in mice by treating the animals with large doses of a carcinogenic compound. When the mice were freed of their tumors by surgery and subsequently injected with cells of the same tumor, they did not suffer a recurrence. The mice did acquire cancer after being injected with cells from other tumors, however. Those observations suggested that cells of carcinogen-induced tumors carry antigens that can elicit a response by the immune system.

For about 20 years after those pioneering experiments were completed, hope ran high that human cancers, too, might bear tumor-rejection antigens. The prospect for antigen-based therapy seemed even better when, toward the end of that period, *T* lymphocytes were found to be particularly important for ridding the body of abnormal cells. Jean-Charles Cerottini and K. Theodor Brunner of the Swiss Institute for Experimental Cancer Research in Lausanne showed that when mice reject tissue transplanted from an unrelated donor, the animals produce cytolytic *T* lymphocytes that can destroy cells from the transplant. By then it was apparent as well that when the specialized antigen receptors on cytolytic *T* lymphocytes bind to foreign antigens on a cell, the lymphocytes both lyse the cell and multiply, amplifying the immune reaction. These discoveries intimated that cancer researchers might make major strides if they concentrated on finding the antigenic targets of cytolytic *T* lymphocytes and on augmenting the activity of the cytolytic cells.

In the mid-1970s, however, experiments reported by Harold B. Hewitt, then at Mount Vernon Hospital in London, ushered in an era of pessimism. In contrast to the earlier experiments, which examined tumors induced by exposure to massive doses of carcinogens, Hewitt looked for evidence of tumor-rejection antigens on spontaneously arising malignancies. His careful work, conducted on many types of cancer, strongly suggested that spontaneous tumors in mammals did not evoke any immune rejection. Hence, he argued, the observations made in the earlier studies had little relevance to human tumors; people are rarely exposed to the high levels of carcinogens with which scientists produce malignancies in the laboratory.

Reasonably, many investigators then turned their attention elsewhere. Yet between 1972 and 1976 my colleagues and I had seen indications that tumor-rejection antigens were present on several mouse tumors that failed to elicit

THIERRY BOON has been director of the Brussels branch of the Ludwig Institute for Cancer Research since 1978 and professor of genetics and immunology at the Catholic University of Louvain since 1980. After earning a Ph.D. in molecular genetics from the Rockefeller University in 1970, he worked as a research associate at the Pasteur Institute in Paris. In 1975 he became an associate professor at the University of Louvain and also established a laboratory at the International Institute of Cellular and Molecular Pathology (ICP) in Brussels. His laboratory is now part of the Ludwig Institute.



**WHITE BLOOD CELLS called cytolytic *T* lymphocytes (*small spheres*) are attacking two cells from a mouse tumor called P815 (*large spheres*). Such lymphocytes bind to tumor cells when they recognize specific targets known as tumor-rejection antigens on the cell surface. Investigators have now found ways to identify the antigens with certainty. They hope to incorporate such antigens into therapies that will incite a patient's own *T* lymphocytes to eradicate tumors.**

an immune rejection response. In addition, we discovered that the initially ineffective antigens could become useful targets for a defensive assault if the immune system were somehow made more aware of their existence. And so, even after Hewitt published his data, we remained hopeful that immunotherapies based on tumor-rejection antigens might be possible for humans.

As often happens in science, we were studying a totally unrelated problem in 1972 when we stumbled onto those first clues. We were trying to identify genes that control the way cells in mammalian embryos differentiate to become the specialized cells of mature organisms. My colleague Odile Kellermann and I, then at the Pasteur Institute in Paris, had exposed a culture of mouse tumor cells to a potent mutagen, a compound that introduces random, permanent changes (mutations) in genes. Then we put individual treated cells in separate plastic dishes and allowed them to proliferate so that each dish eventually contained a population of identical cells (a clone). That done, we transferred the clonal populations into mice and examined the cell types present in the tumors that resulted.

To our disappointment, the experiments did not lead to a better understanding of the mechanisms of differentiation. But they did turn up a highly intriguing phenomenon. The original, or parent, tumor cells (those not yet exposed to the mutagen) almost always yielded cancerous growths when injected into mice. Yet many of the mutagen-treated clones produced no malignancies. Although I was a geneticist by training and knew little about cancer, I felt impelled to find out why the mutagen-treated cells did not form tumors. For simplicity's sake, my associates and I referred to cell clones that failed to generate tumors as $tum^-$ variants.

We found that the $tum^-$ variants

caused no cancer because the immune system of the injected mice had destroyed them, much as it might reject a mismatched kidney transplant. We found as well that the rejection occurred because the mutagen induced the tum$^{-}$ cells to display one or more antigens (tum$^{-}$ antigens) that elicited a potent *T* lymphocyte response; these tum$^{-}$ antigens were not present on the parental, tumor-inducing (tumorigenic) cell line and appeared to be different for every tum$^{-}$ variant.

The results were interesting by themselves, but what truly captivated us was a second finding I obtained with Aline Van Pel, after we joined the International Institute of Cellular and Molecular Pathology (ICP) in Brussels. As was true of the spontaneous cancers studied by Hewitt, the cells of our original tumor were totally incapable of eliciting an immune attack. Yet often when we injected these cells into mice that had rejected one or another tum$^{-}$ variant, no cancer developed. In mounting an immune response to a tum$^{-}$ variant, the mice somehow acquired resistance to the original tumor cell. The mice did not resist unrelated cancers, however, indicating that rejection of the original tumor cells was caused by an antigen shared by the tum$^{-}$ variant and its parent but not by other cancers.

Our findings were later confirmed in several follow-up studies involving many different mouse tumors. Most important, Van Pel observed that she could reproduce our results with the very spontaneous tumors Hewitt had examined. Clearly, the conclusion that spontaneous cancers did not display tumor-rejection antigens had to be revised.

No one has fully explained how tum$^{-}$ variants manage to induce a powerful immune response to the initially ineffectual, or weak, antigens on the original cells. We suspect that small proteins called interleukins play a role. A lymphocyte that has bound to an antigen releases interleukins. These proteins, in turn, promote proliferation of that lymphocyte and nearby ones (such as those bound to another antigen on the same tumor cell or on neighboring cells). It seems probable that the tum$^{-}$ antigens are potent enough to spur *T* lymphocytes to kill tum$^{-}$ cells and to multiply rapidly even in the absence of preexisting interleukins in the local environment. These lymphocytes then produce interleukins, which help other *T* cells become activated by weak tumor-rejection antigens. Consistent with this view is the fact that in recent years several research groups have modified tumor cells to secrete interleukins. In many instances, the workers have seen a considerable increase in the immune response to the tumors.

By the early 1980s, then, our collected evidence suggested the following conclusion: mouse tumors that normally fail to elicit a buildup of *T* lymphocytes nonetheless often carry weak antigens that can become targets for an effective immune assault. Because the immune system of mice is much like that of humans, the data implied that human tumors might be antigenic as well. If so, they might be susceptible to immunotherapy that artificially induced an antigen-specific attack. In other words, immunotherapy for humans was a reasonable goal. At that point, we decided to apply all the forces of our laboratory to the study of tumor-rejection antigens.

Before considering therapy, we would have to identify specific tumor-rejection antigens. All earlier attempts to isolate such structures directly from cell membranes in human and mouse tumors had failed. We therefore decided to try an alternative approach: cloning, or isolating, the genes that direct construction of the antigens. Unfortunately, no one had yet come up with a good way to perform the task. And so in 1983 my colleagues and I, by then members of the Ludwig Institute, set out to develop a method of our own. It took us four years to devise an approach that would work in a test system [*see box on opposite page*].

In our first successful cloning effort, we isolated the gene for the tum$^{-}$ antigen appearing on the cells of a mouse tum$^{-}$ variant. Of course, tum$^{-}$ antigens are not true tumor-rejection antigens, because they are artificially induced to appear on cultured tumor cells and are not found on cancers in the body. But, as will be seen, they were useful for our trial run. We generated the tum$^{-}$ variant from a cell line that was derived from a mastocytoma (mast cell tumor) named P815. The original P815 cell line was appealing for our purposes because the cells replicate rapidly and indefinitely in the test tube. In addition, tum$^{-}$ variants of P815 cells provoke cytolytic *T* lymphocytes into a strong, readily detectable response.

Our gene-cloning plan relied first of all on having a good supply of cytolytic *T* cells reactive to the tum$^{-}$ antigen of the variant. The *T* cells would later lead us to the gene for the antigen. To acquire the cytolytic cells, we injected the P815 tum$^{-}$ variant into mice. Then we removed the spleen (a repository of lymphocytes) from animals that rejected the variant. We knew that if the lymphocytes from these immunized animals were exposed to killed cells of the variant, cytolytic *T* lymphocytes specific for the variant would multiply preferentially; other lymphocytes would disappear. (Tumor cells would be killed to prevent them from overtaking the culture.) When this culturing was done, we had a supply of cytolytic *T* lymphocytes of which some responded to the tum$^{-}$ antigen and others to tumor-rejection antigens present on all P815 cells. By placing individual lymphocytes in laboratory dishes and allowing them to replicate separately, we obtained several clones that would lyse only the tum$^{-}$ variant and could be made to multiply indefinitely in laboratory dishes. We chose one of the clones directed against the tum$^{-}$ antigen to use in the quest for the gene.

In outline, the plan for isolating the gene for the tum$^{-}$ antigen was straightforward. We intended to collect all the genetic material of the variant. Next we would link fragments of this DNA to pieces of bacterial DNA, which would later serve as labels to help retrieve the gene of interest. We would introduce the fragments into cells that do not normally produce the tum$^{-}$ antigen. Then we would test the ability of each of these cells to stimulate our *T* lymphocyte clone. We would know that a recipient cell displayed the antigen (and thus had taken up the corresponding gene) if the cell spurred the lymphocytes to proliferate. By searching for the bacterial label we had attached to the DNA of the tum$^{-}$ variant, we would locate and retrieve from the DNA of the recipient cell the gene for the tum$^{-}$ antigen.

Although the plan was relatively simple conceptually, the implementation was quite laborious. Mammalian cells contain approximately 100,000 distinct genes, spread throughout roughly three billion nucleotides (the building blocks of DNA) in the chromosomes inherited from each parent. Because of inefficiencies in the techniques available for inserting DNA into recipient cells, we had to create a gene "library" containing millions of copies of each gene. These copies were obtained by splicing fragments of the DNA from the tum$^{-}$ variant into 300,000 plasmids, or circular bits of bacterial DNA; each such plasmid carried about 40,000 nucleotides of inserted tum$^{-}$ DNA (containing an average of one or two genes). After allowing the plasmids to multiply in bacteria, we recovered the DNA.

Next we selected as the recipient a cell type that could incorporate such plasmids into its chromosomes. The original P815 line proved suitable. To ensure that at least one copy of each gene in the tum$^{-}$ variant would fit into the DNA of the recipient P815 cells, we had to mix the recovered plasmids with more than 300 million P815 cells. We needed that many because we knew only about one in 10,000 of the cells would take up DNA. We also knew that these few cells would accept a lot of DNA—500,000 nucleotides on average.

Fortunately, we were able to avoid having to test every last cell for its ability to activate the selected clone of *T* lymphocytes. We did so by including in the bacterial DNA a gene that conferred resistance to a particular toxic drug. When we treated the full set of cells with the drug, we eliminated all those that had failed to integrate a plasmid into their DNA. We were thus left to test just 30,000 of the original 300 million P815 cells. By testing small groups of the 30,000 cells, we found the few that stimulated the *T* lymphocytes to multiply. We then homed in on the bacterial DNA of one of these cells and thus picked out the tum⁻ DNA. By repeating much the same process with this DNA fragment, we were soon able to isolate the gene giving rise to the tum⁻ antigen.

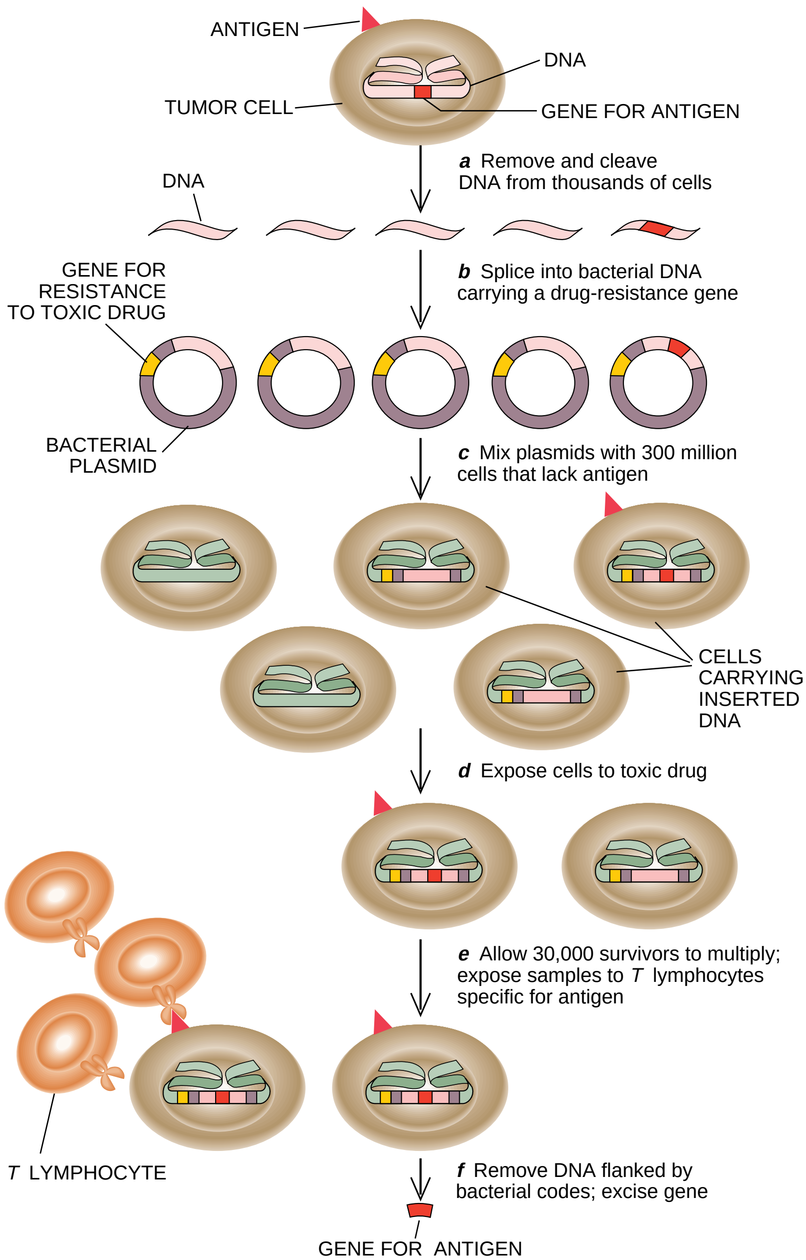
We quickly deciphered the sequence of nucleotides in the gene. The sequence did not resemble that of any gene known at the time. We did find, however, that the gene was expressed not only in the tum⁻ variant but also in the original P815 cells and in normal mouse tissue. That is, the gene, which specifies the sequence of amino acids to be strung together into a protein, was being transcribed into molecules of messenger RNA that were, in turn, being translated into protein.

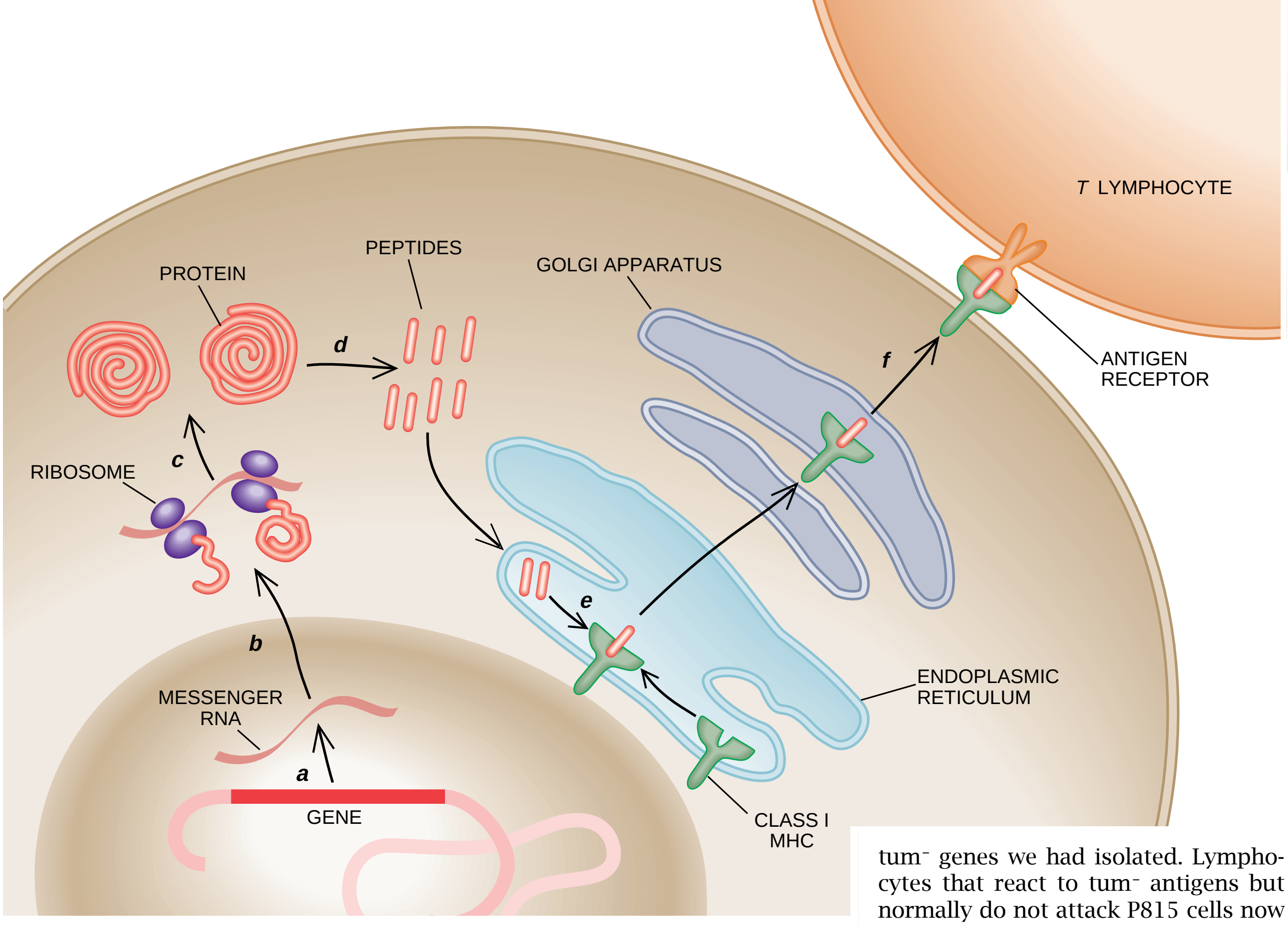
Expression in normal cells meant that our gene specified a standard component of cells. But all was not normal in the tum⁻ variant. There the gene had suffered a point mutation, causing one amino acid to be substituted for another in the protein product. The same was true of two other tum⁻ genes we cloned later. We were puzzled. How could a single amino acid substitution transform a constituent of normal cells into a strong antigen recognized by cytolytic *T* lymphocytes?

Just when we were asking this question, Alain R. M. Townsend of John Radcliffe Hospital in Oxford, England, and his colleagues made a discovery that led us to the answer. In 1986 they demonstrated that cytolytic *T* lymphocytes can often detect viral proteins hidden within cells. In contrast, antibodies respond only to proteins that exert their functions on the cell surface. The *T* cells can accomplish this feat because, in the course of mammalian evolution,

## How Genes for Antigens Recognized by *T* Lymphocytes Are Cloned

Cloning, or isolation, of a gene (*red band in nucleus*) for an antigen (*red triangle*) on a tumor cell begins with removal and cleavage of DNA from multiple copies of the cell (*a*). Workers insert the resulting DNA fragments into plasmids (rings of bacterial DNA) bearing a gene (*yellow*) that confers resistance to a toxic drug (*b*). They mix the plasmids with cells that lack the antigen, causing some of those cells to take up one or more plasmids (*c*). Next investigators expose the cells to the toxic drug (*d*), thereby eliminating any cells that have failed to incorporate the plasmid DNA into their own DNA. The surviving cells are allowed to multiply, and samples are exposed to *T* lymphocytes that specifically recognize the antigen of interest (*e*). Any cell that induces a lymphocyte response (such as proliferation) can be assumed to produce the antigen, which means it also harbors the corresponding gene. Hence, researchers remove the foreign DNA from an identical cell, excise the bacterial DNA and fish out the desired antigen-specifying gene (*f*).






**CELLS PRODUCE ANTIGENS (*red and green complex at top right*) in a multistep process. Once a gene (*red band at bottom*) directs synthesis of a protein (*a–c*), cellular enzymes chop these proteins (*large red coils*) into fragments (*small red bars*) called peptides (*d*). Some of these peptides are then transported into an intracellular compartment (the endoplasmic reticulum) (*e*), where they may combine with so-called class I major histocompatibility (MHC) molecules (*green*). Such peptide-MHC complexes are transported to the cell surface (*f*), where *T* lymphocytes (*orange body at top right*) can examine them.**

an elaborate protein-surveillance system has arisen. Cellular enzymes routinely chop a fraction of all the proteins in the cytoplasm into small fragments known as peptides. These peptides are transported to a special intracellular compartment, the endoplasmic reticulum. There some of them fit themselves into a groove within specialized proteins known as class I major histocompatibility (MHC) molecules. (In humans, MHC molecules are also called human leukocyte antigens, or HLA molecules.) The MHC-peptide complexes move to the surface and become anchored in the cell membrane, ready to be scrutinized by cytolytic *T* cells. Lymphocytes whose antigen receptors can bind to such a complex may then attack the cell. Thus, peptides derived from normal proteins are continuously displayed. This presentation does no harm because of natural tolerance: early in life the body eliminates all *T* lymphocytes that recognize the constituents of the self. But if a peptide is derived from a foreign protein, such as that of a virus hiding within a cell, then a *T* lymphocyte will notice it and attempt to kill the cell.

On the basis of these discoveries, we surmised that the point mutations in the three tum⁻ genes had converted peptides that were not seen by *T* lymphocytes to ones that were seen. To test this idea, we made use of a crucial observation of Townsend and his colleagues. They had found that healthy cells could be rendered instantly recognizable to antiviral cytolytic *T* lymphocytes if the cells were put in a medium containing a synthetic version of a small peptide belonging to a viral protein. Presumably, the healthy cells stimulated the lymphocytes because a few MHC molecules on the surface had taken up the peptides and presented them to the *T* cells.

We conducted similar experiments to reveal the role of tum⁻ mutations. We mixed P815 cells with small peptides (of nine to 10 amino acids) coded for by the mutated regions of the three tum⁻ genes we had isolated. Lymphocytes that react to tum⁻ antigens but normally do not attack P815 cells now lysed the cells. But the lymphocytes did not lyse P815 cells that were mixed with peptides encoded by the normal sequences of the genes. Later we showed that the point mutations in two of the tum⁻ genes had rendered the affected peptides capable of binding to MHC molecules. The normal versions of these peptides do not bind and consequently are never displayed to the immune system. For the third mutated gene, the situation was different. The normal version of the altered peptide does in fact bind to MHC molecules. But because it is a constituent of the self, the process of natural tolerance had eliminated any *T* lymphocytes responsive to it. The mutation changed the shape of the exposed part of the peptide so that the peptide could now be detected by an existing *T* cell population.

Conceivably, a mutation in virtually any gene can result in the appearance of a new antigen on a cell. Accordingly, an infinite variety of antigens can be produced by random mutations. The diverse antigens that appear on rodent tumors induced by chemical carcinogens probably arise through such a mechanism. In addition, mutations can occasionally transform normal genes into ones that cause cancer (oncogenes). Some of these oncogenic changes may well generate antigenic peptides that will one day serve as targets for specific immunotherapy.

Having demonstrated the merit of our

cloning technique, we set about isolating a gene of a bona fide tumor-rejection antigen—one present on a cancer that grows in an animal. Fortunately, we had at our disposal a cytolytic *T* lymphocyte clone that lysed the original P815 cells and did not lyse normal mouse cells. Clearly, the gene specifying the tumor-rejection antigen (named P815A) recognized by these lymphocytes was a logical target for our gene search.

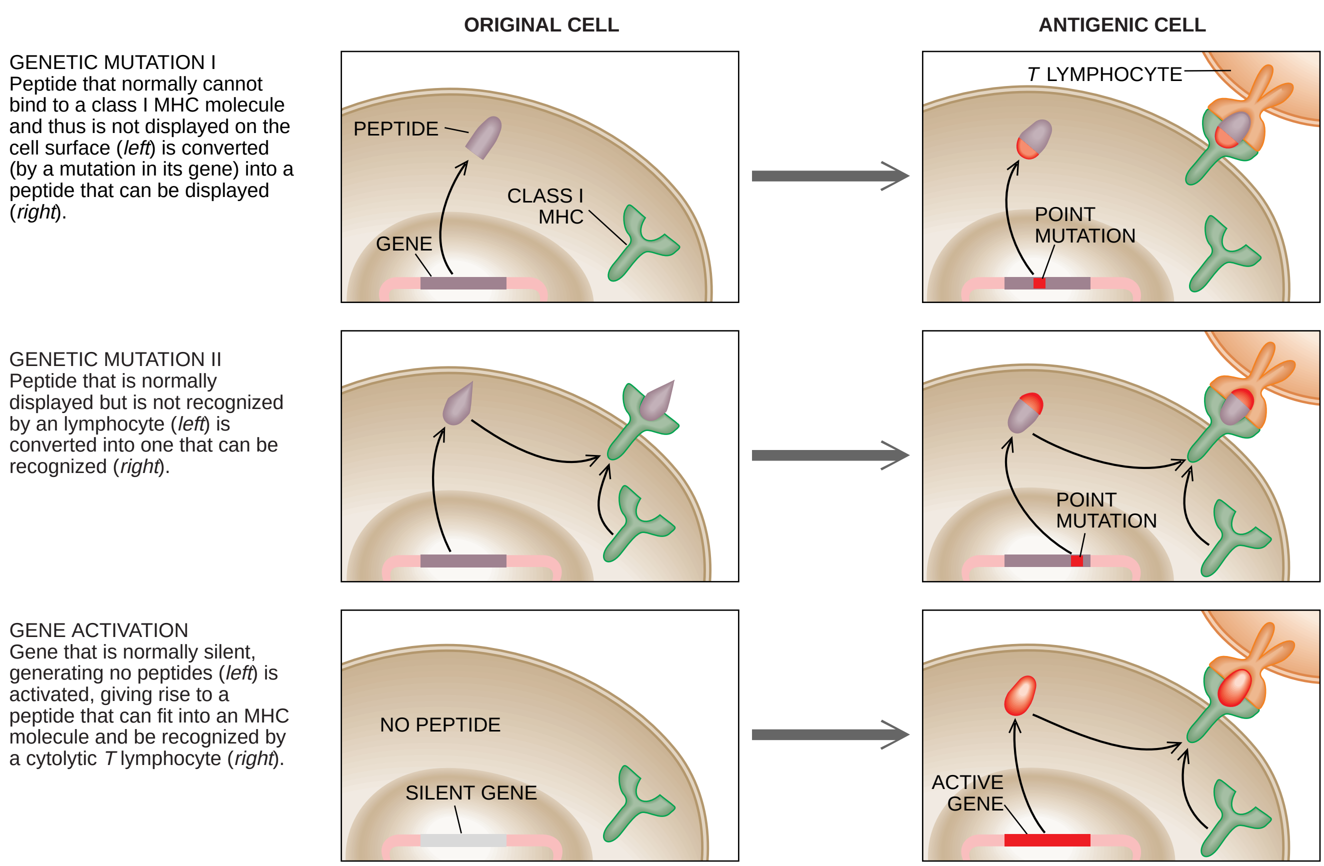
Before starting, however, we wanted to be sure this antigen—which was identified by cytolytic *T* lymphocytes in the test tube—could also direct an immune response to a tumor in the body. We were able to address this question because we had observed an odd effect of P815 cells. Usually when mice are injected with those cells, tumors appear within a month. Yet a few mice formed tumors only after a long delay. When malignancies finally emerged, they resisted attack by the cytolytic *T* lymphocytes responsive to P815A. We concluded, correctly as it turns out, that these animals had rejected almost all the P815 cancer cells because, in the body, *T* lymphocytes identical to those in the clone had recognized antigen P815A. But a few tumor cells had stopped displaying P815A because they had lost the gene specifying it. These so-called antigen-loss variants had proliferated, accounting for the eventual tumor formation. This work demonstrated that an antigen recognized by cytolytic *T* lymphocytes in a laboratory dish might also be of value for eliciting a tumor-rejection response in the body.

Conveniently, such antigen-loss variants could be used as DNA recipients in our efforts to clone the gene coding for antigen P815A. We isolated the gene by applying our by then well-tuned cloning procedure. We built a gene library with DNA from P815 cells and transferred this DNA into cells of an antigen-loss variant. We then fished out the gene from one of the few recipients that incited proliferation of our *T* lymphocytes responsive to antigen P815A. We named the gene *P1A.*

The nucleotide sequence of the *P1A* gene was found to be identical in P815 cells and in normal mouse cells. But in normal cells the gene is inactive; it produces no protein and therefore no antigenic peptide. P815 tumors express the gene and thereby generate an antigen that does not appear on normal cells. Thus, expression of usually silent genes is yet another mechanism of antigen formation. We expected that this last mechanism would generate antigens common to tumors of many different individuals. After all, probably only a relatively limited set of genes can help cancer cells multiply and spread throughout the body. Therefore, we were not surprised to observe that several mastocytoma tumor cell lines express the *P1A* gene, whereas normal mast cells do not.

By 1989 we were ready to begin searching for genes encoding tumor-rejection antigens on human cells. We focused on a cell line named MZ2-MEL, derived from a melanoma tumor (a form of skin cancer) that had formed in a 35-year-old woman known as patient MZ2. We isolated a gene on the cell line in much the same way we obtained the mouse *P1A* gene.

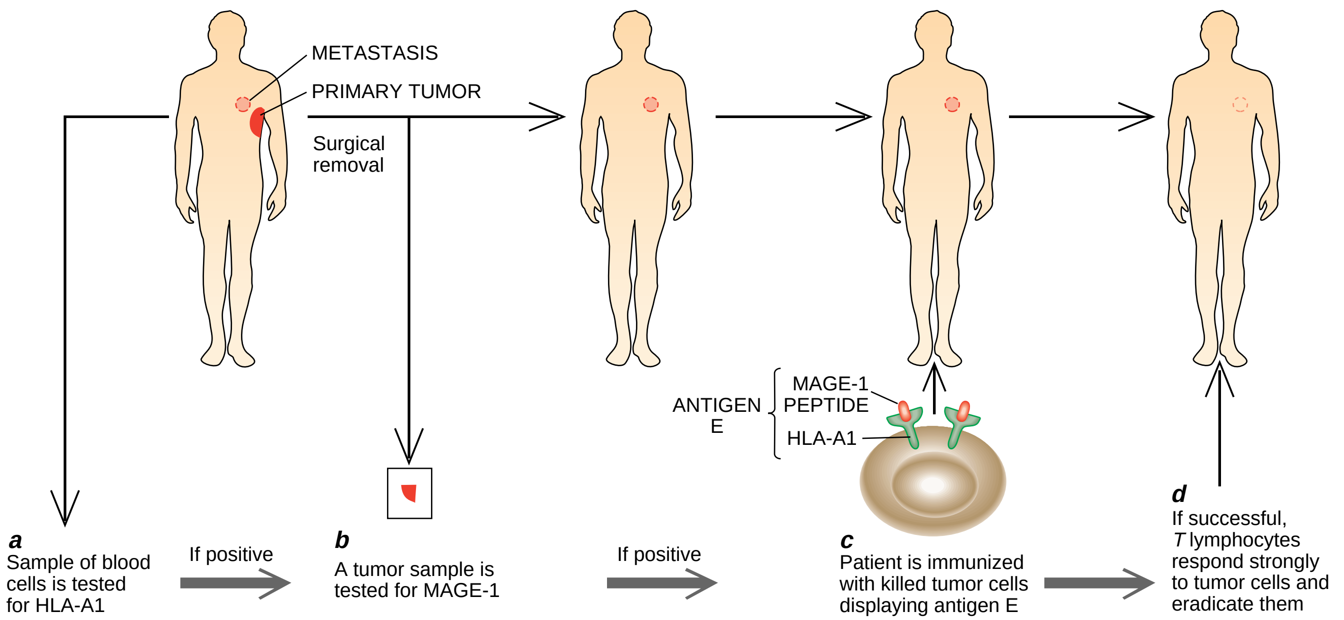
As a first step, we isolated from the patient's white blood cells cytolytic *T* lymphocytes that reacted to the MZ2-MEL cells. Like several other groups working with other tumors, we managed to garner such lymphocytes by culturing the patient's white cells with killed cells from her tumor. Although the original tumor failed to induce rejection in the body, culturing the cells for a few weeks enabled us to isolate cy-




**MUTATION OR ACTIVATION OF CELLULAR GENES can cause cells that do not display antigens recognized by cytolytic *T* lymphocytes (*left column*) to produce antigens *(right column)* that can be recognized by *T* cells.**

## Scheme for Specific Immunotherapy

One immunotherapy now being considered is based on the discovery that cytolytic *T* lymphocytes isolated from some cancer patients can be induced to react to a molecular complex called antigen E. Antigen E is formed by a specific MHC molecule (called HLA-A1) and a peptide derived from a protein called MAGE-1. Melanoma patients whose cells produce the HLA-A1 molecule (*a*) and whose tumors additionally produce the MAGE-1 protein (*b*) will be injected with killed cells displaying antigen E (*c*). If all goes well (*d*), *T* lymphocytes specific for antigen E will proliferate markedly and eradicate tumors. The diagram at the far right represents screening results obtained by the polymerase chain reaction,




tolytic *T* lymphocytes that selectively lysed the tumor cells. From this potentially mixed population of antitumor lymphocytes, we generated clonal populations that were each reactive to a single antigen.

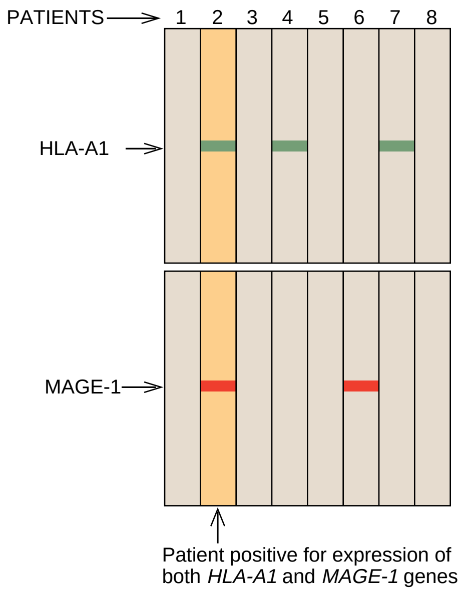
We also needed an antigen-loss variant that could serve as the recipient for DNA from MZ2-MEL cells. This time we obtained the variant by exposing several million MZ2-MEL cells to a similar number of lymphocytes from one cytolytic *T* cell clone—called the anti-E clone because its target antigen was named (arbitrarily) "E." Most of the tumor cells died, but about one in a million lived. These survivors turned out to have lost antigen E. The antigen-loss variants proved sensitive to other *T* cell clones directed against MZ2-MEL cells. Eventually this finding led to the discovery that the MZ2-MEL tumor displays at least four distinct tumor-rejection antigens.

So far we have isolated only the gene that gives rise to antigen E. As might be expected from the *P1A* work, we did so by inserting plasmids carrying the DNA of MZ2-MEL cells into cells of a variant that had lost antigen E. Then we withdrew the gene from one of the few antigen-loss cells that activated the anti-E lymphocyte clone. We named the gene *MAGE-1,* for *m*elanoma *a*nti*ge*n-*1*.

As soon as we knew the nucleotide sequence of this gene, we rushed to determine whether normal cells of the patient carried the sequence. They did, but the gene was not expressed. Here again a tumor-rejection antigen had arisen through the activation of a gene that is silent in normal cells. This finding intimated that, in analogy with *P1A,* the gene might be active in tumors of other patients as well. Indeed, analyses of a large selection of tumor samples suggest that more than 30 percent of melanomas carry an active form of the *MAGE-1* gene. More than 15 percent of breast and lung tumors also express the gene. We have not yet discerned how the MAGE-1 protein promotes tumor progression.

Do these figures mean that all patients who express the *MAGE-1* gene also display antigen E on tumor cells? The answer is no, for reasons that have to do with how antigens form. Recall that the *T* cell receptor actually recognizes not a solitary peptide but a complex consisting of a peptide and the surrounding region of the class I MHC molecule. Now, human class I molecules are encoded by three genes (named *HLA-A, -B* and *-C*), and these genes are polymorphic; that is, they can differ from one person to another. Each gene, in fact, comes in 10 to 40 different forms, called alleles. Because a person inherits one set of A, B and C alleles from the mother and another set from the father, an individual can manufacture six different varieties of HLA proteins—such as HLA-A1, -A10, -B7, -B24, -C4 and -C6—all of which might differ from the six varieties produced by someone else. The protein products of the alleles differ from one another in the shape of the peptide-binding groove and of the surrounding region. Consequently, in any given cell, a peptide typically binds to only one of the available class I molecules, if it binds at all. Hence, only patients who produce the MAGE-1 protein and a particular HLA molecule will display antigen E. We now know the MHC component of antigen E is HLA-A1. We have also found that the MAGE-1 peptide that binds to this HLA molecule is nine amino acids long, and we know its sequence.

Might patients who lack HLA-A1 but produce the MAGE-1 protein also display antigens that can be recognized by *T* lymphocytes? At this point, we do not know. In theory, such antigens a test that can detect expression of the genes giving rise to the HLA-A1 and MAGE-1 proteins. Of eight patients tested, three expressed the *HLA-A1* gene, and two bore tumors that express the *MAGE-1* gene. Only one individual (patient 2) had positive test results in both categories.




could be created if peptides belonging to the MAGE-1 protein were capable of binding to HLA molecules other than HLA-A1. But we cannot be certain that such antigens exist until we identify cytolytic *T* lymphocytes that react to them. So far we have been unable to obtain such lymphocytes. The *T* cells that recognize antigen E would not respond to those antigens because they bind only to the specific shape formed by the peptide in antigen E and the part of the HLA-A1 molecule that surrounds it.

The identification of the gene coding for a human tumor-rejection antigen opens a new phase in the search for an effective specific immunotherapy for cancer. For the first time, we can select as candidates for therapy those patients who have a chance of benefiting from immunization. We can be selective because it is possible to readily identify individual patients whose tumors carry the known antigen. Further, having the gene for a tumor-rejection antigen means we can devise many innovative ways to immunize patients. Finally, we also have the opportunity to determine rapidly whether the immune system is responding to our interventions, because we can measure changes in the number of a patient's cytolytic *T* lymphocytes instead of waiting until clinical effects become apparent (such as the absence of relapse).

We are now initiating clinical studies designed to immunize melanoma patients against antigen E. In these initial studies, we will concentrate on evaluating the cytolytic *T* cell response to the antigen. If we find reliable ways to elicit a good response, later trials will examine cancer remission.

Our methods of identifying candidates for therapy are simpler than might be imagined. We just need to know that their tumors express both the *HLA-A1* and the *MAGE-1* genes. Patients who are about to undergo surgery to remove a tumor can be tested for their HLA type in a couple of ways. One of these methods, based on a small sample of blood, yields results in a few hours. In individuals who test positive for HLA-A1, a sample of tumor can be frozen immediately after surgery. Within two days, a sophisticated technique called the polymerase chain reaction will reveal whether the tumors also express the *MAGE-1* gene [see "The Unusual Origin of the Polymerase Chain Reaction," by Kary B. Mullis; SCIENTIFIC AMERICAN, April 1990]. About 26 percent of white individuals and 17 percent of black individuals carry the *HLA-A1* allele. Considering that some 30 percent of melanoma patients express the *MAGE-1* gene, we can predict that roughly 8 percent of melanoma patients will display antigen E on their tumor cells.

A number of immunization modes can be tested on candidates who fit our dual criteria. Because the *MAGE-1* gene and the antigenic peptide have been identified, we can induce various cell types to express antigen E. Killed versions of the cells can be injected into patients to spur their anti-E lymphocytes into action. Our first clinical studies will follow such a protocol.

We also hope to evaluate the effectiveness of inserting a gene for an interleukin, such as interleukin-2, into cells expressing antigen E. The interleukin should facilitate the activation of *T* lymphocytes around these cells. Synthetic E peptides or purified MAGE-1 proteins that have been mixed with an immune stimulatory substance called an adjuvant will also be tried. Finally, we might insert the *MAGE-1* gene into the DNA of a harmless virus that can penetrate into human cells but cannot reproduce there. After such recombinant viruses are administered to patients, a relatively small number of cells should become infected. These cells should produce the MAGE-1 protein and display antigen E for a while. Immunization with peptides, proteins and recombinant viruses has already proved quite effective for other purposes.

I do not know whether these treatments will cure patients, but I believe there is a good chance that some form of specific immunotherapy will be helpful. My associates and I are encouraged by mouse studies in which strong antitumor responses have been obtained without hurting the general health of the animals. But it is difficult to predict whether the specific immunotherapies I have described will eradicate human cancers, particularly in patients who harbor large tumors. Malignant cells that have lost the ability to produce the MAGE-1 or HLA-A1 protein may arise. Such cells would no longer make antigen E and would thus escape notice of the anti-E lymphocytes. Success, then, may have to wait until we can immunize cancer patients with several tumor-rejection antigens simultaneously. These multiple immunizations should strengthen the immune reaction and also help to prevent variants that have lost one antigen from escaping destruction.

We are confident that the gene-cloning techniques we have developed will lead in the near future to identification of additional genes specifying tumor-rejection antigens. The advances will make it feasible to attack tumors through several antigens. And they will render increasing numbers of patients eligible for trials of specific immunotherapies. Thus, even though success is by no means assured and the work ahead remains considerable, a clear strategy has now been mapped out for the specific immunotherapy of cancer.


FURTHER READING

THE EPITOPES OF INFLUENZA NUCLEOPROTEIN RECOGNIZED BY CYTOTOXIC T LYMPHOCYTES CAN BE DEFINED WITH SHORT SYNTHETIC PEPTIDES. A. R. M. Townsend, J. Rothbard, F. M. Gotch, G. Bahadur, D. Wraith and A. J. McMichael in *Cell,* Vol. 44, No. 6, pages 959–968; March 28, 1986.

A GENE ENCODING AN ANTIGEN RECOGNIZED BY CYTOLYTIC T LYMPHOCYTES ON A HUMAN MELANOMA. P. van der Bruggen, C. Traversari, P. Chomez, C. Lurquin, E. De Plaen, B. Van den Eynde, A. Knuth and T. Boon in *Science,* Vol. 254, pages 1643–1647; December 13, 1991.

A MOLECULAR MODEL OF MHC CLASS-I-RESTRICTED ANTIGEN PROCESSING. John J. Monaco in *Immunology Today,* Vol. 13, No. 5, pages 173–179; May 1992.

TOWARD A GENETIC ANALYSIS OF TUMOR-REJECTION ANTIGENS. Thierry Boon in *Advances in Cancer Research,* Vol. 58, pages 179–210; 1992.

# Flat-Panel Displays

*Recent advances in microelectronics and liquid crystals make possible video screens that can be hung on a wall or worn on a wrist*

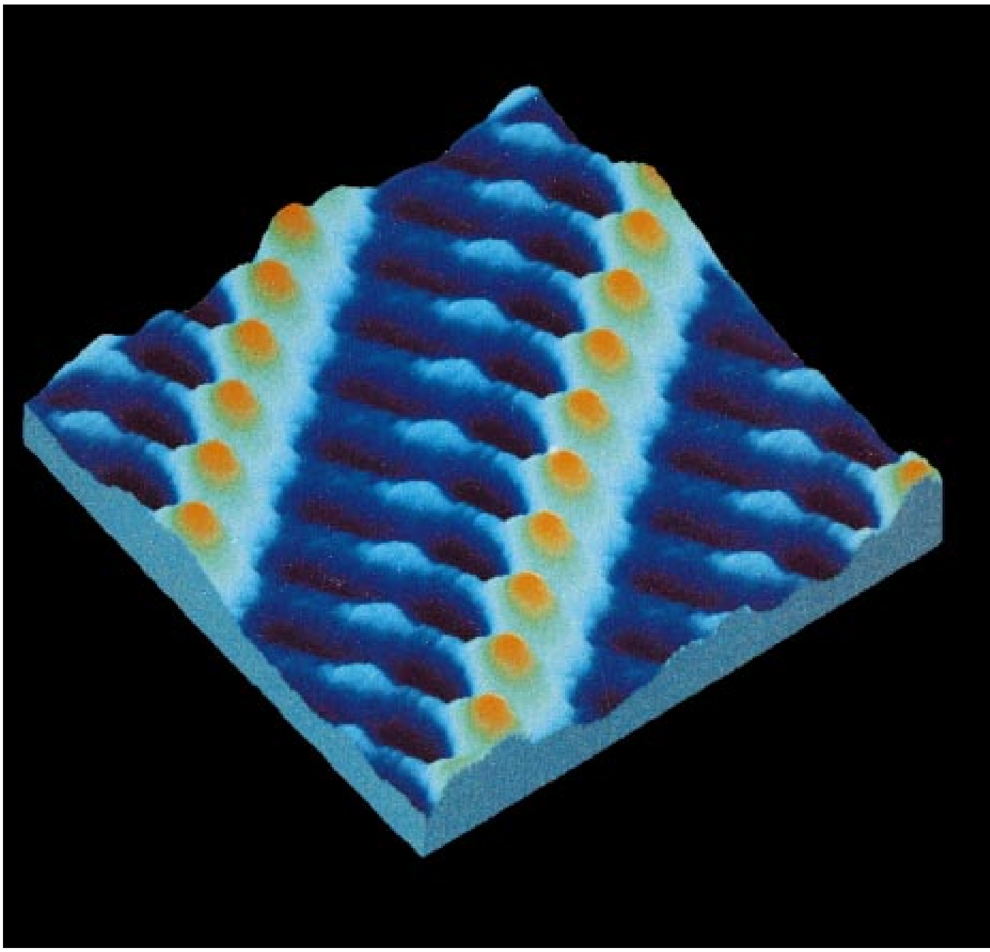
by Steven W. Depp and Webster E. Howard

From the family television set to the computer terminal, the electronic display has become an indispensable way to deliver information. No other medium offers its speed, versatility and interactivity. These attributes are being used to create a wide variety of products that can provide information in any combination of text, graphics, still images or video. Further evolution of this technology will depend, to a great extent, on advances in flat-panel displays.

Although the conventional cathode-ray tube (CRT) remains the dominant display, it has been difficult to modify this technology into a form that is portable, sparing in its use of power and yet capable of producing a superior image. Attempts to squash the CRT into a flat panel have led either to inferior picture quality or to complex designs with excessive manufacturing costs. The latest such attempt replaces the scanning electron gun with an array of tiny electron emitters fabricated on one plate and capped by a second, with the plates kept apart by spacers. But to support the vacuum between the plates, the spacers must be large or numerous; in either case, they tend to obscure the displayed image. Engineers are still seeking a satisfactory and manufacturable solution.

By their nature, flat-panel displays occupy small volume, weigh little and require modest amounts of power. Some can even be written on, like paper. Once a flat panel can be carried about or even worn on one's wrist, like a watch, an individual user will be able to have access to any information, in any place, at any time. Such displays are now at hand. Some airlines use them to provide passengers with armrest movies; in Japan,

STEVEN W. DEPP and WEBSTER E. HOWARD study display technologies at the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y. Depp is department group manager of entry system technologies. He earned a doctorate in physics from the University of Illinois in 1972 and then worked for five years at the Los Alamos Scientific Laboratory. In 1977 he joined IBM's research laboratory in San Jose, Calif., where he specialized in display technologies. In 1982 he moved to Yorktown Heights. Howard manages research on flat-panel display technologies. He earned a doctorate in physics at Harvard University and joined IBM in 1961. He has worked at the Watson Center ever since, conducting research on semiconductor physics, thin-film electroluminescence and active-matrix liquid-crystal displays.



**LIQUID-CRYSTAL MATERIAL is sealed between two glass plates, one bearing transistors to control the electrodes of each cell, the other bearing color filters and an electrode to complete all circuits. Polarizers in the front and rear complete the array, which is illuminated from behind. Liquid-crystal molecules, fixed to a substrate, appear in the scanning tunneling micrograph above.**

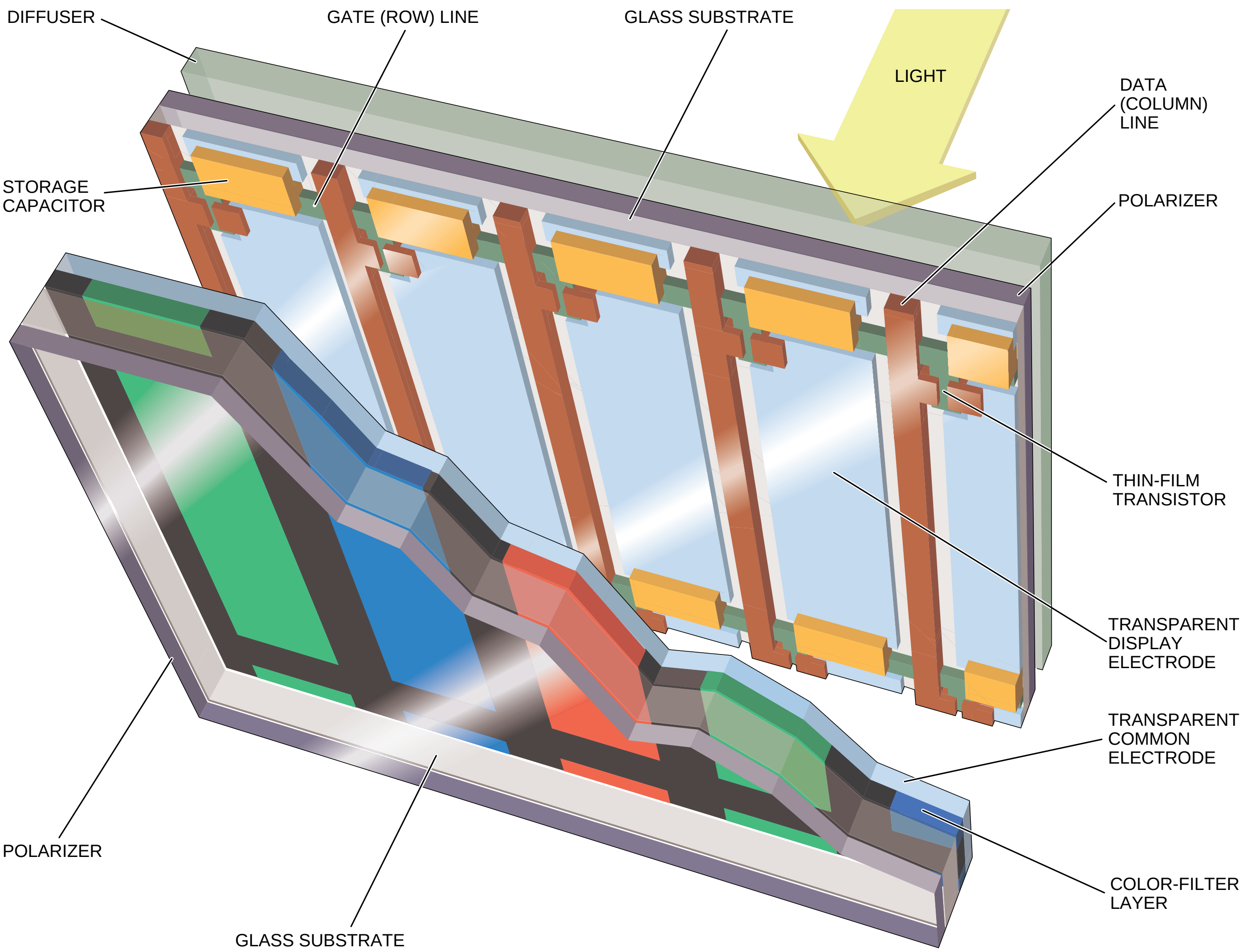
hotels use them in elevators to advertise their restaurants and shops. Within 10 years, high-definition television and dashboard-mounted navigation systems will be commonplace. Flat displays will also spur the development of entirely new products, such as pocket-carried notepads that can store all the memos one might ever write and search them for information keyed to words or dates.

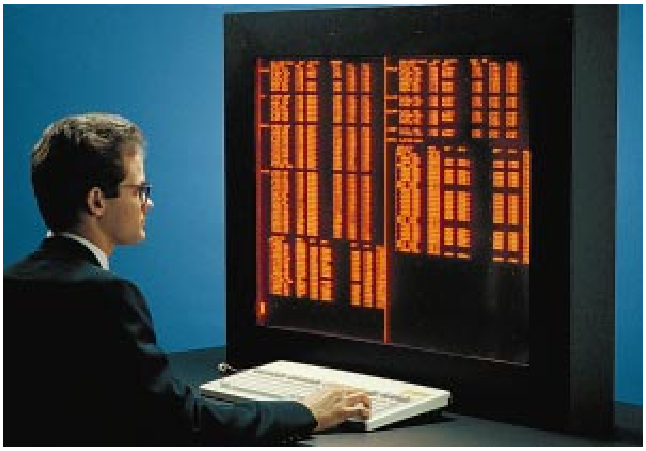
The ideal flat display would portray images with good brightness, sharp contrast, high resolution, quick response, many shades of gray and all the colors of the rainbow. It would also be rugged, long-lasting and inexpensive. The three technologies that have achieved some market acceptance—plasma panels, electroluminescent displays and liquid-crystal displays—all fall short in one or another of these respects, but a fourth just reaching market provides an image that rivals and in some ways surpasses that of the CRT. It is the active-matrix liquid-crystal display (LCD), which currently garners most of the research and development resources and manufacturing investment.

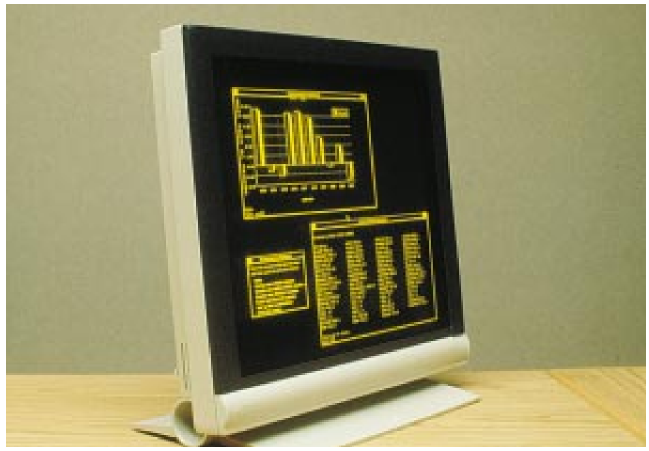
When workers first began to develop flat-panel displays, no one had conceived of active-matrix liquid-crystal devices, and many approaches for making or modulating light in controlled patterns were tried. The first technology to be used in flat panels with high information content employed plasma, or gas discharges. It appeared in the late 1960s.

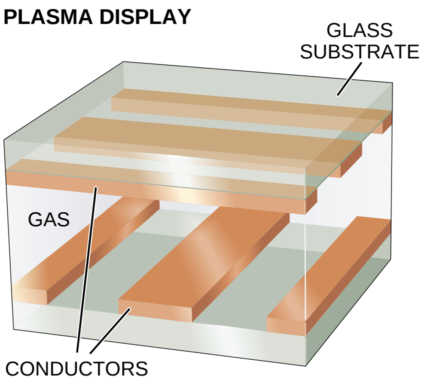
A plasma display begins with two sheets of glass fabricated with a set of parallel ribbons of conductive film. The sheets are placed so that the sets cross. The sheets also enclose a small space filled with a mixture of gases that generally includes neon [*see illustration on next page*]. At any intersection point, a sufficiently large voltage will cause the gas to break down into a plasma of electrons and ions, which glows as it is excited by the current. In effect, one has an array of miniature neon lamps that provide their own light, thus constituting an "emissive" display.

Because the gas ionizes at a well-defined voltage, it is easy to control which intersections light up and which do not. One merely applies a voltage equal to half the firing threshold to a given row and also to its matching column, ensuring that only the crossover point has enough voltage to discharge. By scanning sequentially, a row at a time, and repeating the process at least 60 times per second, one can trick the eye into perceiving a steady image. The image will necessarily be somewhat dim because a given point cannot glow more than a small fraction of the time. Despite this drawback, the design has found some application in portable computers and a handful of other products.

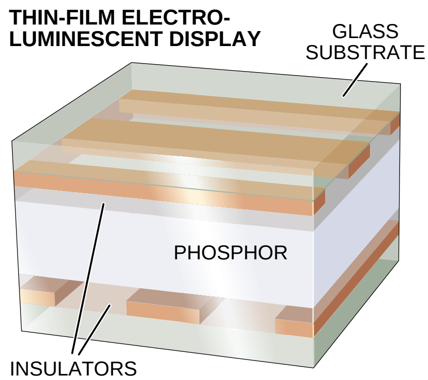









**EMISSIVE DISPLAYS glow with their own light by energizing a plasma or a phosphor. Plasma displays (*left*) consist of an array of miniature neon lamps that are discharged by the combination of one voltage from the cell's row and another from its column. Alternating-current designs (*not shown*) use insulated electrodes to retain a charge, creating a memory effect. Thin-film electroluminescence displays (*right*) replace the gas with a solid phosphor film. Emissive displays are rugged but consume too much power for some portable applications and cannot easily provide full color.**




An elaboration of this approach achieves a brighter image by employing alternating current. Its inventors, H. Gene Slottow and Donald L. Bitzer of the University of Illinois, found that alternating current would produce an inherent memory effect because a cell that has just fired will briefly retain some charge on its insulated electrodes. When the voltage reverses, this charge will add to any new voltage that is applied and will trigger the discharge again. Because the picture elements (pixels) fire each time the voltage reverses, they emit light for a much larger proportion of the time. Originally, interest in alternating-current plasma centered on its memory, but with today's inexpensive semiconductor memory, the display's main advantages are brightness, sturdiness and reliability. It is particularly prized in military applications.

Most plasma displays are orange. It is not easy to modify the design to produce full color, a feature that consumers have come to expect (more than 80 percent of the computer monitors now sold provide full-color displays). For color, three sources—one emitting red light, the others, blue or green—are required. But it would hardly be practical to achieve these colors by filling each pixel with a different gas. Engineers have therefore emulated the fluorescent tube by getting their light in one form and using phosphorescent coatings, or phosphors, to convert it into several others. One gas, common to the entire array, emits ultraviolet light. This invisible radiation then causes each pixel to glow red, blue or green, depending on its phosphor.

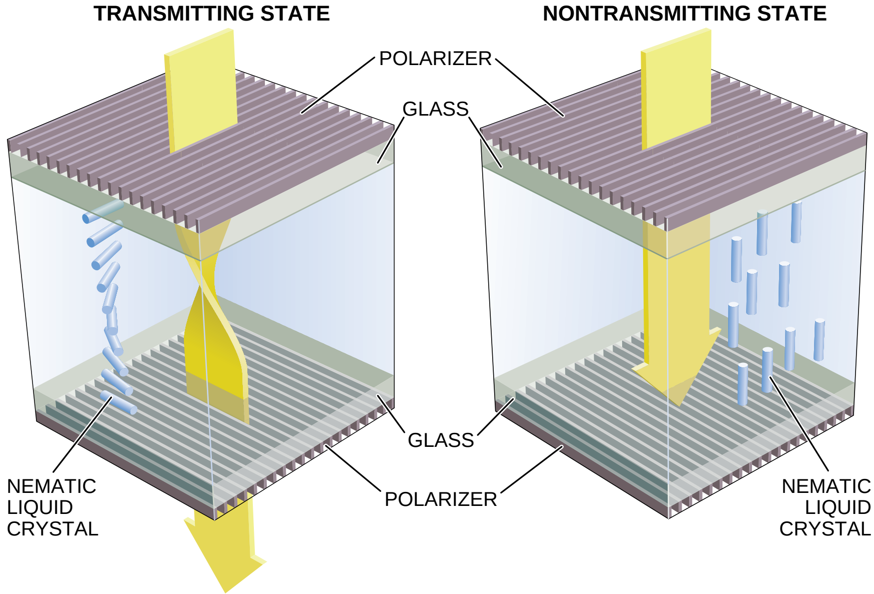
Because plasma displays consume a great deal of power, they appear mainly in applications in which energy efficiency and portability are not required. But such displays can be fashioned into panels measuring as much as 1.5 meters along the diagonal, enough to make a high-definition television screen that might hang on a wall. If color plasma panels can achieve reasonable cost and high reliability, they may be used in such screens.

Another important emissive flat-panel technology is thin-film electroluminescence. Engineers originally hoped to use electroluminescence as a sophisticated source of home lighting. The technology proved to be not nearly efficient enough, however, and the devices ended up in alphanumeric displays, competing with plasma displays.

Like gas-discharge devices, an electroluminescent display resembles a sandwich. A light emitter, or phosphor—typically zinc sulfide doped with manganese—is placed between insulating layers that bear orthogonal electrodes [*see illustration above*]. When the voltage exceeds a well-defined threshold, the emitter breaks down and conducts current. The current then excites the manganese ions, which give off a yellow light, analogous to the glowing neon of the plasma display.

Although such displays are also quite durable, they suffer from two drawbacks: they cannot provide full color, and they consume almost as much energy as do plasma displays. Full color requires a good blue-emitting phosphor, but none having suitable brightness, efficiency and longevity has yet been found. Full color also requires a smoothly varying gray scale, and this has been difficult to control. In addition, the efficiency of such panels declines as the number of picture elements increases. Each element functions as a capacitor, and significant amounts of energy are needed simply to charge and discharge them all. This problem militates against using high-information-content electroluminescent panels in battery-powered applications.

The primary nonemissive technology makes use of the electro-optic properties of a class of organic molecules known as liquid crystals (LCs). As the name implies, liquid crystals are com-




**LIGHT-TWISTING SHUTTER is used in most nonemissive flat-panel displays. Threadlike, or nematic, liquid-crystal molecules are arranged so as to rotate the polarization of light. Nonenergized pixels enable the light to pass a second polarizer (*left*). When a voltage is applied, the twist is disrupted, and the light is blocked (*right*).**

pounds that flow like liquids but have a crystalline order in the arrangement of their molecules. This phase of matter seems improbable, but it is not so rare as one might think. It has been estimated that an organic chemist randomly synthesizing compounds would produce molecules with liquid crystalline properties about once in every 1,000 experiments. These molecules have been known for about 100 years. They appear in many forms, from cell membranes to soap scum.

The most important of the LCs are the nematic, or threadlike, compounds. Their rod-shaped molecules are free to move with respect to one another, yet subtle intermolecular forces tend to keep their long axes aligned. The effect somewhat resembles the schooling of fish. One can set the direction of the molecules' alignment by placing them in an electric field or close to a specially prepared surface. By controlling the alignment, one effectively controls the optical properties as well—above all, the ability of the material to affect the transmission of light.

In one popular configuration, a liquid-crystal compound is sandwiched between glass sheets having different patterns of molecular alignment [*see illustration above*]. To construct such a "twisted nematic" LC, fabricators coat the inner face of each sheet with a transparent, electrically conductive film such as indium tin oxide. They then add a thin layer of an organic polymer and finally brush the polymer in a desired direction. Brushing aligns the polymer by combing its chains or making tiny scratches, or both. The surface effects communicate themselves to the adjacent liquid-crystal molecules through a mechanism that remains obscure. The twist is achieved by placing the glass sheets so that their preferred orientations run at right angles to one another.

If polarized light is directed through the liquid-crystal cell, the direction of polarization will tend to follow the twist. It therefore emerges from the sandwich rotated by 90 degrees. Crossed polarizers on either side of the cell will then pass light rather than blocking it, as would be expected if the cell were not present. This is the "on," or transmitting, state. The "off," or nontransmitting, state occurs when an electric field is applied via the two transparent conductors. The field arranges the liquid-crystal molecules with their long axes parallel to the field lines. This realignment removes the twisted configuration, and thus the cell no longer rotates the incoming polarization. The cell then has no significant optical effect, and the crossed polarizers block the light as usual. In an actual cell, the transmission varies smoothly with voltage.

Thus far we have described a single pixel. A complete liquid-crystal display incorporates many of them by putting the row electrodes on one glass substrate and the column electrodes on the other; each crossover point defines the address of a pixel. Because the light comes from a reflector or light source, the matrix operates as an array of tiny, electronically controlled shutters. The simplicity of design and the modest voltage and power requirements of this "passive matrix" LCD have made it the best-selling flat-panel technology today.

Yet the simplicity of this approach places an inflexible limit on its performance. As Paul M. Alt and Peter Pleshko of IBM have demonstrated, one can improve resolution only by sacrificing contrast. The fault lies in cross talk, an inherent feature of the system by which pixels are energized. Here is how it happens. A voltage is applied to a single row, and the column voltages are adjusted to produce a large combined voltage across the selected pixels in that row. Unselected pixels receive a smaller total voltage [*see illustration on page 96*].

The addressing continues as the next row down is energized, a process that sweeps from top to bottom in about one sixtieth of a second, then repeats for the next frame. During the depiction of a frame of video, each selected pixel receives a large voltage pulse when its row is activated; unselected pixels receive moderate voltages, and all pixels receive a series of smaller cross-talk pulses when the remaining rows are addressed.

Because cross talk rises with the number of rows, the difference in effective voltage between selected and unselected pixels declines as the size of the array increases. For arrays having 240 rows—half as many as a typical television screen—the difference comes to only 6.7 percent. For a nematic cell twisted through 90 degrees, an acceptable contrast would require a difference in voltage of at least 50 percent.

There are three strategies for getting around the trade-off between contrast and resolution. First, one can try to make the transmission-voltage curve steeper, so that small differences in voltage yield large changes in transmission. "Supertwisted nematic" LCDs do this by twisting the liquid crystal through 180 degrees or more. Second, one can use liquid crystals that exhibit a memory effect, enabling one to address many rows without sacrificing contrast. Ferroelectric liquid crystals have this property, and they have been assembled into displays of 1,000 rows and more. But they are somewhat slow. Moreover, because they have only two stable transmission states, they cannot easily reproduce various shades of gray. These drawbacks render the passive-matrix approach less suitable for the display of realistic images and video.

The third and most radical approach separates the functions of addressing and transmitting, so that each may be optimized. Such "active matrix" addressing employs an array of transistors, each of which activates a single pixel. Each pixel receives a voltage from its column line only when its own transistor is switched on. The other rows can be addressed when the transistor switches off; meanwhile the pixel holds the voltage it had been given. Because this method effectively isolates the pixel from cross talk, the number of addressable rows can be extremely large. In addition, the technique can easily provide color. Pixels and their respective transistors are simply grouped into triads in which each pixel has a filter for one of the three primary colors [*see illustration on page 91*].

The entire display is rewritten after all the rows have been addressed. Such refreshing, as it is called, prevents the distortions that might otherwise appear as charge gradually leaks from the liquid-crystal cells, reducing their voltages and changing their transmission of light. More important, by refreshing every sixtieth of a second, the pixels can track video and other rapid screen changes.

Active-matrix displays closely resemble dynamic random-access memory (DRAM) chips. Both are complex integrated circuits that store charge in a million or so discrete locations, each one under the control of a single transistor. Yet a computer reads a DRAM chip a row at a time by sensing the charge at each location, whereas the human eye in effect reads the entire display in parallel.

Another difference is that DRAM stores digital data; active-matrix displays store analog data. In the former, each cell must be either mostly "on" or "off"; in the latter, cell voltages must vary along a gray-scale continuum. This feature requires the display to "encode" far more information than a DRAM chip of similar size can manage.

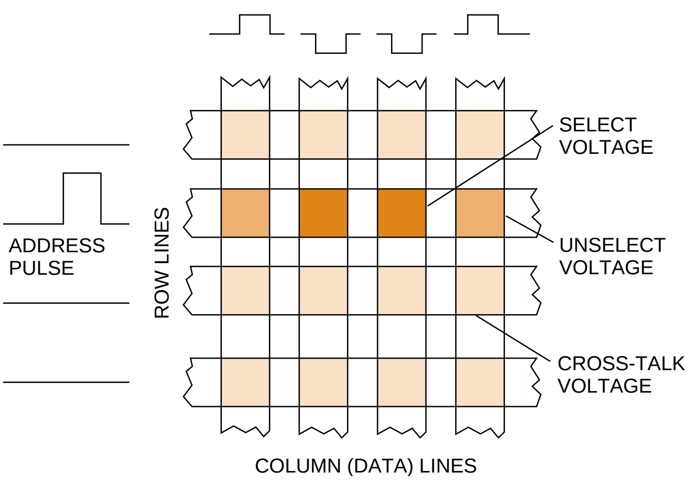
It has been possible to meet this constraint by careful design and controlled manufacturing tolerances. Fortunately, these tolerances need apply only on a small scale, from pixel to pixel. The reason is that although the eye is sensitive to the local variations produced at the edges of objects, it hardly notices the variations that build gradually as it scans from one side of the screen to the other.

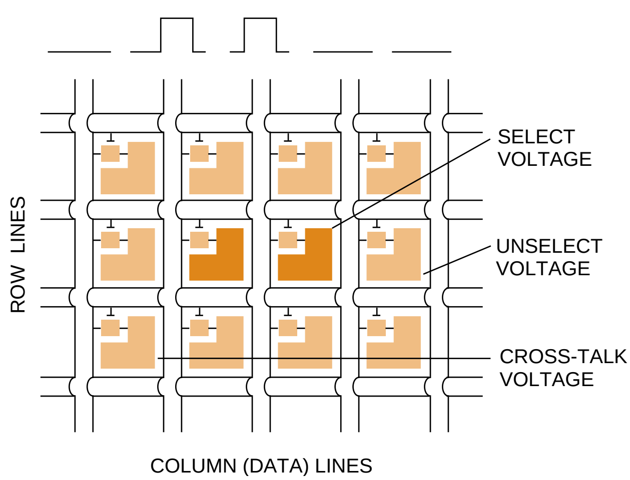
An active-matrix display is controlled by thin-film transistors (TFTs), a technology that has come into its own only in the past decade. Such transistors derive from solar-cell manufacturing technology and can be made in large arrays at comparatively low cost. The only drawback—a lower current-carrying capacity—hardly matters in display applications because very little current is needed to control a liquid crystal.

A thin-film transistor uses deposited layers of different materials to constitute the semiconductor, the insulators and the electrodes. As in conventional transistors, two terminals conduct current, and the third switches the transistor on and off. But conventional transistors form most of these parts on the surface of a single, semiconducting crystal, whose electrical properties are modified in particular regions by the addition of charge-donating atoms, called dopants. The TFT can be fabricated on virtually any surface, including inexpensive glass.

It took about 20 years to turn the idea of the TFT into a commercial product. When Paul K. Weimer of RCA invented the technology in 1962, experts expected that it would find application in conventional electronic circuits. But other electronic devices quickly superseded it, discouraging most researchers from working on TFT technology. Only in 1974 did T. Peter Brody and his colleagues at Westinghouse demonstrate that TFTs might be used as switches for liquid-crystal displays. Even that application proved difficult at the time because the available materials and processes could not produce large arrays of stable, defect-free devices.

Researchers advanced by trying a variety of semiconductors. Polycrystalline silicon became the first to find its way into a commercial product in 1984, when Seiko-Epson used it in a pocket color television set with a remarkably attractive two-inch screen. But the display had to be fabricated with expensive integrated-circuit processes and high-temperature materials. Already it was apparent that the future lay with another material, amorphous silicon. In 1979 P. G. LeComber and his colleagues




**PASSIVE VERSUS ACTIVE: passive liquid-crystal matrices (*left*) give all pixels in a given row an address pulse, then add or subtract a column voltage. Unselected cells thus accumulate an average "cross-talk voltage" that rises with the number of rows, producing a trade-off between contrast and resolution. Active matrices (*right*) avoid this problem by using an array of thin-film transistors to address the liquid crystals. While the transistors are switched on, the associated liquid-crystal cells are charged to the voltages on the column lines, also known as data lines. When the transistors are turned off, the voltages are held constant. In this way, an image can be written row by row and maintained until the next "refresh cycle."**

at the University of Dundee had the insight to see that amorphous silicon, already being developed for inexpensive solar cells, could be suitable for TFTs in displays.

There are several key processes for fabricating the TFT array. First, the substrate is formed of glass that has been purified of alkali metals, which might otherwise contaminate the transistors or liquid crystal. The molten glass is poured out, in a finely controlled process developed by Corning, to produce a sheet so flat that, were it scaled to the size of a football field, its thickness would vary by less than the breadth of a pencil.

Later, the semiconductor layer is deposited by a plasma process using silane gas, $SiH_4$, under low pressure. An electrical discharge ionizes the gas and breaks the molecules into fragments, which condense on the glass as a random network of silicon, rich in hydrogen. The hydrogen is crucial because it ties up broken bonds that might otherwise trap electrons and disrupt the semiconductor. Finally, metal electrodes, insulators and other elements in the TFT are deposited in a manner similar to that used in fabricating integrated circuits. Here, however, a much larger area must be covered.

About two dozen companies, mostly in Japan, developed TFT/LCD technology during the mid-1980s. By the end of the decade, IBM/Toshiba, Sharp, Hitachi and others were demonstrating excellent image quality in color displays spanning diagonals of 10 inches or more. For the first time, a flat-panel display could match a cathode-ray tube in contrast, brightness and color saturation (or vividness). Indeed, when the two kinds of display are placed side by side, most observers prefer the crisp, undistorted, jitter-free image of the TFT/LCD.

Although the active-matrix display is meeting the standards of image quality set by the CRT, it costs several times as much to produce. Further refinements in manufacturing technology are required; to develop the necessary experience, manufacturers will concentrate, in the beginning, on the many emerging applications for which a flat display is indispensable.

Among such applications are tiny televisions and videotape recorders, portable—even wearable—computers, dashboard-mounted screens for automobile navigation systems, avionics displays for military, commercial and space applications, as well as electronic books, notebooks and clipboards. Soon the list will include still more exotic products, such as shopping-cart displays and products for the emerging industry of virtual-reality entertainment.

**WALL-HUNG TELEVISION in the 1990 movie *Total Recall* depicts life in the year 2084, yet comparable products should reach the market within a decade. When not presenting live video, such a panel might revert to the pattern of the surrounding wallpaper, display works of art from a digital library or serve as an interface with a computer network.**

By 1995 such applications are expected to form a market of between $4 billion and $5 billion. That scale of production should eventually reduce unit costs enough to make flat-panel displays become competitive with CRTs in television, desktop computing and other traditional markets. At the end of this decade there should be more revenue from flat panels than from CRTs.

We predict that as flat panels continue to evolve toward higher resolution and reduced power and weight, they will increasingly replace paper. In addition to their flexibility, they offer the advantage of responding to touch or stylus. These qualities all tend to simplify the most cumbersome element in modern electronics, the interface, enabling the user to take full advantage of the computer itself.

One such display, with data storage, might replace a filing cabinet full of forms. Entirely new uses will also arise. When one is not using one's screen as a television or as a computer monitor, it may disappear, like a chameleon, into the wallpaper pattern or serve as an art gallery or photograph album. We can easily foresee a time when even this magazine will be an electronic publication viewed on a personal flat-panel display, enlivened by video graphics and backup information available at a touch of the screen.

FURTHER READING

FLAT-PANEL DISPLAYS AND CRTS. Edited by Lawrence E. Tannas, Jr. Van Nostrand Reinhold Company, 1985.

LIQUID CRYSTAL TV DISPLAYS: PRINCIPLES AND APPLICATIONS OF LIQUID CRYSTAL DISPLAYS. E. Kaneko. D. Reidel Publishing Company, 1987.

LIQUID CRYSTALS: APPLICATIONS AND USES, Vol. 1. Edited by Birendra Bahadur. World Scientific, Singapore, 1990.

FLAT PANEL DISPLAYS USING AMORPHOUS AND MONOCRYSTALLINE SEMICONDUCTOR DEVICES. K. Suzuki in *Amorphous and Microcrystalline Semiconductor Devices: Optoelectronic Devices.* Edited by Jerzy Kanicki. Artech House, Boston, 1991.

THIN-FILM-TRANSISTOR/LIQUID CRYSTAL DISPLAY TECHNOLOGY: AN INTRODUCTION. W. E. Howard in *IBM Journal of Research and Development,* Vol. 36, No. 1, pages 3–10; January 1992.

# How Parasitic Wasps Find Their Hosts

*Besides recognizing odors from their caterpillar hosts, wasps also learn to identify compounds released by the plant on which the caterpillars feed*

by James H. Tumlinson, W. Joe Lewis and Louise E. M. Vet

Hide-and-seek is not simply a game for children. Animals "play" it too, and for most the stakes are life or death. For parasitic wasps, the goal is propagation. Females ready to lay eggs must inject them into a host, often the caterpillar of a moth or butterfly. The caterpillar becomes paralyzed, and the eggs inside develop into wasp larvae. The growing larvae feed on the host, thus killing it. The larvae then form cocoons on or near the caterpillar body and pupate. Needless to say, it is in the best interest of the caterpillar to remain hidden under leafy cover. Secrecy would not seem to present much of a problem in a large, lush crop field.

How, then, do wasps detect their victims within the complex environment, particularly when the caterpillars evade their attackers by hiding, moving or otherwise remaining concealed? Random searching alone certainly seems an implausible way to achieve such a literally vital goal; field studies suggest that a wasp might spend its entire lifetime searching for a single host if the wasp had to find it by chance. For the past 10 years, our research groups have been collaborating in pursuit of the answer. We think we have found it. Furthermore, the basic principles we have discovered seem to apply to other kinds of parasitic wasps, including species that parasitize eggs or other insect stages of development.

Briefly the answer is this: the wasps home in on unique volatile compounds that they can detect over substantial distances. These chemicals come from the caterpillar's feces and, perhaps more surprising, from the plants on which the caterpillar feeds. Our quest led us into two areas that had been relatively unexplored with respect to these animals. The first was the ability of wasps to learn. They decipher many indirect and frequently changing cues that indicate the location of hosts. This ability represents a level of sophistication previously unexpected in these insects. The second area is the interaction between the wasps and the plants. When attacked by a caterpillar, a plant produces signals that wasps home in on. Research suggests that investigators can exploit the learning abilities of wasps and the wasp-plant conspiracy to control caterpillar pests in the field.

JAMES H. TUMLINSON, W. JOE LEWIS and LOUISE E. M. VET have been collaborating on studies associated with the foraging behaviors of beneficial parasitic insects for more than a decade. Tumlinson and Lewis received their doctorates from Mississippi State University and work for the Agricultural Research Service, U.S. Department of Agriculture. Tumlinson, who received his Ph.D. in chemistry, is research leader in insect chemistry at the department's branch in Gainesville, Fla. Lewis, who received a doctorate in entomology, is senior research entomologist at the department's Tifton, Ga., location. Vet is associate professor at the Agricultural University in Wageningen in the Netherlands. She received her Ph.D. in animal ecology from the University of Leiden.

Our findings about the complicated interactions among the parasites, hosts and plants began with the assumption that a specific set of odors from the hosts would guide the wasps to their victims. This supposition was rooted in many studies that indicated parasitic wasps respond to certain chemical cues. Early efforts identified and synthesized several kinds of chemicals that can act as guides. Such compounds belong to a class of chemical mediators called kairomones. These substances, which come predominantly from the oral secretions and feces of the caterpillar, act at close range as trail odors or searching stimulants.

These kinds of kairomones, however, could not be the whole story. Because they are nonvolatile, they do not disburse well in the atmosphere. Thus, they cannot be detected over long distances. If wasps had to rely solely on short-range cues from the hosts, their success in finding caterpillars would only be slightly better than chance.

In this regard, S. Bradleigh Vinson and his co-workers at Texas A&M University uncovered a clue. They found that parasitic wasps are also attracted to volatile chemicals. Such compounds can be detected over several meters and could provide the beacon wasps need. Only after work done in one of our laboratories in Wageningen, the Netherlands, clearly demonstrated that wasps can exploit volatile chemicals to find hosts did we have the first indication of the importance of learning in foraging.

To prove it, we used a device called an airflow olfactometer, which was essentially a square container. The olfactometer can provide up to four different odor fields (one from each corner) to a wasp placed in the center. We found that *Leptopilina clavipes* wasps responded to the odor of decaying mushrooms, which are likely to harbor its host, fungus-eating *Drosophila* fruit fly larvae. We also demonstrated that wasps can develop preferences for odors from other sources, such as fermenting fruit, if they have been reared on hosts that have fed on that food and have laid eggs in those hosts.

Although experiments with the airflow olfactometer showed that wasps respond to volatile signals, they did not indicate how the wasps actually used the odors in the field. The wasps did not fly in the olfactometer, and the device was too small to mimic natural conditions realistically.

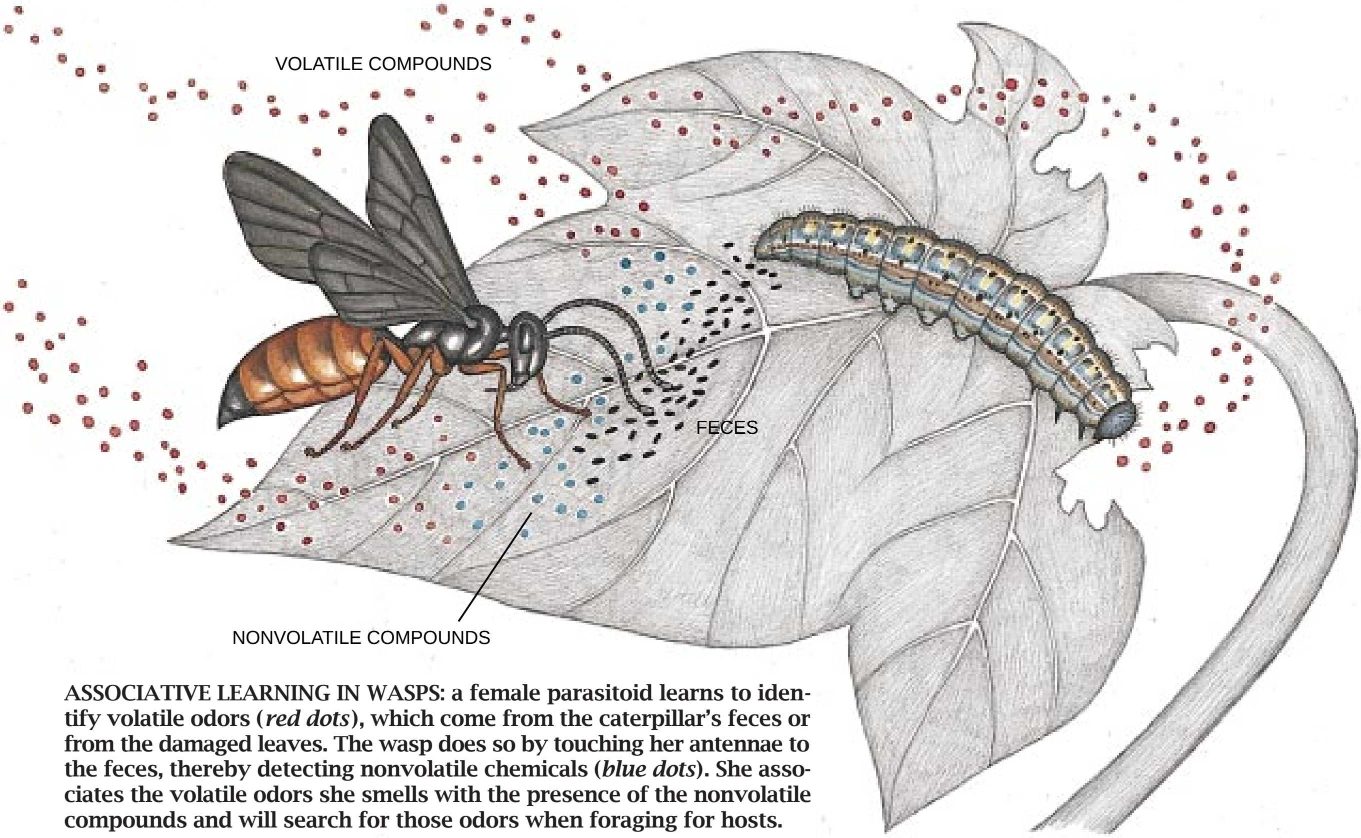
We therefore decided to use wind tunnels to study the responses of flying wasps to volatile chemicals. Such tunnels are essentially long containers that insulate the interior from extraneous odors and air currents. Investigators have used such devices to analyze the responses of moths to pheromones and other attractants. One can also use them to test other insects, such as fruit flies [see "The Amateur Scientist," page 144]. A small fan blows the scent from one end of the tunnel to the other. We placed caterpillar-infested plants at one end and released the wasps at the other. We reasoned that the wasps would find their hosts by flying upwind into an odor plume originating from the host.

Our initial results were disconcerting. Females of the parasitic wasp *Microplitis croceipes* responded poorly to odors of their hosts, corn earworm larvae, which were feeding on leaves of cowpeas. Yvonne Drost and Oliver Zanen, working in our laboratory in Tifton, Ga., discovered that only about 10 percent of laboratory-reared female wasps responded to odors from the larvae-infested plants.

This curious outcome was rooted in the fact that the *M. croceipes* wasps had been reared on hosts maintained on a laboratory diet (generally of ground beans supplemented with vitamins). Even in their larval stages, the wasps had never come in contact with hosts in natural conditions, host products or cowpea plants. Only after we allowed the female wasps to contact the corn earworm larvae did they respond to the odors in the wind tunnel.

This result was a major break-

**FEMALE PARASITIC WASP stings a plant-feeding caterpillar, in the process laying eggs in her victim. By studying how the wasp tracks down her host, researchers believe they can use the wasps to help control caterpillar pests.**




**ASSOCIATIVE LEARNING IN WASPS: a female parasitoid learns to identify volatile odors (*red dots*), which come from the caterpillar's feces or from the damaged leaves. The wasp does so by touching her antennae to the feces, thereby detecting nonvolatile chemicals (*blue dots*). She associates the volatile odors she smells with the presence of the nonvolatile compounds and will search for those odors when foraging for hosts.**

through in our investigations of the foraging behaviors of these wasps. It alerted us to the fact that experience played a more important role in the exploitation of odors than we had realized. Even more surprising, the experience need not include direct physical contact with a host. After touching its antennae to the feces of the caterpillar for about 30 seconds, wasps could later locate the larvae.

Subsequent tests in the wind tunnel showed that wasps are attracted to the hosts' feces. Indeed, Fred J. Eller, working in our laboratory in Gainesville, Fla., found that the feces odor enabled the wasps to determine the plant on which the host has fed. He discovered that females familiar with the feces of hosts fed one plant would fly to the odors of those feces more often than to the odors from hosts feeding on a different plant species.

Exactly what in the feces attracted the wasps? The volatile compounds in them seemed to be the obvious candidate. We thought the wasps might have an innate ability keyed to the smell of the hosts' feces. To confirm that hypothesis, we extracted the volatile compounds from the feces with hexane, a liquid solvent derived from petroleum. In flight-tunnel tests, the volatile extracts attracted the wasps as strongly as did the host-infested plants.

Yet more thorough experimentation soon presented us with a puzzle. We tested females who had never been exposed to caterpillar feces. Rather these wasps had contact only with the volatile extract. To our surprise, these females responded poorly. In fact, they did no better than wasps that had not touched anything.

We decided, then, to examine the remaining, nonvolatile components of the feces. We removed those compounds with water and tested to see whether they would attract the wasps. The females were only slightly attracted to the material extracted by water. Yet after touching these compounds with their antennae, the wasps responded to the volatiles extracted with hexane.

These results, taken together, indicate that a crucial component lies in the water-soluble, nonvolatile compounds. This component is specific to each type of host. The wasps must associate the volatile chemicals with the nonvolatile ones.

To determine conclusively that wasps can learn volatile chemicals by association, we experimented with completely novel odors. For example, we allowed the wasps to smell vanilla extract as they touched the water extract. Sure enough, they subsequently flew to the smell of vanilla; they "thought" that hosts were nearby. This ability to link a wide variety of odors to their hosts greatly broadens the range of chemicals parasitic wasps can use to locate their prey.

Hans Alborn, working in our Gainesville laboratory, showed that the chemicals in the water-soluble extract are not related to diet. He demonstrated that the chemicals that the wasps recognize could be obtained from feces from corn earworm larvae raised on artificial diets. The food could even be filter paper. That the nonvolatile chemical does not depend on the host's food contrasts sharply with the volatile compounds, which are specific to diet. Current studies are trying to identify the nonvolatile component.

Although chemical cues are the primary sources of information that guide parasitic wasps, the shapes, colors and patterns of the area around the caterpillar also play a role. Felix L. Wäckers, working in our Tifton laboratory, studied how female *M. croceipes* wasps use such visual cues in a wind tunnel. Using odor, he taught wasps to associate a caterpillar with various color patches, such as solid orange or black-and-white squares, attached to plant leaves. When the odor was removed, the wasps responded to the color design alone.

Wäckers also showed that the females quickly developed preferences for combinations of odor and visual stimuli. Odor and visual effects proved to be additive—that is, the learned preference for the combinations was stronger than either the odor or visual cue alone.

This sophisticated ability to learn visual and chemical cues and to use them in combination equips the wasps with a powerful system for detecting and tracking their prey. With the acquired information, they can readily locate likely sites and concentrate their search on the most profitable plant parts.

During the wasp's hunt for caterpillars, the plant does not remain a passive observer. In fact, its reaction to an attacking caterpillar may be the most surprising element in this predator-prey system. When chewed on, plants actively produce and release volatile compounds from both damaged and undamaged tissues. This response suggests a defensive mechanism designed to repel invaders: studies have shown that many of the volatiles released by damaged plants are toxic to insects. The wasps, however, are not repelled by these odors. Instead, they exploit the smells to find their hosts.

The first indication that plants enlist the help of natural enemies to fight off herbivorous attackers came in 1988. Marcel Dicke and Maurice W. Sabelis, colleagues of one of us (Vet), and their co-workers studied predatory mites that prey on plant-feeding spider mites. They found that when herbivorous spider mites feed on lima bean leaves, the plant releases a blend of volatiles that attracts predatory mites. The blend is specific for both the plant species and the species of spider mite that feeds on it. The predatory mites can discern these differences and thereby recognize their prey.

The volatile chemicals released by caterpillar-infested plants have proved to be potent attractors of wasps. In fact, recent studies have shown that wasps are more sensitive to the chemicals from plants than they are to those from the host and host feces. Ted C. J. Turlings, working in our Gainesville laboratory on *Cotesia marginiventris,* a parasitoid of moth larvae, found in wind-tunnel tests that corn plants damaged by hosts were the most important source of volatile attractants. Removing the hosts and host products, including feces, only slightly diminished the attractiveness of the plants. Neither hosts nor feces were as attractive as the plants alone.

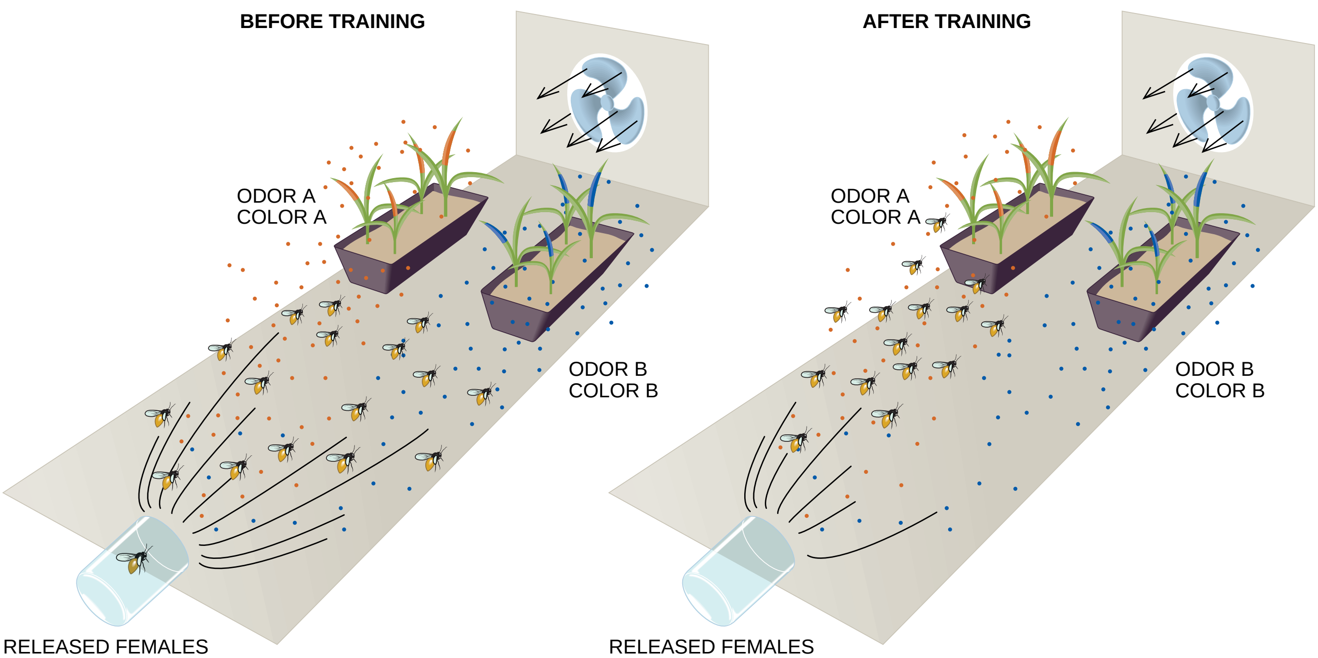
One might suspect that the "green-leafy" odor reminiscent of freshly cut grass attracts the wasps. Yet such signals are intermittent; they emanate only as a caterpillar feeds. Once it stops eating and moves away, the damaged section of the leaf no longer emits the freshly cut scent.

More detailed analysis, however, revealed that plants produce a more consistent set of volatile attractants. These compounds—hydrocarbons called terpenes and sesquiterpenes—are produced several hours after the attack and persist for several hours, perhaps days. The delay in forming the terpenes and sesquiterpenes indicates that the physiology and biochemistry of plants actively change in response to herbivore damage. The mechanisms behind this change are unknown.

Chemicals from the caterpillar are essential if the plant is to release the volatile compounds. When we mimicked caterpillar damage on leaves with a razor blade, the plants did not emit large amounts of the volatile compounds. But if we applied oral secretions from the caterpillar to the wounds, the plants would release terpenes and sesquiterpenes several hours later. Caterpillar regurgitant on undamaged leaves did not induce such emission. Wind-tunnel tests revealed that oral secretions placed on artificially damaged leaves made the plants as attractive to the wasps as the plants with real caterpillar damage.

Turlings also found that the response of the plant to the oral secretions is systemic—that is, the entire plant releases volatile compounds when one or more leaves are attacked by caterpillars. Dicke had earlier found a similar effect, in which intact leaves of a plant injured by a spider mite attracted predatory mites. This systemic effect is significant in that it makes the plant under attack stand out from its neighbors.

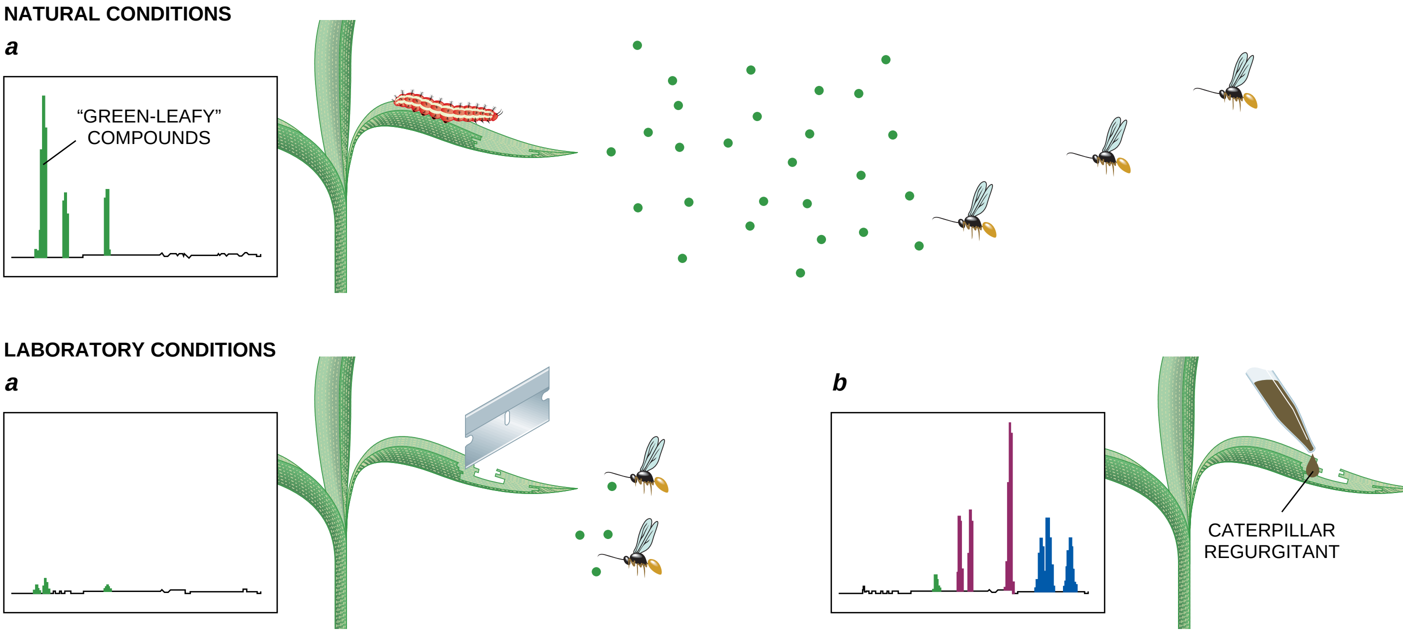
Whether the release of volatiles by the plant has evolved to attract the natural enemies of the herbivores or whether the insect carnivores only exploit a plant defense mechanism aimed at the herbivore is being investigated. The discussion is




**WIND TUNNEL is used to show that female wasps can be trained to respond to a particular odor and color. During the training period (*left*), two sets of plants, each marked with a different scent and color, are presented. Rewards of caterpillars are given to the wasps only if they fly to the set labeled "odor A" and "color A." So trained (*right*), the wasps, when given the same choice, will fly immediately to that set, even if no caterpillars are present.**

## Distress Signals from Plants

Under natural conditions, corn seedlings attacked by beet armyworm caterpillars attract parasitic wasps by releasing various volatile compounds (*upper row*). The chemicals, identified through gas chromatography, are displayed as colored peaks. As a caterpillar feeds, the plant initially releases "green-leafy" chemicals (*a*). Such chemicals, responsible for the odor




akin to debating whether the chicken came before the egg. The current phenomena result from a prolonged evolutionary process of action and reaction. The larvae have evolved to become as inconspicuous as possible to avoid parasitization and predation. They must, however, feed to survive, and in feeding they damage the plant. The damage induces a reaction from the plant. The plant's reaction provides information that facilitates the wasp's attack on the caterpillar. In this process the carnivores exert an important influence on the defensive strategy of the plant.

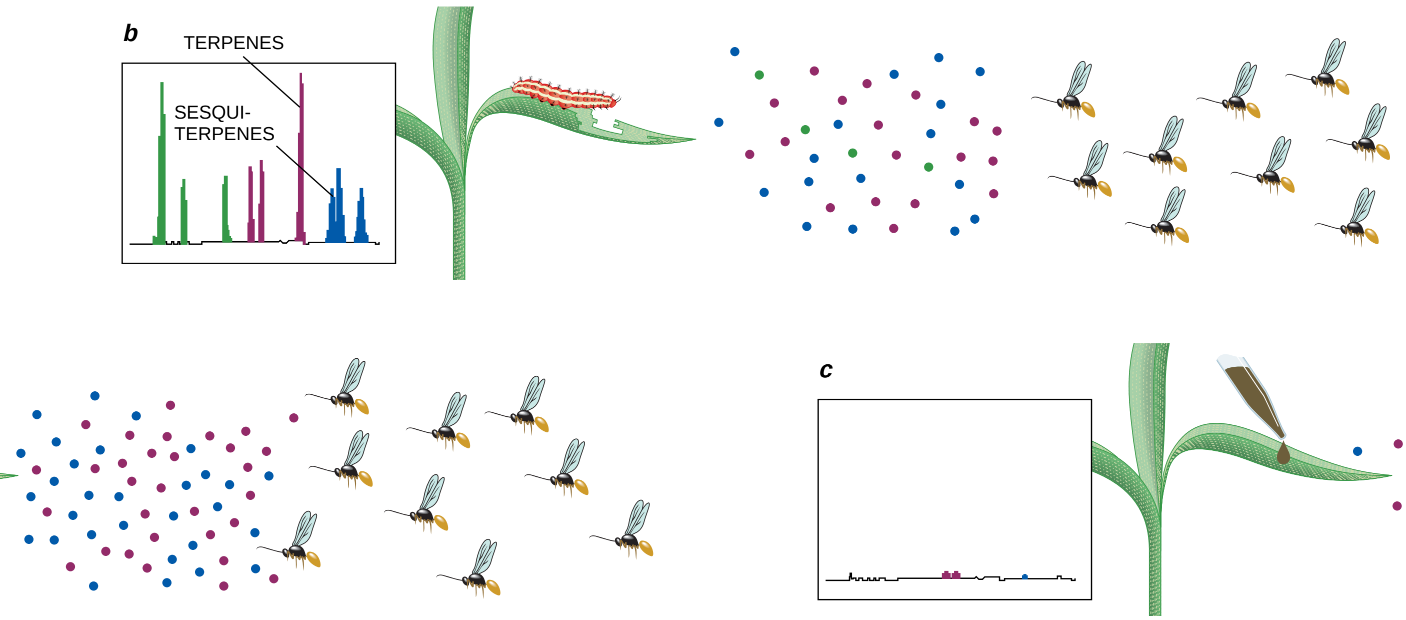
This evolutionary game of hide-and-seek leads to another question. Why do parasitic wasps learn? The answer lies in the incredibly complex collection of chemicals with which the wasps must contend. The signals will vary considerably if the hosts are feeding on different plant species. For parasitoids such as *M. croceipes,* whose victims feed on a wide variety of plant species, the different plants will send out completely different signals. For example, cotton, cowpea and soybean each produce a unique blend of volatile chemicals when eaten by corn earworm caterpillars. Moreover, the composition of the volatile compounds can differ when the hosts feed on different parts of the same plant. The picture is even more complicated if a parasitoid attacks caterpillars of several different species of moths, as does *C. marginiventris.* The same plant may release two different blends when fed on by two different species of caterpillar.

In such a complex, dynamic chemical environment, a simple, rigid search procedure would not permit the parasitoids to exploit the available resources efficiently. They must be able to detect a great variety of potentially important chemical cues and determine their significance in the context of the surrounding environment. They may encounter different host species or growth stages of plants of varying suitabilities in the same location, and they may have to "decide" whether to continue to search in a given area or to abandon it for another. The probability of survival of their progeny depends on making the right choices during foraging. When all the complexities of the environment in which the parasitoids must search are considered, it becomes clear why learning is an essential element of the foraging strategy of the wasp.

That female wasps need experience to learn the odor of their hosts raises a question. How do females newly emerged from their cocoons, which are formed on the outside of or near the caterpillar, find their first hosts? The answer is somewhat speculative, but enough experiments have been performed to enable us to piece together a reasonable conjecture. There appear to be two possibilities.

First, wasps seem to have some innate preference for certain odors. We know from the work with *Drosophila* larval parasitoids that odors from some sources are more attractive to inexperienced female wasps than those from others. In our laboratories, Eller observed that *M. croceipes* females that had had no contact with hosts were nonetheless slightly attracted to volatile chemicals from some plants. Wasps thus seem to have a genetic inclination for odors from substances that have served as food sources for hosts.

The second possibility is that compounds in their cocoons can to some extent train the wasps. Franck Hérard, working in our Tifton laboratory on *Microplitis demolitor,* a species from Australia that attacks corn earworm larvae, found that wasps reared on plant-fed hosts responded well to the odors of host-infested plants in a wind tunnel. The response, however, occurred only if the wasps had been allowed to emerge naturally from their cocoons. In contrast, wasps cut out of their cocoons responded poorly to odors from plant-feeding hosts. Their reaction improved only when they could walk on and touch their antennae to their cocoons. The cocoons undoubtedly contain odors and of freshly cut grass, attract wasps only slightly. But later, the seedling also produces terpenes and sesquiterpenes, which are highly attractive to wasps (*b*). The plant continues to produce such compounds even after the caterpillar has stopped feeding. In the laboratory (*lower row*), simulated caterpillar damage using a razor blade (*a*) produces few compounds. But when oral secretions from the caterpillar are placed on artificially damaged leaves, a plant produces terpenes and sesquiterpenes, though no green-leafy compounds (*b*). Regurgitant on an intact leaf elicits virtually no response (*c*).




other substances from the hosts and their food plants. This initial conditioning provides the "experience" that aids the wasps in locating their first host.

A level of innate ability and contact with the cocoon seem sufficient to prime the wasps to respond to the odors of host-infested plants. In studies, wasps usually had trouble finding their first hosts, often requiring 20 to 30 minutes. Efficiency improved significantly with each success for the first few encounters. The wasps began to uncover subsequent hosts sometimes in less than five minutes.

Because parasitic wasps learn readily and respond efficiently to chemical cues, researchers have begun to explore ways to exploit the wasps' abilities. Specifically, they are trying to determine whether the natural proclivity of parasitic wasps to control caterpillar pests can be significantly enhanced. Appropriately trained, the wasps could "police" plants and protect them from crop-feeding caterpillars.

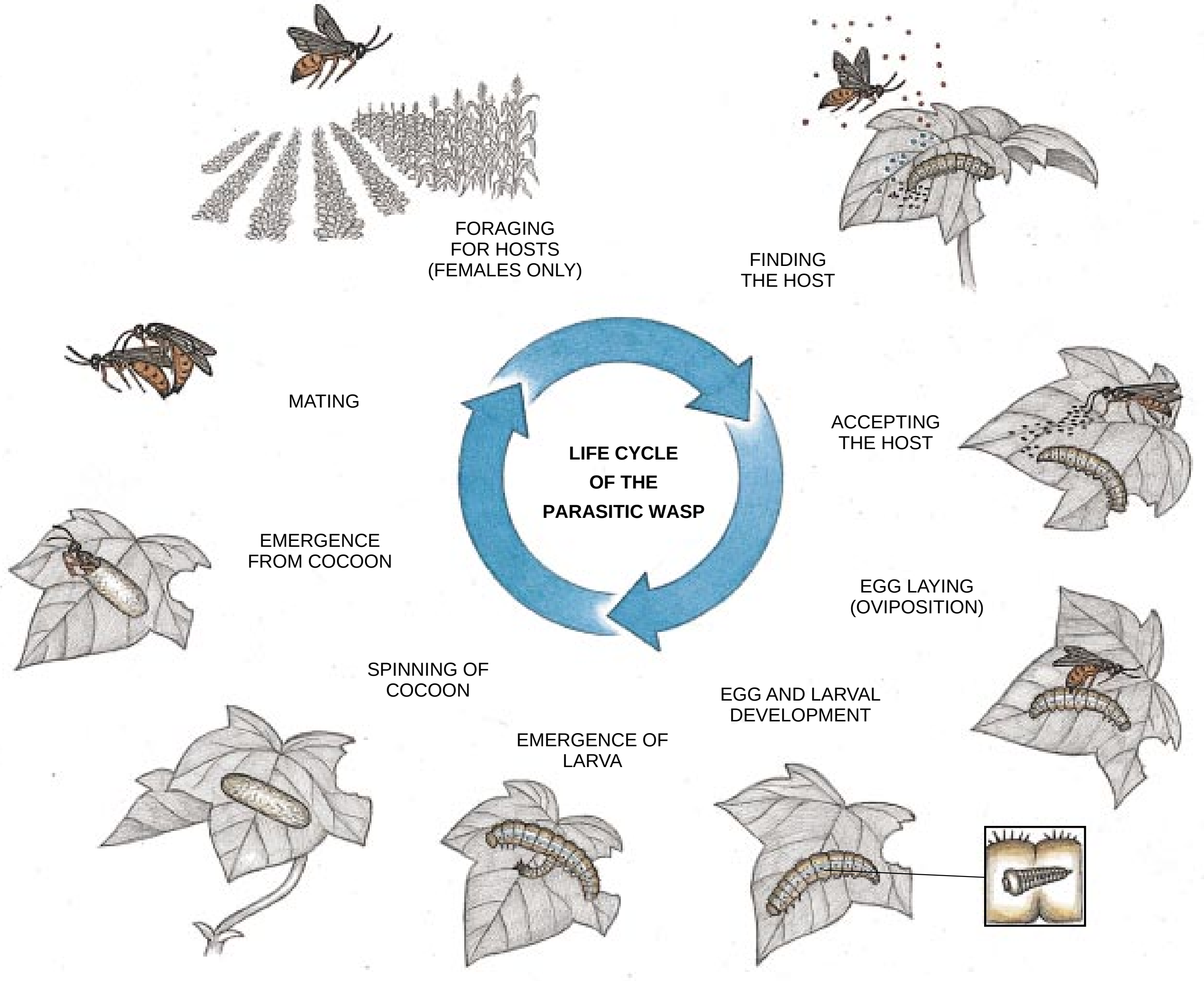
Using beneficial insects to control pests is not a new idea. The first truly notable example is the biological control of the infamous cottony-cushion scale, which was on the verge of destroying the citrus industry in California in the late 1800s. Importation of the vedalia beetle (ladybug), a predator from the pest's native home of Australia, saved the crops. More recently, successful pest control combines beneficial insects with such sound management practices as crop rotation. Within the past 20 years, this latter approach has become the primary means in Europe for averting the ravages of whiteflies and other pests in greenhouses. In the Netherlands the greenhouse area protected in this way increased from 400 hectares in 1970 to 14,000 hectares in 1991. In this case, biological control has clearly proved to be cheaper and more reliable than chemical control.

Yet the vast potential of these natural weapons remains for the most part untapped—in part because researchers have failed to unlock the secrets of how beneficial insects find their prey. The success of biological control has often been hit or miss. Consequently, control for most of our major crop pests remains centered on synthetic pesticides, despite the urgent problems associated with the extensive use of toxic chemicals: environmental pollution, possible food contamination and development of resistance by the pests.

Recent experiments have shown that properly trained parasitic wasps are vital components in ecologically based pest control. Two of our colleagues, Randy Martin and Daniel R. Papaj, each conducted small-scale field studies showing that training in the laboratory improved the initial foraging success of wasps released in the field. Furthermore, the wasps tended to stay in the field for which they were trained. Keeping the beneficial insects on a plot had been a difficulty in some previous attempts at biological control.

Still, it is not yet possible to provide a specific prescription for developing reliable control with parasitic wasps. We think two criteria must be met before wasps become suitable pest-control weapons. First, female parasitic wasps must be lured into and retained in a target area. Second, the wasps must sustain a high level of efficiency in finding and parasitizing hosts.

Several possible ways exist to achieve these goals. Wasps learn easily, so great care must be exercised to make sure the stimuli that appear during laboratory rearing accurately reflect those encountered in the field. Laboratory artifacts could lead to incorrect "programming" of parasites, a problem that might have contributed to failure of such biological control in the past. For example, Vinson and his co-workers found that laboratory-reared *Bracon mellitor*, a parasitoid of the boll weevil, would look for material containing methyl parahydroxybenzoate. This substance, an antimicrobial agent, had been added to the diet used to rear the hosts




**LIFE CYCLE of a parasitic wasp, from egg to adult death, typically takes about five weeks. After locating a host, a pregnant female wasp will confirm her find by touching her antennae to the caterpillar's feces. She then injects one of her eggs into the caterpillar. In about a week, the egg develops into a larva, which feeds on the caterpillar and thereby kills it. After emerging from its host, the larva spins a cocoon around itself. It pupates and emerges as an adult in about 10 days.**

and is not present under natural conditions. Aileen R. Wardle and John H. Borden of Simon Fraser University in British Columbia encountered a similar laboratory problem: wasps trained on caterpillars reared artificially (in plastic egg cups) had difficulty finding hosts.

To maintain an efficient response to volatile stimuli, wasps need to receive steady rewards of caterpillars while they are learning to forage. Studies with *M. croceipes* and *Drosophila* parasites suggest that actually laying eggs is the most important form of reinforcement. Wasps seem to "remember" the smells better after laying eggs than if they had only come into contact with host products. Preliminary field results suggest that such a reward system can be an effective tool.

Plant breeding or genetic engineering may also play a role in biological control. The techniques could produce strains of plants that generate greater amounts of the attractive signals. Although more information is needed to put these ideas into routine practice, the underlying principles are sound enough to develop useful methods. For useful applications, we will need to learn as much about the wasp and its world as the wasp needs to learn to find its host.


FURTHER READING

VARIATIONS IN PARASITOID FORAGING BEHAVIOR: ESSENTIAL ELEMENT OF A SOUND BIOLOGICAL CONTROL THEORY. W. J. Lewis, Louise E. M. Vet, J. H. Tumlinson, J. C. Van Lenteren and D. R. Papaj in *Environmental Entomology,* Vol. 19, No. 5, pages 1183–1193; October 1990.

SEMIOCHEMICALS AND LEARNING IN PARASITOIDS. Louise E. M. Vet and Alex W. Groenewold in *Journal of Chemical Ecology,* Vol. 16, No. 11, pages 3119–3135; November 1990.

EVOLUTIONARY THEORY OF HOST AND PARASITOID INTERACTIONS. Peter W. Price in *Biological Control,* Vol. 1, No. 2, pages 83–93; August 1991.

LEARNING OF HOST-FINDING CUES BY HYMENOPTEROUS PARASITOIDS. Ted C. J. Turlings, Felix L. Wäckers, Louise E. M. Vet, W. Joseph Lewis and James H. Tumlinson in *Insect Learning: Ecological and Evolutionary Perspectives.* Edited by Daniel R. Papaj and Alcinda C. Lewis. Chapman and Hall, 1992.

# Ice Age Lamps

*The invention of fat-burning lamps toward the end of the Ice Age helped to transform European culture. It coincided with several other major technological advances*

by Sophie A. de Beaune and Randall White

The controlled use of fire, first achieved at least half a million years ago, is one of the great innovations in human culture. Although archaeologists and anthropologists generally emphasize the importance of fire for cooking, warmth and protection from predators, the light accompanying fire was also a precious resource, one that made it possible to extend human activity to times and places that are naturally dark. The invention of stone, fat-burning lamps, which happened in Ice Age Europe nearly 40,000 years ago, offered the first effective, portable means of exploiting this aspect of fire. The appearance of lamps broadly coincides with a number of other extraordinary cultural changes, including the emergence of art, personal adornment and complex weapons systems.

Many scholars have hypothesized about how Ice Age lamps functioned and were used, but nobody had ever undertaken a systematic study of them. One of us (de Beaune) therefore set out to examine these lamps in detail and to classify them by type. In conjunction with that project, we built working replicas of stone lamps in order to analyze their effectiveness as light sources and to learn about their design, fabrication and use. The results of this investigation provide a provocative insight into the technology and behavior of some of the earliest modern humans in Europe.

The first object explicitly identified as an Ice Age lamp was discovered in 1902, the year researchers authenticated the wall art in the cave at La Mouthe, France. Archaeologists had presumed that the creation of paintings and engravings hundreds of meters underground must have required an artificial light source. In the course of exploring La Mouthe, they uncovered compelling support of that notion: a carefully fabricated and heavily burned sandstone lamp bearing the engraved image of an ibex on its underside.

Since then, hundreds of more or less hollowed-out objects have been excavated and rather indiscriminately lumped into the category of lamps. The initial research goals were to sift through the potpourri, establish criteria for identifying lamps and examine variation within this category of objects. A search of the literature and of museum collections turned up 547 artifacts that had been listed as possible lamps. The first hurdle was to distinguish lamps from other similarly shaped implements, such as grinding stones. It quickly became obvious that the size and shape of an object are insufficient as defining criteria. For example, lamps need not have a bowl-shaped depression; many perfectly flat slabs show clear traces of localized burning, which in these and other instances provide

SOPHIE A. DE BEAUNE and RANDALL WHITE share a fascination with the culture and technology of Ice Age humans. De Beaune is a member of the Laboratory of Prehistoric Ethnology at the National Center for Scientific Research (CNRS) in Paris. She also teaches biological and cultural anthropology to high school students in France. Her current research focuses on nonflint artifacts from the Paleolithic era; she has participated in and directed several excavations in southwest France. White is an associate professor of anthropology at New York University. He specializes in Upper Paleolithic art and technology and is currently preparing a monograph on the earliest forms of personal adornment among European cultures. White co-edited *French Upper Paleolithic Collections in the Logan Museum of Anthropology,* which brought to light important late Ice Age art and artifacts in an important U.S. repository. De Beaune contributed most of the new research for this article, which White has helped place in a broader scientific context.

**CARVED-HANDLE LAMP, 17,500 years old, represents one of the most elaborate designs used by Paleolithic humans in France. The lamp consists of abraded red sandstone. The bowl holds a fatty fuel and wick; the long handle keeps the end of the lamp cool. Engravings on the handle resemble patterns painted on the walls in the cave at Lascaux, where the lamp was found.**

the only incontrovertible evidence that an object served as a lamp.

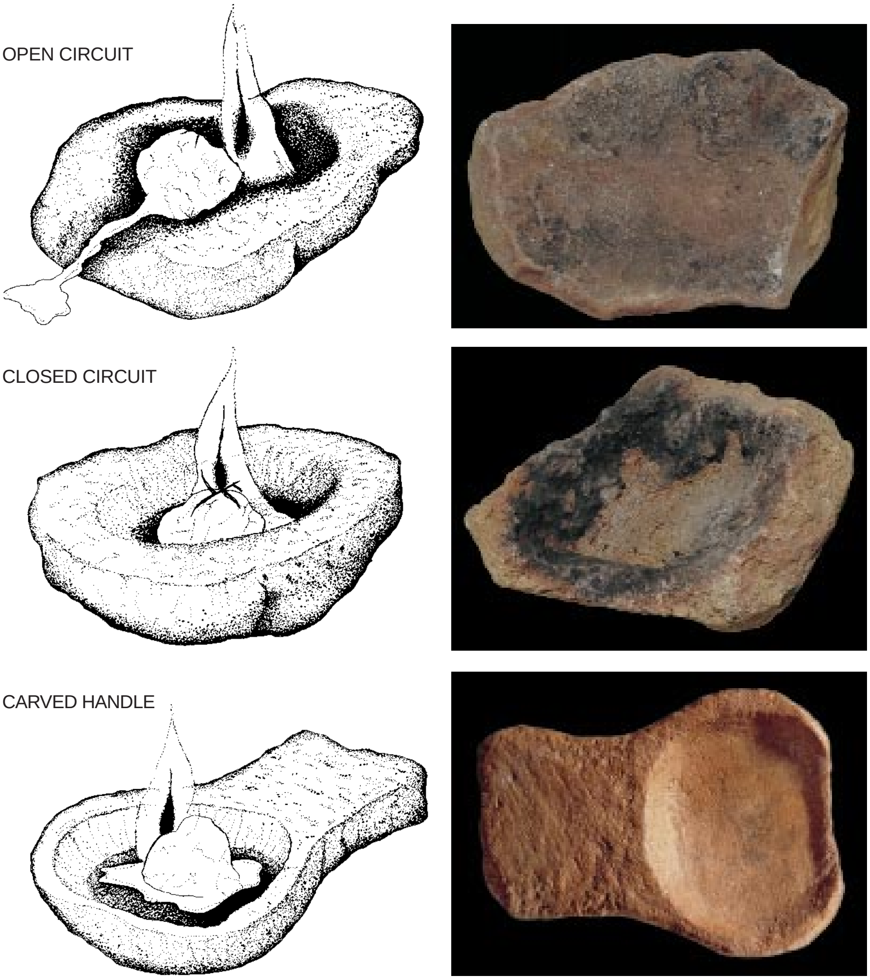
We judged that 245 of the 547 putative lamps clearly served other purposes (mortars, ocher receptacles and so on). The remaining 302 objects were of uncertain status as lamps. We then divided that sample (285 of which have a well-known site of origin) into two categories. We considered 169 of the items to be certain, probable or possible lamps. The other 133 we classified as doubtful or unavailable for study. Markings left by the burning of fuel and wick tend to disappear over time, so the oldest lamps were the most likely to fall into the dubious category. The lamps that we consider here all date from the Upper Paleolithic era, between 40,000 and 11,000 years ago.

The 285 lamps of known origin come from 105 different archaeological sites, mainly in southwest France. The Aquitaine basin has yielded 60 percent of the lamps, the Pyrenean region 15 percent. Considerably fewer lamps have been recovered from other parts of France, and lamps found outside France—in Spain, Germany and Czechoslovakia—are exceedingly rare. Although this pattern may be explained in part by the historically greater intensity of research and the greater number of sites in southwest France, it seems that lamp-producing cultures were in fact restricted to a particular European region.

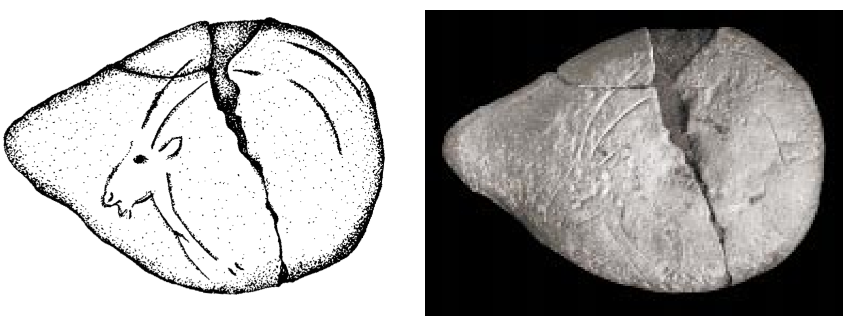
The vast majority of the known stone lamps consist of limestone or sandstone, both of which are fairly abundant. Limestone has the advantage of often occurring naturally in slablike shapes that require little alteration. Moreover, limestone conducts heat poorly, so lamps of this material do not get hot enough to burn the user's fingers. Sandstone is a much better heat conductor, so simple sandstone lamps quickly become too hot to hold after they are lit. Paleolithic people solved this problem by carving handles into most sandstone lamps. Perhaps part of the appeal of sandstone lay in its attractive red color and smooth texture.

Our experiments suggest that the size and shape of the bowl are the primary factors that control how well a stone lamp functions. Setting bowl shape as our primary criterion, we divided the 302 Upper Paleolithic lamps into three main types: open-circuit lamps, closed-circuit bowl lamps and closed-circuit lamps with carved handles.

Open-circuit lamps are the simplest kind. They consist of either small, flat or slightly concave slabs or of larger slabs having natural cavities open to one side to allow excess fuel to drain away as the fat melts; the largest ones are roughly 20 centimeters across. Be-




**LAMP DESIGNS fall into three main categories. Open-circuit lamps (*top*) consist of largely unaltered slabs of rock. When the lamp is lit, melted fat runs off through natural crevices in the rock. Closed-circuit lamps (*middle*) have carved depressions to contain the runoff. Carved-handle, closed-circuit lamps (*bottom*) also have bowl-shaped fuel chambers but are more finely finished and have formed extensions for easier handling. Burn marks indicate that the wick was placed away from the handle.**



**ENGRAVED DECORATIONS often appear on the sides or bottoms of closed-circuit lamps. This carved-handle lamp, which features the incised image of an ibex, was found at La Mouthe in 1902. It was the first object explicitly identified as a lamp.**

cause open-circuit lamps show no noticeable signs of carving or shaping, large numbers of them may have gone unrecognized in premodern excavations. As a result, open-circuit lamps probably are underrepresented in the current sample.

Any slab of rock will work as an open-circuit lamp, so fashioning one requires extremely little effort. The trade-off is that these kinds of lamps inevitably waste a lot of fuel. Open-circuit lamps may be best interpreted as makeshift or expedient devices, easily made and freely discarded. Studies of the modern Inuit show that human groups, even those capable of building large, elaborate lamps, occasionally burn a piece of fat on a stone slab when no alternative lies readily at hand.

Closed-circuit bowl lamps are the most common variety. They are found in all regions, in all periods and in all types of sites where lamps have been recovered. Closed-circuit bowl lamps have shallow, circular or oval depressions designed to retain the melted fuel. The recovered lamps of this kind range from crude to elaborate. Some bowl lamps are entirely natural, some have a slightly retouched bowl and others are completely fabricated. The exterior part of the lamp also may be natural, partly retouched or entirely sculpted. These lamps consist of oval or circular pieces of limestone that are usually the size of a fist or slightly larger. The bowl has sloping sides capable of retaining liquid when the lamp is placed on a horizontal surface. A typical bowl measures a few centimeters across but only 15 to 20 millimeters deep. The largest bowls can hold about 10 cubic centimeters of liquid.

Ice Age closed-circuit lamps resemble those employed by certain Inuit peoples—such as the Caribou, Netsilik and Aleut—who had access to wood for fuel and were therefore not dependent on lamps for heat. Inuit living north of the treeline, where wood was scarce, designed large lamps from slabs of soapstone that were up to a meter across. Those giant lamps (perhaps more correctly thought of as stoves) served many of the same functions as hearths elsewhere, including drying clothes, cooking and heating. There may be direct relations between the quality and abundance of locally available wood for fuel, the presence of fireplaces and the form of lamps at a site.

The most intricate lamps are those we classified as closed-circuit lamps with carved handles. The 30 such lamps in our sample are shaped, smoothed and finely finished entirely by abrasion. Each has a carved handle; 11 of them are decorated with engravings. These lamps appear in the archaeological record somewhat later than the others. The first carved-handle lamps show up in either the Solutrean (22,000 to 18,000 years ago) or Lower Magdalenian (18,000 to 15,000 years ago) cultures. They are particularly abundant in the Middle and Upper Magdalenian (15,000 to 11,000 years ago). Most carved-handle lamps are found in the Dordogne region of France. They are most abundant in rock-shelter sites but are also found in caves and open-air camps.

The elegant design, rarity and limited

distribution in time and space of carved-handle lamps may imply that they served primarily ceremonial purposes. A well-known example from Lascaux, which has been dated to 17,500 years ago, was found on the cave floor at the bottom of a vertical shaft, below a drawing of a hunter confronting a wounded bison. This lamp was discovered by the Abbé Glory, a Catholic lay priest who suggested that such lamps were used to burn aromatic twigs and hence were analogous to incense burners. Too few chemical analyses have been performed, however, to test this hypothesis adequately. The other kinds of stone lamps probably served exclusively as sources of light.

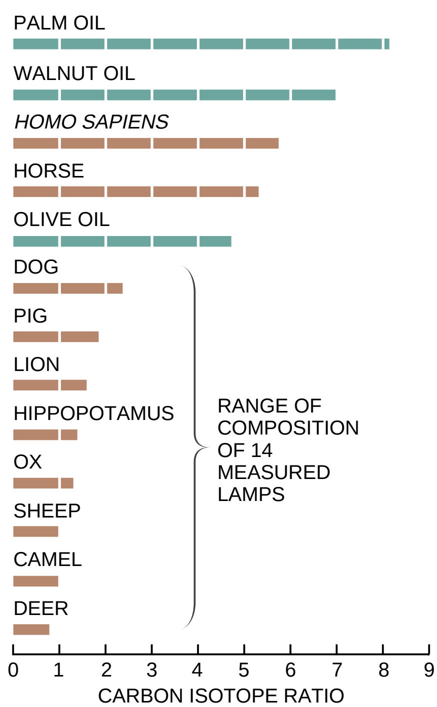
To be effective, a fat-burning lamp must be reliable, easy to handle and bright enough to throw usable light a distance of a few meters in, for example, a darkened cave. The form of lamp that predominates in our sample of Paleolithic lamps is precisely that which our experiments revealed to be optimally efficient. It is a closed-circuit lamp having an oval or circular depression and gently sloping rather than vertical sides. Sloping the side of the bowl facilitates emptying the lamp (so that the wick does not become swamped in melted fat) without dislodging the wick. Carving a gap or notch in the rim of the lamp offers an alternative way to empty the bowl while keeping the wick in place. Eighty percent of the Paleolithic lamps we studied use the sloped-side approach.

Anthropologists have long assumed that animal fat was the fuel burned in Ice Age lamps. From our experiments, we learned that the best fats are those that melt quickly and at a low temperature. Also, they must not contain too much adipose tissue, the connective tissue in fat. Fat from seals, horses and bovids proved most effective in experimental lamps. But were these in fact the fuels favored by Paleolithic humans?

Guy L. Bourgeois of the University of Bordeaux and de Beaune analyzed residues from several Paleolithic lamps to identify the substances they contained. Using two sensitive chemical analysis techniques (vapor-phase chromatography and mass spectrometry), they measured the carbon isotope ratios in fatty acids in the residues. The abundance ratios resemble those in animal fats from modern herbivores, such as cattle, pigs and horses. Unfortunately, scientists have no samples of fat from the actual animals that lived during the late Pleistocene. Nevertheless, the observed ratios of carbon isotopes are quite unlike those in vegetable fats, proving that animals were indeed the source of fuel for Ice Age lamps.

Our investigations also provided new information about the materials from which wicks were made. A good wick must be able to attract melted fat by capillary action and convey it to the free, burning end without being too quickly consumed. Of the wicks we tested, lichen (known to be used by modern Inuit), moss and then juniper worked best. Fritz H. Schweingrüber of the Swiss Federal Research Institute for Forest, Snow and Landscape analyzed several lamp residues. He detected remnants of conifers, juniper and grass, as well as nonwoody residues, possibly lichen or moss. In our experience, juniper wicks are never completely consumed by the flame and so may be better preserved than wicks composed of other plants.

The traces of use on our experimental lamps make it possible to interpret with confidence the markings observed on Paleolithic lamps. Those signs of usage come in three broad forms: light accumulations of soot, deposits of charcoal and reddening of the rock itself, a process known as rubefaction. In 80 percent of all the lamps observed, soot and charcoal deposits are situated within or on the rim of the fuel chamber, where one would expect the wick to lie. Occasional blackening of the side or underside of the lamp can be produced by trickles of melted fat that carried with them small particles of soot. Charcoal deposits result from carbonization




**EXPERIMENTAL CLOSED-CIRCUIT LAMP (*left*) clarifies how these objects were used in Paleolithic times. A lump of fat serves as the fuel; the wick consists of bits of bark, lichen or moss. Melted fat collects in a depression in the rock and must occasionally be poured off. Chemical analysis of Ice Age lamps reveals the presence of residues whose composition resembles that of fat from animals that were common in Paleolithic France (*right*); vegetable fats clearly were not used.**

of the wick or from the heat alteration, or calcination, of adipose tissue in the burning fat.

Thermal reddening often appears on the sides and undersides of lamps, but it, too, most frequently appears in or on the rim of the fuel chamber (in 67.5 percent of the cases). Experience with modern replicas indicates that such reddening took place when hot, melted fat ran onto the side or bottom of the lamp, either as the lamp was being emptied or when it overflowed on its own. Thermal reddening evidently can occur after only a few uses and so provides a helpful indicator of which artifacts served as lamps.

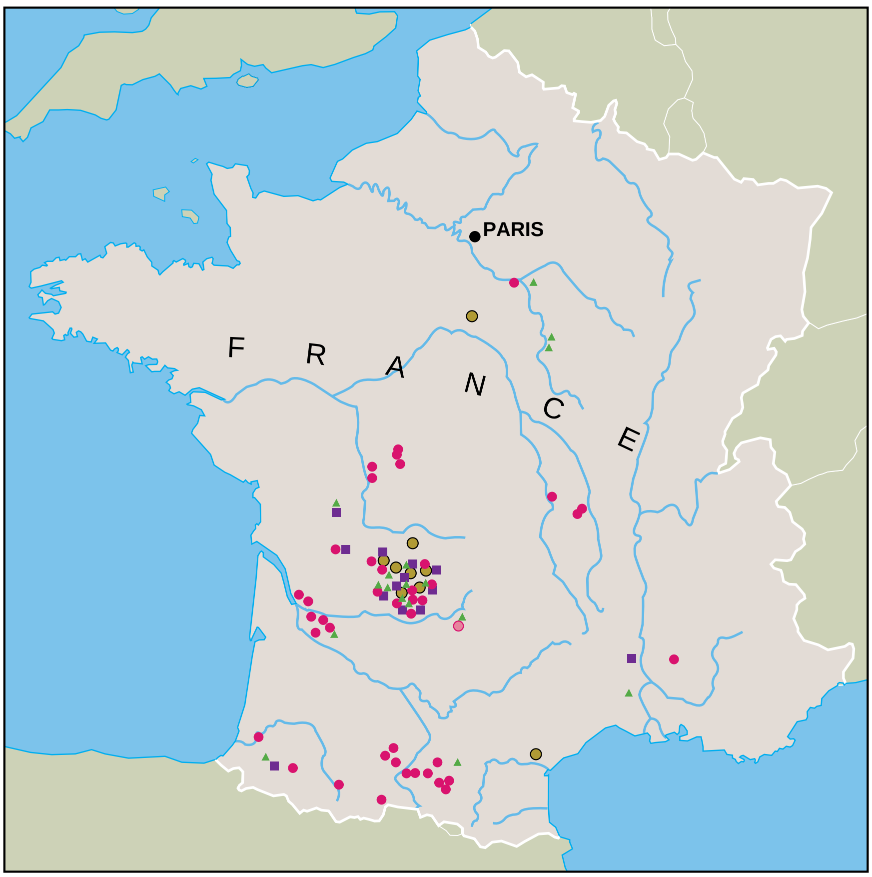
Repeated reuse of a lamp leaves distinct patterns. If a standard open- or closed-circuit lamp is lit on several occasions, the placement of the fat and wick tends to change from one time to the next. Because there is no preferred orientation for those simple lamps, they eventually become blackened and reddened over the entire bowl or surface. The carefully worked closed-circuit lamps that have handles display strikingly different signs of usage. They are oriented the same way each time they are lit, so soot deposits build up on one part of the bowl only, generally the area opposite the handle.

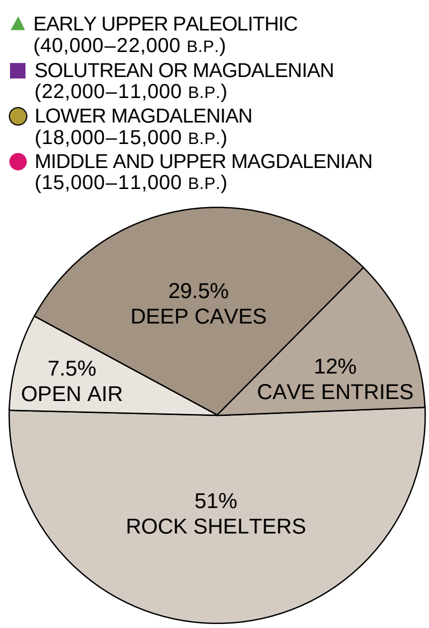
Open-circuit and simple closed-circuit lamps probably were lit only a few times before being discarded. They are so easy to manufacture that there would have existed little incentive to carry them from site to site; we found that we could make a decent lamp in about half an hour. Decorated, carved-handle lamps, which represent a greater investment of labor, were more likely to have been used repeatedly.

To evaluate the effectiveness of Paleolithic fat-burning lamps, one needs to know how much light those lamps could provide. De Beaune investigated this matter by measuring the light output of modern replicas in the metrology laboratories of Kodak-Pathé, France. In quantity, intensity and luminescence, the experimental lamps provided distinctly less light than a standard candle but nonetheless would have been sufficient to guide a person through a cave or to illuminate fine work when placed nearby—assuming, of course, that the visual acuity of Paleolithic people was the same as ours.

The limitations of Ice Age lamps suggest that the creators of cave drawings never saw them as they appear in modern photographs. Human color perception is constrained and distorted at levels less than 150 lux (for comparison, 1,000 lux is typical in a well-lit office). It seems doubtful that the creators of the cave art worked under such bright conditions. Achieving full and accurate color perception of the cave images along a five-meter-long panel would require 150 lamps, each of them placed 50 centimeters from the cave wall. Torches could have provided supplementary light, but few traces of torches have been found in deep caves. On the other hand, the absence or scarcity of lamps in vast cave galleries such as those at Rouffignac, Niaux and Les Trois Frères implies that the creator of the paintings had access to some alternative light sources.

Today when one views the famous cave art in France and Spain, the artificial illumination creates an effect fundamentally unlike that experienced by Paleolithic visitors. Electric lights in the cave of Font de Gaume yield a steady light level of about 20 to 40 lux across a full panel of drawings. Ten to 15 thoughtfully placed stone lamps would be needed to attain 20 lux. A person carrying a single lamp would get a very different impression of the cave art and could view only small portions of the wall at a time. The dim illumination produced by flickering lamps may well have been part of the desired effect of viewing art deep within a cave. The illusion




**ICE AGE LAMPS have been found primarily in southwest France (*left*). Lamps appear in all eras of the Upper Paleolithic (40,000 to 11,000 years ago); more of them have been recovered from the later periods. Surprisingly, most lamps have been retrieved not from deep caves but from open-air sites and from under rock shelters (*right*).**

of animals suddenly materializing out of the darkness is a powerful one, and some cave images are all the more convincing if one cannot see them too well.

Of course, fat-burning lamps were employed for many tasks other than creating and viewing cave art. Lamps are found in such abundance at sites throughout southwest France that they must have been a fairly ordinary item of day-to-day existence. Only about 30 percent of the known lamps were recovered in deep caves. Open-air sites, rock shelters exposed to plentiful daylight and cave entries have provided the rest. The number of lamps at each site (two to three, on average) does not differ significantly from caves to rock shelters to open-air sites.

The location of lamps within sites provides clues to how people exploited them. In deep caves, lamps are often recovered from places where people had to pass, such as cave entrances, the intersections of different galleries and along walls. It would seem that lamps were placed at strategic or predictable points where they could easily be found and reused. The discovery of many lamps lying together—most notably at Lascaux, where 70 lamps have been recovered—implies that lamps were stored in particular locations between uses. Unfortunately, one cannot deduce how many of the lamps were lit at any one time.

Lamps are frequently discovered near fireplaces. Perhaps they were preheated in the fire in order to warm the fat and make it easier to ignite or were abandoned and reused as hearthstones. More likely, fireplaces served as central points of heat and light from which people departed into and returned from the darkness. Many lamps are found inverted in the soil, implying that on returning, people extinguished them simply by turning them over.

In at least one location, a lamp seems to have provided a permanent, fixed source of light within a campsite. Archaeologists found two lamps in a small, natural cavity in the wall of the rock shelter of La Garenne. One lamp had been turned over as if to extinguish the flame. The other was placed upright in a natural hollow in the rock that held it level. The cavity itself would have served as a natural reflector that maximized the lamp's light output.

Sorting through the sample of fat-burning lamps, we sought to learn how their abundance and design changed over time. That analysis is somewhat restricted by the paucity of data. Accurate radioactive dates are available for only the most recently discovered lamps. In most cases, ages are inferred from the archaeological levels in which the lamps were found, and in many early excavations even that information was not recorded. Nevertheless, enough information exists for us to make some general observations.

**CAVE ART must have required artificial illumination both to create it and to view it. The dim, flickering light provided by fat-burning lamps may have been integral to the intended appearance of these subterranean paintings.**

Many more lamps appear in the last cultural period of the Upper Paleolithic, the Magdalenian, than in preceding periods. This may reflect the fact that there are simply more Magdalenian sites known than is the case for earlier periods, as well as the fact that most deep-cave painting took place in the Magdalenian. Older lamps are also harder to identify with certainty.

The form of lamps seems to have evolved surprisingly little through the ages. Some variation in form, material and design occurred, but there is no clear progression from crude to elaborate. Although carved-handle lamps are more common in the later eras, all three primary types of lamp are found throughout the Magdalenian, and even the most elaborate lamp designs date back to the earliest Upper Paleolithic periods, which roughly corresponds to the time when Cro-Magnon, anatomically modern humans, appeared in Europe. The various forms of lamp most likely represent functional responses to particular contexts of use; the need for both simple, easy-to-make lamps and carved, aesthetically pleasing ones apparently was common to all Paleolithic cultures in France.

It is difficult to overstate the importance of artificial light in freeing humans from their evolutionary adaptation to the daylight world. Cave art specialist Denis Vialou of the Museum of Natural History in Paris lauds the Magdalenian cave artists as the people who conquered the world of the underground. But perhaps it is more accurate to see them as the most daring of a long line of our Cro-Magnon ancestors, who, through intelligence and technological innovation, changed the human experience forever by domesticating the realm of darkness.


FURTHER READING

DARK CAVES, BRIGHT VISIONS: LIFE IN ICE AGE EUROPE. Randall White. American Museum of Natural History and W. W. Norton & Company, 1986.

PALEOLITHIC LAMPS AND THEIR SPECIALIZATION: A HYPOTHESIS. S. de Beaune in *Current Anthropology,* Vol. 28, No. 4, pages 569–577; August/October 1987.

TECHNOLOGICAL CHANGES ACROSS THE MIDDLE-UPPER PALEOLITHIC TRANSITION: ECONOMIC, SOCIAL AND COGNITIVE PERSPECTIVES. P. Mellars in *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans.* Edited by P. Mellars and C. Stringer. Princeton University Press, 1989.

NONFLINT STONE TOOLS OF THE EARLY UPPER PALEOLITHIC. Sophie de Beaune in *Before Lascaux: The Complex Record of the Early Upper Paleolithic.* Edited by H. Knecht, A. Pike-Tay and R. White. CRC Press, 1993.

# Flooded Forests of the Amazon

*Parts of the vast rain forest are as much aquatic as terrestrial ecosystems. Unique adaptations allow creatures to thrive in these inundated woods*

by Michael Goulding

If you fly over the rain forest that fringes the rivers of the Amazon during the rainy season, you can catch glimpses of your aircraft between the trees. The giant mirror lying below the canopy is a vast sheet of river water. For an average of six or seven months a year, the lowland rivers rise and invade the enormous floodplain areas that border them. The forest becomes inundated with as much as 10 meters of water; understory plants become completely submerged.

These flooded forests, which make up about 3 percent of the total Amazon rain forest, are one of the keys to understanding the world's richest ecosystem. In the past two decades, scientists have attempted to formulate testable hypotheses to explain the incredible biological diversity found in the rain forest. Many scientists believe that in drier times, especially during periods of heavy glaciation, the area was divided into patches, or refugia, where species evolved because of the geographic isolation of gene pools. When wetter conditions returned in interglacial epochs, the new species supposedly dispersed from their centers of origin into the expanding rain forest.

MICHAEL GOULDING has lived in the Amazon basin for 16 years and has written five books on various aspects of the region's rivers. After receiving his Ph.D. in 1978 from the University of California, Los Angeles, Goulding conducted ecological research at the National Institute for Amazonian Research in Manaus, Brazil, and at the Goeldi Museum in Belém. He is currently a staff scientist at the Rainforest Alliance in New York City and, with his Brazilian colleagues, is developing a project for the management and conservation of Amazonian rivers and fish.

More recently, Paul A. Colinvaux of Ohio State University has argued that the hypothesis of extreme aridity and rain-forest contraction is based mainly on circumstantial evidence [see "The Past and Future Amazon," by Paul A. Colinvaux; SCIENTIFIC AMERICAN, May 1989]. He states that the wealth of life found in the Amazon is better explained by recognizing that the forest undergoes constant change and that it is far from homogeneous. Many different ecosystems thrive because there are myriad kinds of soil, climate patterns, erosional basins and rivers. Communities change when local conditions do—for example, when a river alters its course or when rainfall patterns vary.

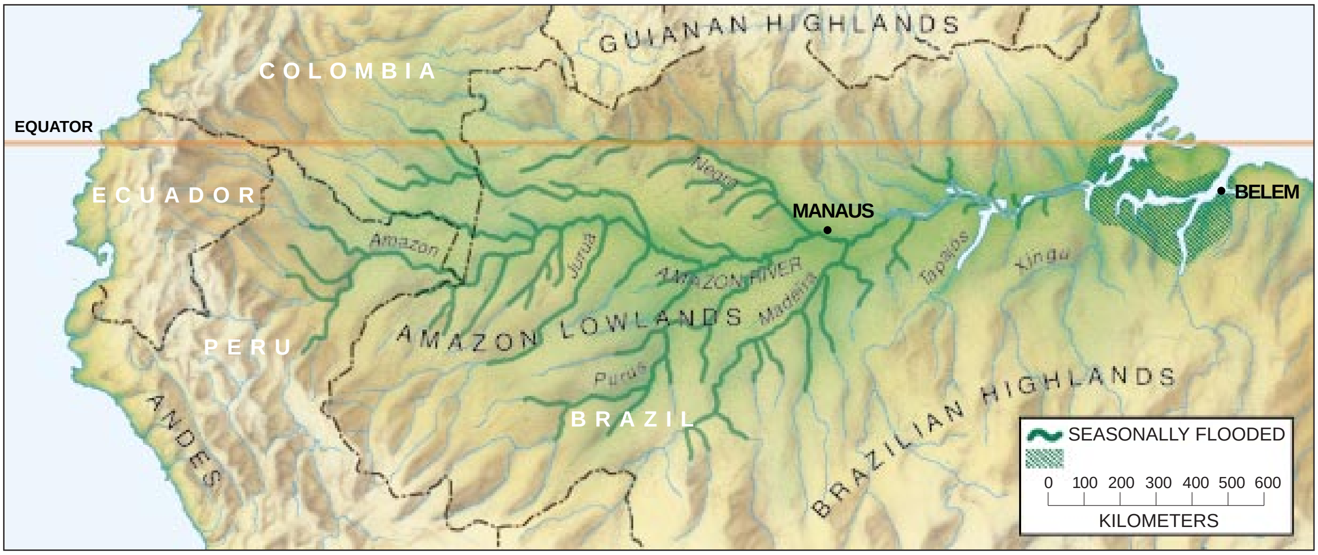
I believe flooded forests provide a striking example of communities that undergo enormous, constant change. The flooded forests promote biodiversity because plants and animals need special adaptations to live in these habitats. Many species found in the floodplains are not common to the upland forest, or *terra firme*. At the same time, many genera have different species in both flooded and dryland forests, a situation that provides an excellent opportunity for comparative phylogenic and biochemical studies.

Flooded forests make up at least 150,000 of the five million square kilometers of the Amazon rain forest. The lowland rivers fluctuate an average of seven to 13 meters a year, and in the central part of the basin, rising river water can penetrate as far as 20 kilometers into the forest on either bank. The woods of the estuary are also subject to flooding, but on a daily rather than seasonal basis. Twice a day, tides push some of the Amazon River's huge freshwater discharge into the estuarine forests. Mangroves are found in parts of these forests where highly brackish or salt water from the Atlantic Ocean is concentrated.

The distribution of the flooded-forest flora is inextricably linked to the history of the river. Before the rise of the Andes, the Amazon flowed westward and is thought to have emptied into the Pacific Ocean via the Gulf of Guayaquil in Ecuador. The uplift of the Andes, which began 15 million years ago in the Miocene epoch, caused the Amazon to excavate a new course eastward through a low-lying stretch of land between the Guianan and Brazilian shields, or highlands. It was probably only after this reversal that the immense expanse of tidally flooded forest came into existence. Silt carried down from the Andes formed most of the islands in the estuarine archipelago, providing a large area for trees to colonize.

The changing course may also have helped increase the floral diversity of many flooded forests while at the same time allowing some species to become widely disseminated. Trees of the flooded forests produce fruit mostly during the high-water periods, when their seeds can be dispersed by currents. Before the rise of the Andes, floodplain seeds were generally carried aquatically from east to west. After the river began to flow toward the Atlantic, species isolated in the west would have been able to colonize the eastern floodplains. This process may help explain why many species, such as some of the rubber trees (genus *Hevea*), are now so widespread in the lowlands.

**AMAZON RAIN FOREST is partially submerged in some areas by about 10 meters of water for a period of up to seven months a year. Special adaptations permit the unique flora and the associated fauna of these flooded forests to survive such heavy inundation.**





**AMAZON BASIN, covering some 6.5 million square kilometers, is essentially a geographic funnel that drains the eastern Andes, the vast Amazon lowlands and the Guianan and Brazilian highlands. Flooded forests, which grow along the seasonally swollen rivers of the region as well as in the tidally washed area of the estuary, make up about 3 percent of the total rain forest.**

Changes in sea level since the early Pleistocene, 1.8 million years ago, undoubtedly led to great shifts in the configuration of flooded forests. The rivers cut deeper in their channels and scoured floodplains when sea levels were low, or they caused flooding when levels were high. These varied patterns suggest that forests along riverbanks were subjected to alteration. These changes, in turn, probably promoted speciation because the forests constantly rebuilt themselves with new combinations of species best adapted to the flooding system of any particular geologic time.

Another factor that has influenced the diversity of flooded forests is the chemistry of rivers. The 19th-century naturalist Alfred Russel Wallace divided Amazonian rivers into three groups: white water, clear water (or blue water) and black water. White water rivers, such as the Amazon and the Madeira, are muddy, or café-au-lait to whitish in color, because their headwaters transport large quantities of silt out of the Andes. These rivers tend to have a neutral or basic pH and are well endowed with nutrients. Clear water rivers, including the Tapajós and the Xingu, originate on either the Brazilian or Guianan shields to the north and southeast of the Amazon lowlands. Because little erosion takes place in those parts, these tributaries have a small load of sediment. They are also slightly acidic.

Unlike the white and clear water rivers, black water rivers arise in the lowlands. Although black water rivers transport little sediment, they are stained by the tea-colored compounds of plants. Black water is formed when the rate of carbon fixation through photosynthesis and carbon's partial decay into soluble organic acids exceeds the rate of complete decay into carbon dioxide. In the case of the Rio Negro, one of the world's four largest rivers, the plant acids originate in communities of stunted vegetation—known variously as *caatinga, campina* or *campinarana*—that thrive in the sandy soils around the river. Such soils, unlike the clays found in most of the Amazon basin, are not equipped to decompose or to trap large quantities of secondary plant compounds. As these materials find their way into streams and rivers, they stain them brown. These waterways tend to be very acidic as well as low in nutrients.

Both the soil and water chemistry of flooded forests are directly linked to river type. The exact chemical factors influencing the composition of species in the watery forests, however, are not understood. Botanists, such as Ghillean T. Prance of Kew Gardens in England and João Murça Pires of the Goeldi Museum in Belém, Brazil, have noted that the floodplains of the various kinds of rivers have major floristic differences. In general, the tallest and most diverse flooded forests are those of the white water rivers. Black water and clear water rivers seem to have more species in common than either type does with its white water counterparts.

The tree species of the flooded forests are often distinct from those found in the *terra firme* rain forest, although species pairs—that is, one adapted to upland and one to flooded forest—are common. For example, *Astrocaryum jauary* is probably the most abundant palm of the floodplain forest in the lowlands; on the *terra firme,* it is replaced by *Astrocaryum tucuma,* a species not found in flooded habitats. The presence of such counterparts suggests that the inundated communities have special biochemical adaptations that enable them to survive long periods of flooding, although what exactly these might be remains unknown. No obvious structural adaptations are universal, and few species have aerial roots, which would seem like an ideal way to cope with the lack of oxygen in the root zone. Yet despite inundation and oxygen-poor conditions during half the year, trees usually retain their green leaves, and the forest, above and below water, remains as verdant as its dryland equivalent.

Some specialized plants are adapted to live in permanently swampy conditions. In these communities, species diversity is greatly reduced, and one or a few palms (*Mauritia flexuosa*) often dominate. Most flooded forests require at least a brief dry spell in order to survive. For the tidally flooded forest, this respite is provided by the low tides.

Just as there are dryland and swamp versions of the plants that thrive in flooded forests, there are animals with similar correlates. The diversity of creatures inhabiting the watery woods is a strong indication that these types of habitat have been present since at least the early Cenozoic era, 65 million years ago. Although no realistic estimate exists of the number of arthropod species in these forests, their diversity far surpasses that of vertebrates.

In addition, the arthropod fauna of flooded forests is different in species composition and behavioral activity from that of adjacent dryland forests. Joaquim Adis of the Max Planck Institute for Limnology in Plön and Terry L. Erwin of the Smithsonian Institution view inundated forests as short-term refugee and long-term evolutionary centers. Rising water brings together creatures that live on the ground and in the trees because soil arthropods migrate upward to avoid drowning. Adis and Erwin believe the concentration of arthropods in the canopy for six months increases competition and predation pressure and that these factors have promoted speciation. They note that terrestrial carabid beetles forced into the trees would have to evolve adaptations for an arboreal existence in order to compete with their canopy cousins. This seasonal adaptive undertaking could lead to speciation.

The extent to which vertebrate animals are ecologically linked to flooded forests varies. The large Amazonian reptiles associated with rivers and forests are noted more for abundance than diversity (in particular, crocodilians, or caimans, and turtles were plentiful before hunting decimated their populations). All six of the side-necked turtles (family Pelomedusidae) found in the lowlands migrate to the flooded forests to feed on fruits and seeds that fall into the water. One of the species, *Podocnemis expansa,* is the largest river turtle in the world. The side-necked turtles could have survived in the Amazon for some 65 million years because of the existence of inundated forests since then. The ability of these creatures to swim through flooded groves and forage successfully may explain why they were so abundant.

The only other large reptiles whose distribution and behavior are associated with the flooded forests are meter-long teiid lizards and iguanas. Both groups jump out of the trees into the water when birds of prey approach. The teiid lizards are adapted to feed in the water: they dive for mollusks and other prey, such as shrimp.

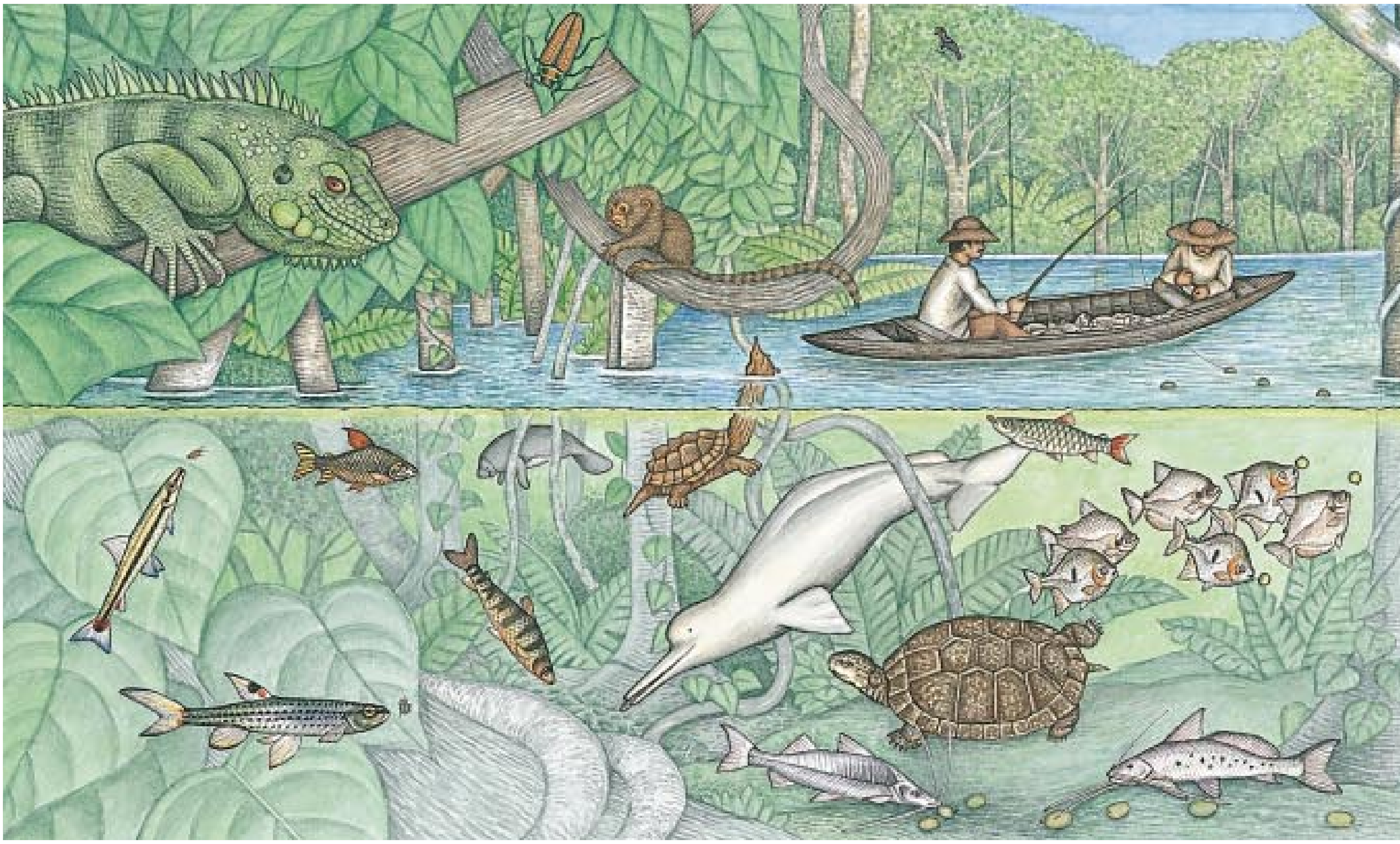
Aquatic mammals also reside in the underwater forests. Manatees occasionally feed on the leaves of some tree species, although their main food consists of herbaceous plants. The boto dolphin (*Inia geoffrensis*), another denizen of these forests, belongs to the primitive cetacean family Platanistidae, of which there are species in Chinese and Indian rivers and a marine form along the coast of Argentina. Unlike the delphinids, the boto and other members of its family have a flexible neck and head because the cervical vertebrae are not fused as they are in all other dolphins. Botos are nearly blind and rely mostly on echolocation, or sonar, to navigate. The ability of the boto to sweep its head from side to side is probably essential to finding its way successfully through the maze of tree trunks.

The inability of the Amazonian relative of the delphinids, the tucuxi dolphin (*Sotalia fluviatilis*), to do this may explain why it avoids flooded forest. The boto follows fish into the forest, while the tucuxi stays in rivers, channels and lakes. What the boto lacks in speed compared with its fleet of fin relatives, it makes up for with its highly adapted echolocation. Perhaps the ability to use flooded woodlands or other complex habitats allowed these dolphins to survive in the wake of the evolution of more advanced delphinids.

The presence of three monkey species that rarely leave the floodplain forests indicates that another form of evolutionary pressure may be at work. Although it remains unclear why the pygmy marmoset (*Cebuella pygmaea*) and the two uakaries (*Cacajao*) restrict their home ranges, one possibility could relate more to disease or parasitism than to competition for food. The epidemiology of floodplain and upland forest animals has not been widely compared. But significant differences in the prevalence of certain human diseases indicate that vectors may be disproportionately associated with one habitat or another. Studies by Jeffrey J. Shaw of the Wellcome Parasitology Unit at the Evandro Chagas Institute in Belém show that leishmaniasis is rare on white water river floodplains because the sand flies that transmit the disease do not occur. The mosquito carriers of malaria are also usually absent in the floodplains of black water rivers because poor nutrient levels and high acidity deter their larvae. Floodplain forests might represent epidemiological refuges for some vulnerable animals.

The most diverse and abundant group of vertebrates in the flooded forest are unquestionably fish: the Amazon basin has the richest freshwater fish fauna in the world. Studies suggest that as many as 3,000 fish species may live there, although no more than 1,800 have been described. A survey I con-

**RIVER CHEMISTRY varies greatly in the Amazon basin, as the contrast between the clear water of the Rio Tapajós and the white, or café-au-lait, color of the Amazon River suggests. Clear water rivers are often slightly acidic and low in nutrients but still sustain a rich flora or fauna because of flooded forests. The silt-laden white water rivers have a lower acidity and higher nutrient levels.**





**FLOODED FOREST is replete with creatures both above and below water: indeed, such woods may be the planet's richest arboreal-aquatic menagerie. Despite the lack of oxygen in their root zones, the specially adapted submerged plants remain as verdant as their dry-land counterparts.**

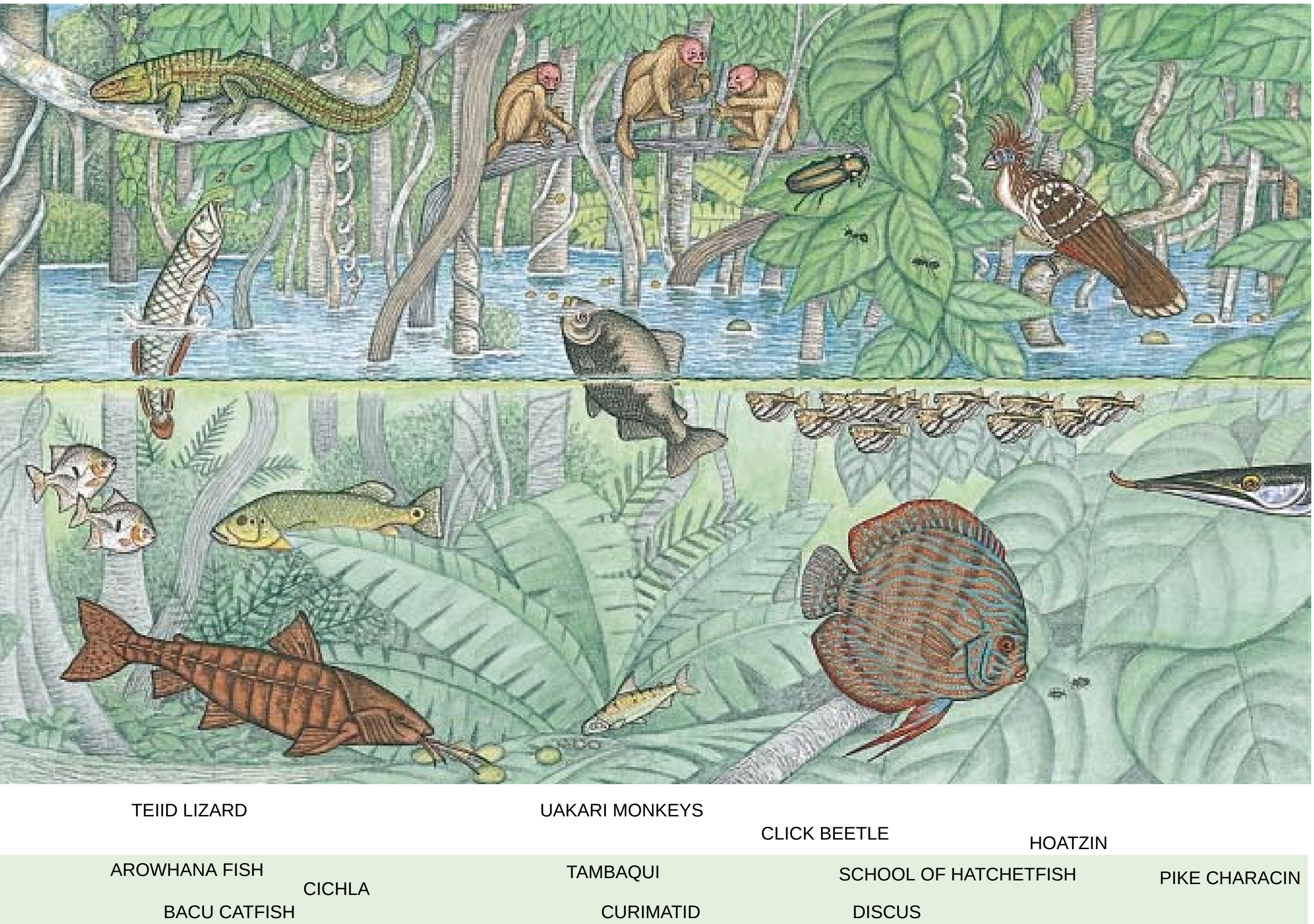
ducted with Mirian Leal Carvalho of the National Secretary of Environment in Brasília and Efrem G. Ferreira of the National Institute for Amazonian Research (INPA) in Manaus revealed that the Rio Negro alone probably has 600 fish species, more than are found in all of North America.

The existence of at least 200 kinds of fruit- and seed-eating fish in the Amazon basin suggests that flooded forests, though subject to constant change, have persisted as large tracts for a very long time. Perhaps such habitats even antedate the appearance of angiosperm plants, which first flourished on a large scale in the Cretaceous period (140 to 65 million years ago) and which currently dominate the rain forests. All Amazonian fruit-eating fish certainly evolved subsequent to the conquest of tropical forests by flowering plants. Interestingly, Klaus Kubitzki of the University of Hamburg found doradid catfish feeding on the fleshy seeds of *Gnetum,* an ancient gymnosperm vine commonly found in flooded forests.

Fossil skulls of fish with teeth similar to some of the nut-eating species of the Amazon have also been discovered in African deposits. Surprisingly, no extant frugivorous fish have been reported in the Zaire basin, despite the fact that this large network of rivers has flooded forests. The Zaire basin, however, is generally higher in elevation and during the Pleistocene suffered more frequent dry periods than did the Amazon basin.

The primary fruit-eating fish of the Amazon forests are the characins and catfish. During the floods, these fish wait under trees or swim through the undergrowth searching for food. Only the characins are equipped with teeth strong enough to crack nuts. Catfish nearly always swallow fruit whole. When fish ingest entire fruits, the seeds pass unharmed through the digestive system and are dispersed. It is tempting to hypothesize that fish and trees have evolved in a mutualistic relation. Although birds, bats and monkeys are also dispersal agents, they take second place to fish because of the much larger populations of the latter. Furthermore, fish commonly feed on the fruits knocked down by animals foraging in the canopy.

The tambaqui (*Colossoma macropomum*), a characin that grows to be a meter long and 30 kilograms, is the best known of the fruit- and seed-eating fish of the Amazon basin. It has been the most important commercial species, and its life history exemplifies the way in which fish use flooded forests and different river types. Larval and young tambaqui (pronounced TOM-bah-key) are confined to the floodplains of the rivers, which receive an annual injection of nutrients from the Andes. These nutrients support relatively high herbaceous plant and phytoplankton production. Indeed, zooplankton blooms can become so extensive that they give floodplain waters an almost souplike




appearance. Tambaqui have numerous elongate and fine gill rakers, which are used to remove zooplankton from the water. Young fish also feed on grass seeds from the floating meadows and on small fruits and seeds from surrounding forests.

Fish-culture experiments show that *C. macropomum* reaches maturity at about four or five years of age. At this time, the fish leave their nursery habitats and migrate to the river channels, where they spawn at the beginning of the annual floods. They then travel to flooded forests of all river types and spend as much as six months feeding. When the forests dry out, adults retreat to river channels, where they spend the other half of the year living off their fat reserves.

Tidally flooded forests have fewer fruit-eating fish than do the inland waters. One of the most abundant species appears to have evolved a life-history strategy that embraces both the tidally and seasonally inundated forests of the Amazon lowlands. The young of *Doras dorsalis,* a large catfish, remain in the estuary, moving in and out of the forest with the tides. They feed on fruits, arum leaves and mollusks. Adults are found upstream, where commercial fishermen catch them as they migrate up river channels, or in seasonally flooded forests, where they feed on succulent fruits.

Although these fish have yet to be tagged and their movements mapped in detail by biologists, the separation of young and adult populations strongly hints at an evolutionary strategy to decrease competition. Such is the case with several predatory catfish species as well. The displacement of the adult segment of the population upstream reduces competition among members of the same species for limited food resources. The fact that there is less time to feed in estuarine forests because of the low tides may contribute to such competition. Heavy fruiting lasts about four to five months in both forest types, but the low tides halve the total amount of time that fish can stay in the tidal communities.

Other fish in the forests rely on a supply of insects from the canopy. We found more than 80 species of fish in the Rio Negro feeding on arthropods, particularly beetles and spiders. Although it is unclear why so many arthropods fall out of the trees, density in the canopy may be a principal cause. Flooding brings about high concentrations on the emergent trunks and in the canopy. Wind and rain also dislodge these animals.

No fish species were found that eat only one or a few kinds of insects or spiders. Most arthropod-eaters are small fish, usually less than 20 centimeters in length. As Stanley H. Weitzman and Richard P. Vari of the Smithsonian have determined, one of the characteristics of the Amazon fish fauna is the large number of small species. Among these miniatures are many of the beautiful aquarium species, such as the much loved hatchetfishes (*Gasteropelecidae*), whose expanded pectoral fins allow them to pop up from below the surface and to snatch insects. The miniaturization of the fish fauna may have evolved as a response to the opportunities to feed on arthropods offered by flooded forests. Small species are more successful at capturing insects and spiders


**TAMBAQUI, one of the most commercially valuable fish in the Amazon region, feeds on rubber-tree seeds in the flooded forest (*left*). The frugivorous fish has extremely strong teeth that are adapted to crush hard nuts (*right*).**

than are bigger fish. In such a complex habitat, it could be difficult for one, or even a few, species to monopolize all the arthropods contributing to the food chain.

Leaves and woody plant material also provide sustenance for the specialized forest fish. Although very few species eat fresh leaves in large quantities, decomposing leaves form the basis of the food chain. In the Rio Negro, we found more than 130 fish species feeding to some extent on detritus. The most peculiar of these detritivores belong to the genus *Semaprochilodus,* a food fish. These characins have fine, bristlelike teeth on the outer part of their lips. The lips are everted to form a suction organ that is used to remove fine detritus.

The feeding ecology of *Semaprochilodus* species is closely linked to flooded forests. The fish remove ample amounts of detritus from submerged tree parts, such as trunks, stems, branches and leaves. In this way, the fish build up fat stores during the rainy season. Although this detritus often contains microorganisms, evidence gathered by Stephen H. Bowen of Michigan Technological University on a similar species supports the idea that the detritivores could not maintain themselves without assimilating the organic material as well.

I believe the trees themselves enrich the attached detritus to some extent by supplying nutrients from decomposing bark and perhaps even from organic exudates. Algae become concentrated near, and add to, flooded-forest detritus communities in nutrient-poor waters. The fish, in turn, help to fertilize the surrounding forests by recycling precious nutrients locked in the detritus. This nutrient recycling by fish is especially significant in black water flooded forests because acidic conditions, long periods of flooding and soils low in nutrients result in very slow decomposition rates.

Within the past decade, international attention has focused on the destruction of the Amazon rain forest. Philip M. Fearnside of INPA estimates that about 6 to 7 percent of the rain forest has been cleared, mostly for cattle pasture. Many people have proposed as an alternative to upland deforestation large-scale development of agriculture on the Andean–Amazon River floodplains, where soils are much better than in *terra firme* rain forests.

In fact, cattle and buffalo ranchers, along with jute farmers and loggers, have already had devastating effects on flooded forests. Only 15 to 20 percent of such areas along the lowermost 2,000 kilometers of the Amazon remain standing today. Fearnside estimates that approximately 325,000 square kilometers of *terra firme* rain forest have been destroyed so far—largely within the past two decades. If these rates of deforestation are redirected to the floodplains, as livestock operations suggest they might be, then the flooded forests will be almost entirely eliminated within the next decade.

The destruction of the flooded forests may be the single most significant threat to Amazonian biodiversity. The fact that *terra firme* deforestation has overshadowed flooded-forest losses in the popular and scientific media is mostly the result of erroneously seeing the Amazon forests as uniform. The tree species of the inundated woodlands are unique and cannot be replaced by *terra firme* counterparts once they are gone. Indeed, it is reasonable to argue that the flooded forests have been more significant as gene pools for *terra firme* than vice versa. Floodplain trees can survive without flooding, as planting them in upland soils demonstrates. But upland species cannot tolerate long periods of aquatic immersion. Thus, it seems speciation occurred from floodplain to upland.

Many of the animals that live in and depend on the flooded forests are also unique. The destruction of these woods could cause the greatest loss of freshwater fish known in human history. These creatures have been a primary source of animal protein for Amazonian cultures and provide a strong economic argument for preserving these habitats. As in so much of the world, however, there is no effective policy to prevent ranchers from transforming complex ecosystems into pasture. That the magnificent biodiversity of the flooded forests could be swept away is a chilling thought, even under the tropical sun.


FURTHER READING

THE FISHES AND THE FOREST: EXPLORATIONS IN AMAZONIAN NATURAL HISTORY. Michael Goulding. University of California Press, 1980.

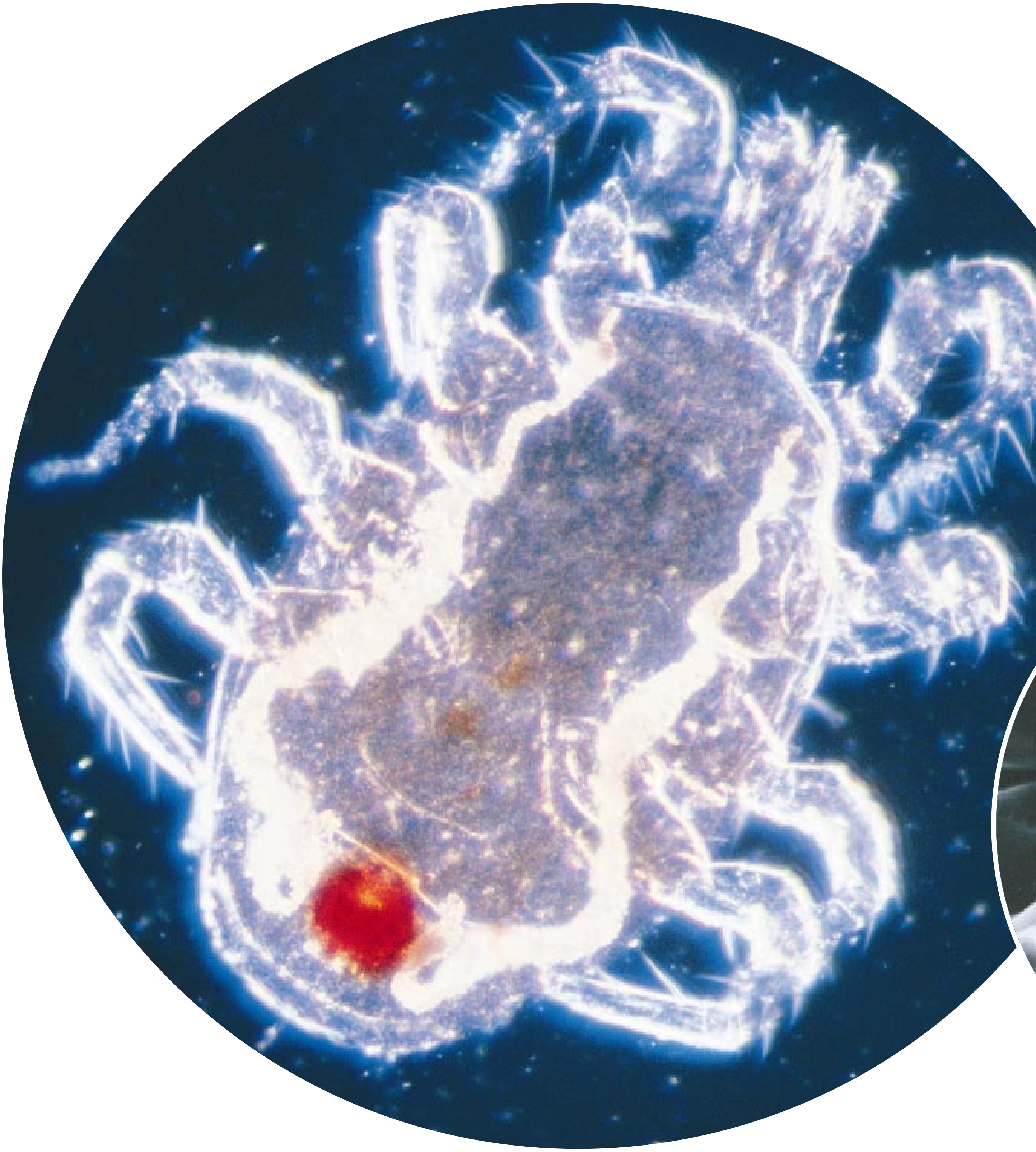
THE AMAZON: LIMNOLOGY AND LANDSCAPE ECOLOGY OF A MIGHTY TROPICAL RIVER AND ITS BASIN. Edited by Harald Sioli. Dr W. Junk Publishers, Dordrecht, 1984.

KEY ENVIRONMENTS: AMAZONIA. Edited by Ghillean T. Prance and Thomas E. Lovejoy. Pergamon Press, 1985.

AMAZON: THE FLOODED FOREST. Michael Goulding. Sterling Publishing, 1990.

# DNA'S NEW TWISTS

by John Rennie, *staff writer*



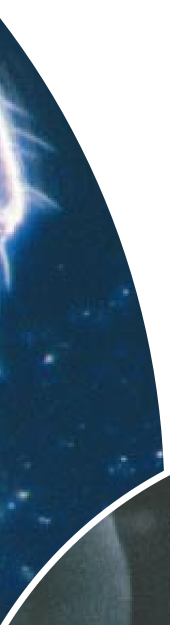
**The known rules of genetics are only the beginning. The newly discovered abilities of a familiar molecule are influencing theories about evolution and the inheritance of disease.**

Margaret G. Kidwell is patiently waiting for a miracle. Bottles in her laboratory at the University of Arizona are the communal homes of two fruit fly species. Because the flies cannot interbreed, they should remain genetically distinct. Yet if Kidwell is right, bits of DNA will somehow pass from one species to the other. Classical Mendelian genetics says that kind of inheritance should be impossible, but Gregor Mendel never reckoned with genes that might hitch a ride inside the hungry parasitic mites that prey on both species. Kidwell and her co-workers have found evidence that such a transfer has already happened at least once in the wild.

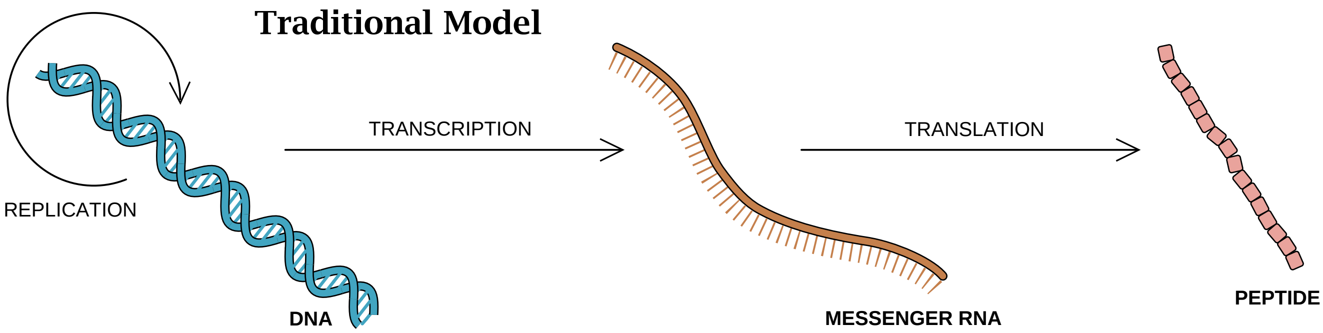
Mendel's laws may have intriguing loopholes, as Kidwell and other modern geneticists have learned. The new tools that molecular biology has provided for assaying chromosomes have helped researchers take the measure of the genetic machinery more accurately. And they are discovering intriguing twists to the traditional, more simplistic models of DNA's behavior. Contrary to expectations, genes sometimes leap from one chromosome to another or expand and contract like accordions. Chromosomes seem to carry chemical tags that identify whether they originated in an organism's mother or its father. Proteins can sometimes be specified by genes that, in the conventional sense, do not exist. Workers have even found indications that organisms may be able to respond to changes in the environment by altering their genes.

"We must pay much more attention to the diversity of detail in the ways that organisms play out the basic properties of DNA, RNA and proteins," says Joshua Lederberg, a pioneering molecular geneticist and Nobelist serving on the faculty at the Rockefeller University. During the 1940s, Lederberg made his reputation by showing that bacteria engaged in a sexual exchange of genetic material—a heretical notion at the time but now an accepted article of genetic principle.

The new heresies being discussed may be the keys to similar major revisions in genetics. Already workers in the field are waking to the fact that DNA is not quite the predictably stable molecule they have often taken for granted. "DNA isn't this inert thing encased in Lucite, sending out instructions," insists Jeffrey W. Pollard, a developmental biologist at Albert Einstein College of Medicine. "It's part of the cell, and it's responding to what's happening around it." In conversation, Pollard and other biologists occasionally refer to DNA as a



*TINY MITE* (left) *may have transferred pieces of DNA from one species of fruit fly to another on its pipettelike mouthparts* (right)*, according to a recent finding. Biologists are becoming increasingly attentive to the possibility of such horizontal transfers and to other phenomena that lie outside traditional genetics. The red spot in the mite's abdomen is pigment taken from a fruit fly.*




**DYNAMIC MODEL of the genetic mechanism is emerging from current research. Under the simple previous model (*left*), DNA replicated itself. Genes in the DNA were transcribed as messenger RNA, which was directly translated into peptides, or pieces of proteins. The new view (*right*) involves the possibility of many additional activities, only a few of which are illustrated here. Transposable genes can change their position within the DNA. RNA is occasionally reverse-transcribed into DNA. A messenger RNA molecule is not complete until it has been chemically processed in any of several ways. The translation of the messenger sometimes involves a recoding, or reinterpretation, of its genetic message.**

"metabolic molecule" to reflect its biochemical responsiveness.

Many researchers are wondering what the full significance of the emerging view will prove to be. "I think a great deal of genetics and evolutionary theory has not yet caught up with the implications of the dynamic chemistry of DNA," Lederberg remarks. He cautions that some of the phenomena may turn out to be will-o'-the-wisps having no more than marginal importance. Nevertheless, every level of molecular activity is clearly far more complex than was once believed. No one is throwing out the old genetics texts, but new chapters are certainly being written.

The science of genetics was born during the mid–19th century in the garden of the Moravian monastery where Mendel experimented with pea plants. The mathematical rules of inheritance that he and subsequent workers developed became the cornerstone of modern biology. None of those researchers knew what genes were, however: they were just convenient abstractions for describing the transmission of traits. Not until the 1950s was DNA (deoxyribonucleic acid) identified as the primary genetic material, thus opening the era of molecular biology.

The substitution of "DNA" for "gene" did little to alter the fundamentals of genetics. In most discussions, DNA was still treated as an ideal entity that, barring the occasional mutation, was a comfortable constant in the turbulent internal environment of a cell. Genetic information, researchers learned, was written in the double-strand helix of DNA as a sequence of its four constituent nucleotide bases. Single strands of DNA acted as templates for making complementary molecules of messenger RNA (ribonucleic acid), which carried the information to the organelles called ribosomes. The ribosomes then read the messenger RNA as though it were a ticker tape and, interpreting its bases three at a time, popped together appropriate amino acids to make a protein. But as investigators have more closely scrutinized the actions of the genome (the complete set of an organism's genes) and its associated molecules, they have found that their properties and behavior are far more diverse.

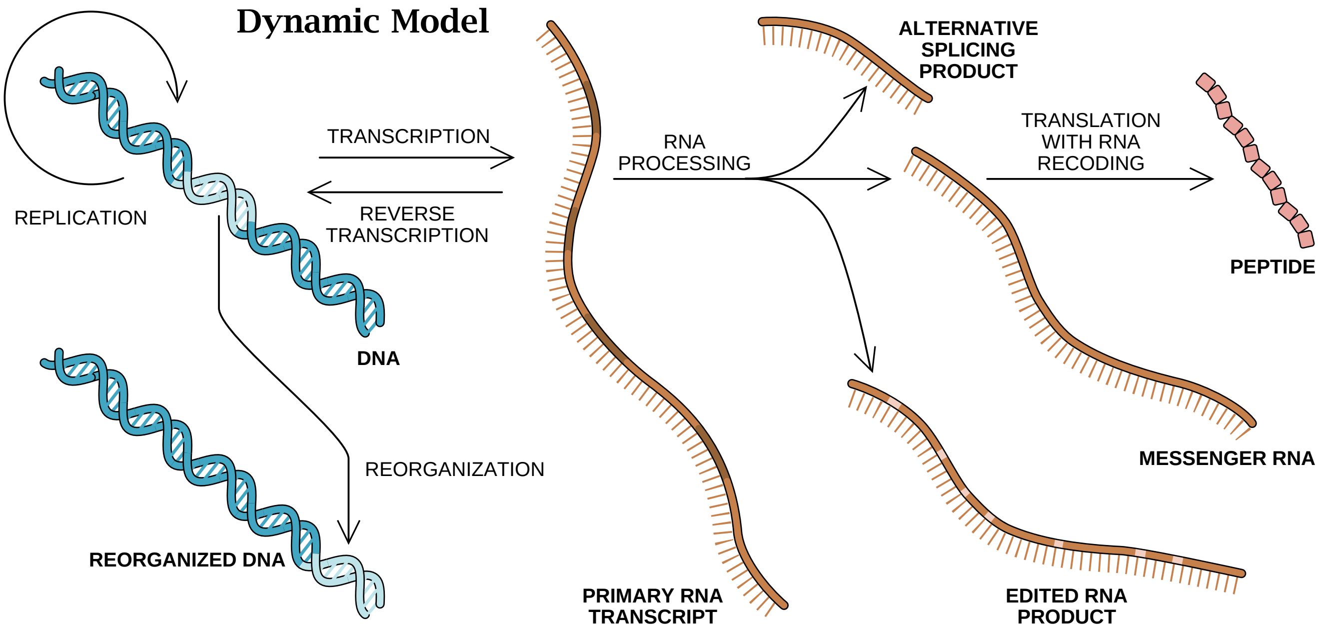
### Jumping Genes

One of the first scientists to appreciate the dynamism of the genome was the geneticist Barbara McClintock, who died in 1992. In 1947, while performing breeding experiments on maize at Cold Spring Harbor Laboratory, McClintock saw odd patterns in the inheritance of pigments that the conventional rules could not explain. After puzzling over her results, she concluded that a few of the genes did not have fixed locations on a chromosome. Rather they seemed to leap from one spot to another between parent and progeny. Environmental stresses such as heat appeared to increase the rate at which such genes transposed themselves.

**GREGOR MENDEL deduced many fundamental laws of genetics from his breeding experiments on pea plants during the 19th century.**

Like Mendel's results, McClintock's idea of transposable genetic elements, often called transposons or "jumping genes," languished for decades. Nothing in the one-way flow of information from DNA to RNA to proteins seemed to allow for the possibility that genes could move. In the early 1970s, however, McClintock's theory was vindicated. Molecular biology experiments proved that small pieces of DNA did sometimes move within or between chromosomes, thereby triggering changes in gene expression. By jumping to a site beside or within a gene, for example, a transposon could shut it off. McClintock's discovery won her the Nobel Prize in 1983.

The origin of transposons is still obscure, but many are believed to be the remnants of viruses that permanently integrated their genes into their host. A few viruses produce an enzyme called reverse transcriptase, which enables them to convert strands of RNA into DNA, thereby reversing the typical flow of information in cells. Jumping genes generally seem to rely on a similar mechanism for their gymnastics: some transposons make their own reverse transcriptases; others bor-




row enzymes produced by more capable elements or viruses.

However they accomplish their jumps, transposons can exert a powerful influence on the organism bearing them. In a variety of plant and animal species, researchers have documented many genetic alterations caused by just such transpositions. Jumping genes have recently been shown to influence human health as well.

In 1991, for example, Francis S. Collins and his colleagues at the University of Michigan identified in one patient the mutation causing neurofibromatosis, a tumor-producing disorder. A gene that normally regulates cell growth had been rendered inactive by the insertion of a common genetic element called Alu. Two months later a team at the Johns Hopkins University School of Medicine headed by Haig H. Kazazian announced that it had caught another gene almost in the act of jumping. While studying a group of hemophiliacs, the researchers discovered that in one child the disease was caused by a transposon that had inactivated a gene for an essential blood-clotting factor. The transposon was essentially identical to a gene at a different location in the child's parents.

Transposable elements are usually passed "vertically" from one generation to the next, like ordinary genes. Yet their ability to take short hops within cells has also prompted speculation about the possibility that under rare circumstances they might jump "horizontally" from one organism to another—and maybe even between species. Bacteria do exchange genes, as Lederberg proved, and transfers between bacteria and plants and insects are strongly suspected.

But no one has yet proved that transposons can move between higher organisms. John F. McDonald, a molecular geneticist at the University of Georgia, notes that for many years biologists have noticed peculiar genetic similarities between various pairs of unrelated but cohabiting species. These similarities suggested that genetic information had passed between them, "but it was a lot of arm waving because you couldn't really prove it was a genetic transfer," he says.

**BARBARA McCLINTOCK, after studying maize, made the revolutionary discovery that some genes could "jump" within and between chromosomes.**

### The Mite That Could

The best evidence, in the eyes of many researchers, comes from work presented in 1991 by Marilyn A. Houck, a mite specialist now at Texas Tech University, and by Kidwell. They have found hints that recently a jumping gene called a P element moved between two species of fruit fly.

During the 1970s, Kidwell and others noticed that a strange genetic incompatibility had arisen between wild and laboratory stocks of the fruit fly *Drosophila melanogaster*. When the researchers tried to crossbreed the fly populations, they found that the crosses were often infertile or produced abnormal offspring. These incompatibilities were eventually traced to the ubiquitous presence of P elements in the wild flies. Only flies that had been isolated in laboratories for many generations were free of them. The researchers deduced that the epidemic of P-element infestation must have started within the previous 30 to 50 years.

Gene-sequencing studies by Stephen B. Daniels of the University of Connecticut revealed that the P elements in *D. melanogaster* were virtually identical to those in a different fruit fly, *D. willistoni*. Those findings created a strong circumstantial case that sometime in the 1940s or so—perhaps right around the time that McClintock was theorizing about jumping genes—P elements hopped from the *willistoni* to the *melanogaster* flies.

How the transfer had taken place would probably have remained a mystery if Kidwell's fly stocks had not become parasitized by an unusually nasty variety of mite. Houck, who worked just down the hall from Kidwell, was given samples of the pests to find a way to exterminate them. The mites were eventually identified as *Proctolaelaps regalis*. After examining them under an electron microscope, Houck noticed that their mouthparts bore a striking similarity to the thin glass tubes used by biologists in gene transfer experiments.

That similarity suggested to Houck that nature might have used the mites for a little genetic engineering of its own. After dining on a *D. willistoni,* a mite with P elements still on its mouthparts or in its digestive tract could have feasted on an egg of *D. melanogaster*. Conceivably, P elements that passed into the egg might have integrated with its DNA. If the egg survived, the result would be a fly that had inherited P elements from another species.

"There's no doubt in my mind," Houck says, "that this will turn out to be horizontal transfer." Kidwell remarks, "The P element spreads like crazy once it gets inserted properly. I think it's that initial step of getting integrated that's the tricky one."

The theory involves many suppositions and long shots, but Houck and Kidwell have already established that at least part of it stands up. They have shown that mites preying on the flies do pick up recognizable P elements in their gut. Houck and Kidwell are now separately doing experiments to see if they can complete the transmission process. Kidwell is raising *D. melanogaster* and *D. willistoni* together in the presence of mites to see whether she can reproduce the interspecies transfer. Houck is looking at more specific aspects of the problem.

At this point, the scientists have not ruled out the possibility that viruses might have participated in the cross-species jump. For many years, viruses have been discussed as theoretical vectors for transposons, which they might have woven into their own genetic material and carried to different host cells. McDonald notes that researchers have found at least one example of a transposon in a virus that infects several species of insects.

If transposable elements can be shown to move between species, the finding will further improve the case for them as agents of evolutionary change. Last summer hundreds of geneticists gathered in Athens, Ga., to explore the ramifications of transposable elements for evolution theory. "The traditional Darwinian view is that evolution proceeds gradually through the accumulation of point mutations and that selection pulls out what is not good," explains McDonald, who organized the conference. "But the thing about transposable elements is that they tend to produce macromutations—sudden, substantive changes in phenotype—in shorter periods of time."

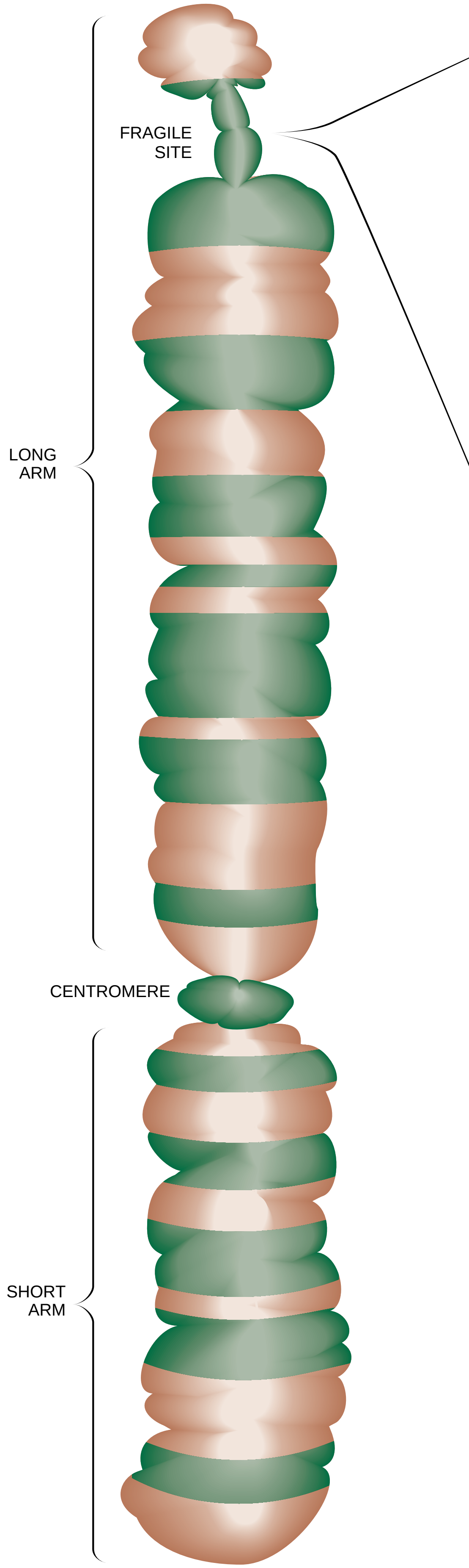
By suppressing the expression of other genes, McDonald offers, transposons could function as genetic regulators. "On a molecular level they're producing new regulatory networks, but on a phenotypic level that may translate into new developmental patterns," he says.

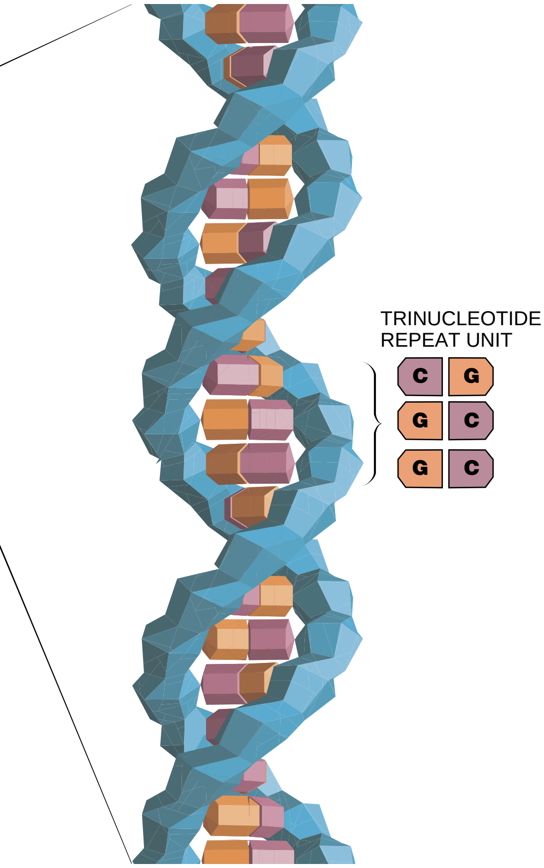
Diane M. Robins of the University of Michigan has found what may be one bit of such a network. As she explained at the Athens conference, she has discovered that the regulatory sequence for one gene in mice strongly resembles part of a transposon found elsewhere in the animal's genome. The insertion and retention of the transposon component at that site seem to have made the gene responsive to levels of circulating hormones.

Other hints that a transposon may have had a more clearly beneficial effect in human evolution have been discovered by Linda C. Samuelson, who works with Robins. In many mammals the pancreas secretes the enzyme amylase to digest starches in the diet. But humans also secrete amylase in their saliva, which seems to broaden the range of foods they can eat. Samuelson has shown that a transposon may have enabled this dual expression of amylase by altering the regulation of the gene for the enzyme.

Some biologists even speculate that transposons play a major role in the origin of species. Jeffrey Pollard thinks transposons may be partly responsible for the punctuated pattern of evolution that many paleontologists see when they look at the fossil record. In his view, organisms that find themselves in unfamiliar environmental niches may experience stresses that hasten the rate of gene jumping. In effect, the organisms would become more mutable and might therefore evolve faster.

Kidwell takes a more cautious view of the role of transposable elements in evolution. "I think at the moment it is still very much up in the air," she comments. "Personally, I have to admit a bias toward there being a functional role, but I think it could be really com-






**FRAGILE X CHROMOSOME is named for a spindly region near the tip of its long arm. The fragile site is created by a bizarre mutation that greatly enlarges unstable forms of the gene *FMR-1*. The expanded region contains hundreds to thousands of tandem repeats of three specific nucleotides.**

plex. We can't make any great claims for it yet."

### The Amazing Colossal Genes

Whatever the extent of their effects, jumping genes represent a type of mutation unforeseen by the founders of genetics. In recent years, researchers have discovered another one that is in many respects even more unorthodox: abnormal genes that suddenly balloon in size, with tragic consequences. The discovery of this type of mutation is helping geneticists to fathom the odd patterns of inheritance associated with several diseases. One of them is fragile X syndrome, the most common inherited cause of mental retardation, which draws its name from a chromosomal deformity in its sufferers. They have an X chromosome in which the tip of its long arm is attached by only a slender thread of DNA.

In 1991 groups led by Jean-Louis Mandel of INSERM in Strasbourg, Grant R. Sutherland of Adelaide Children's Hospital in Australia and Stephen T. Warren of the Emory University School of Medicine jointly discovered that the cause of the disorder was a mutation unlike any previously seen. In normal individuals, a gene designated *FMR-1* contains about 60 or fewer tandem repeats of a particular trinucleotide base sequence. Healthy carriers of fragile X syndrome may have as many as 200 tandem copies. In sick individuals the tandem repeat region is fantastically larger: many hundreds to thousands of the base triplets appear. Because children with fragile X syndrome are the offspring of healthy carriers, the mutant genes must be growing from one generation to the next.

Similar expansions have been found to cause myotonic dystrophy, the common form of muscular dystrophy in adults, and a rare condition called spinal and bulbar muscular atrophy. The mechanism of this explosive change is still unknown, but researchers suspect that an aberrant form of polymerase—an enzyme that adds nucleotides to growing DNA strands—is the culprit.

"The big question, I think, is that there are a lot of repeat regions in the genome," Warren reflects. "Why don't you see those undergoing this massive expansion?" In fact, he notes, David E. Housman of the Massachusetts Institute of Technology has found evidence that a number of sequences in the genome of mice do expand and shrink, though less dramatically than the one in *FMR-1*.

Warren concludes that some singular instability in certain forms of *FMR-1* may predispose it to extreme growth. Over one or more generations, the unstable allele, or form of the gene, grows to the size seen in the healthy carriers of fragile X syndrome. When the gene reaches a critical length, it is primed for more spectacular elongation, which occurs in the affected individuals. For reasons not yet clear, the gene seems to expand most dramatically in individuals who have inherited the fragile chromosome from their mother. "It seems quite clear that first there is a change from a stable to an unstable allele," he says.

An interesting consequence of the expanding mutation phenomenon, Warren muses, is that "it makes you look again at quantitative genetics in humans." Most mutations reach a stable equilibrium frequency in a population within a few generations because selection can favor or remove them proportionally to the mutation's effects. "That's textbook genetics," he says. "But now we're saying, no, there's a mutation that has no effect on offspring or on their offspring, but somewhere later down the road it does."

The expanding mutations illuminate the highly peculiar pattern of inheritance that marks fragile X syndrome, also known as Sherman's paradox. The classical genetics for X-linked diseases predicts that all males carrying the fragile chromosome will be affected by it. In fact, more than 20 percent of those men are perfectly normal because they have shorter "premutation" forms of *FMR-1*. Their children, too, are normal because their genes have expanded only minimally. Both the male and female grandchildren of the original carriers, however, are often retarded because the repeat region in their genes has spectacularly expanded.

Warren and many others maintain that the observed pattern of gene expansion is sufficient to explain Sherman's paradox. A few dissenting biologists, however, argue that something more is going on. "The fact that you see the biggest changes in nucleotide sequence occurring only through Mom's germ line is still difficult to explain," insists Carmen Sapienza of the Ludwig Institute for Cancer Research in La Jolla, Calif. Sapienza is one of a group of researchers who think another phenomenon that violates the traditional genetic dogma may also be at work: sex-specific imprinting of chromosomes.

### Gene Imprinting

One fundamental assumption of Mendelian genetics is that the effect of a gene is totally independent of whether it came from an organism's mother or from its father. Yet geneticists have found compelling instances in which males and females seem to imprint, or mark, the genes that they pass on. Experiments on mice have shown that embryos with a full set of chromosomes derived from just one sex inevitably fail to reach birth, even though they are genetically identical to normal mice. Without a maternal imprint on its chromosomes, an embryo becomes abnormal; without a paternal imprint, the placenta fails to develop.

In humans, imbalances in the complement of maternally and paternally imprinted chromosomes occasionally lead to illness. Robert D. Nicholls and his colleagues at the University of Florida have found interesting evidence for the effects of genome imprinting in at least two diseases. Children who have Prader-Willi syndrome, which is characterized by mental retardation and obesity, often seem to have inherited both of their copies of chromosome 15 from their mothers. Conversely, children lacking portions of their maternal chromosome 15 (and whose paternal chromosome is therefore disproportionately represented) exhibit the mental retardation and staccato movements of Angelman syndrome. Sapienza and others have also linked abnormal imprinting to several childhood cancers.

Precisely how genes can be imprinted is not yet certain, but the process may involve the chemical linkage of methyl groups to cysteine, one of the bases in DNA. Methylation of DNA seems to inactivate the marked genes. Moreover, the pattern of methylation is retained when the DNA replicates. Imprints in a fertilized egg can therefore be passed on to all the cells of the developing body.

Charles D. Laird of the University of Washington postulates that imprinting of the fragile X chromosome, rather than genetic expansion, is the key to Sherman's paradox. He has developed a model for the disease that he believes describes its behavior more accurately than do the molecular data alone.

It is built around the observation that females, who have two X chromosomes in each cell, typically turn off one of them through imprinting. Normally, this imprint is fully removed before the chromosomes enter meiosis, the cell-division process that produces ova. Laird suggests that in the carriers of fragile X, a mutation in the chromosome sometimes prevents the removal of the imprint from the fragile site.

Consequently, about half of the sons of a healthy woman carrying the genetic defect will receive an X chromosome that is partially but permanently inactive and will be mentally retarded as a result. Some of her daughters will also receive an inactive fragile X, and they too will suffer. In Laird's view, the expanding mutations seen by Warren and his colleagues could be no more than secondary effects of the deactivating imprint on the fragile site.

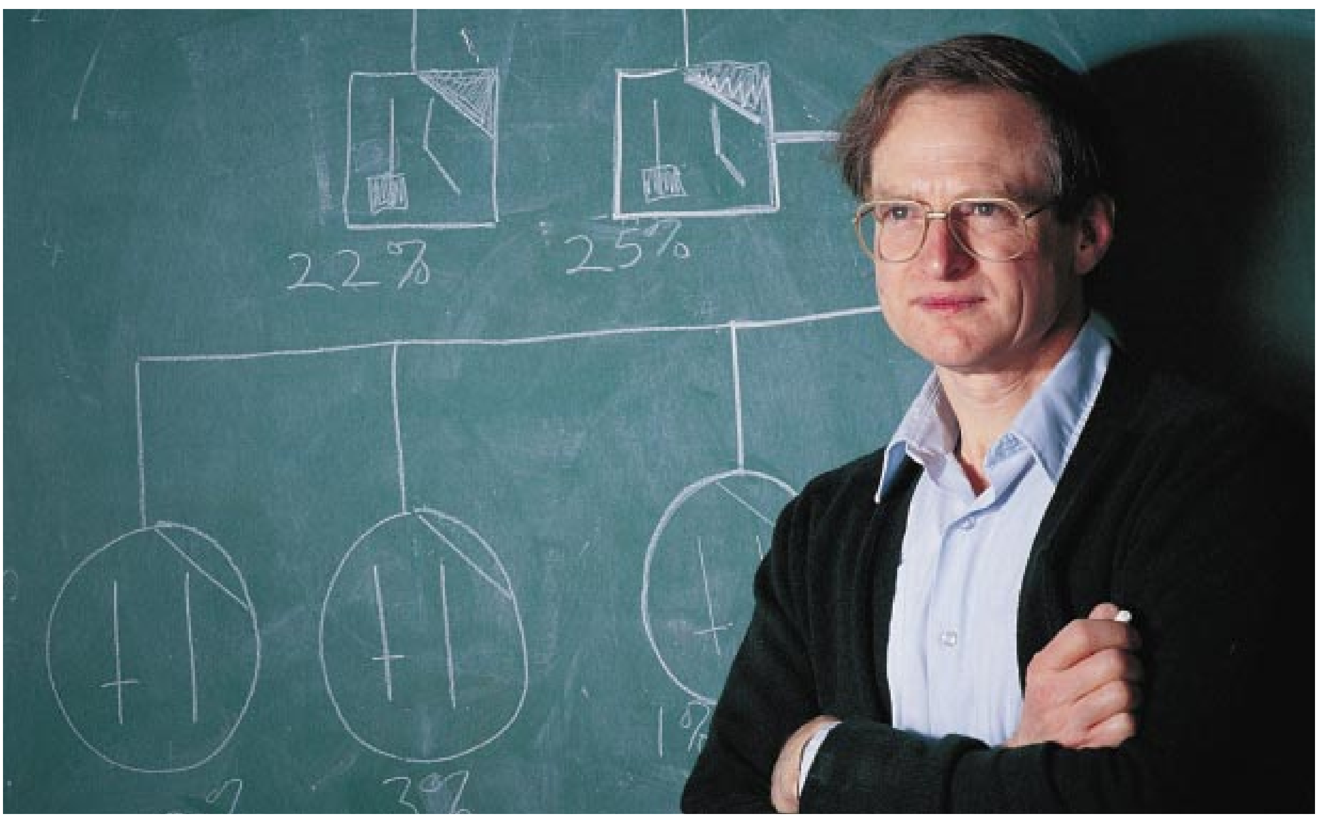
Sapienza hails Laird's analysis as "a real detective-work tour de force." Warren and his colleagues are generally less impressed. They say Laird has amended his theory several times to make it agree more closely with the observed data. Further research may help settle the controversy: Warren is currently trying to insert an expanded repeat region into a chromosome to see whether that is enough to make the *FMR-1* gene less stable.

### Directed Mutations

A still greater controversy rippling molecular genetics these days concerns the effects of the environment on genomes. Radiation, carcinogenic chemicals and other agents are well known for their ability to induce random mutations. Yet a few biologists are investigating the possibility that environmental stresses can sometimes direct the kinds of mutations that occur. The idea is often seen as deeply heretical because it seems to resurrect the great bugaboos of Lamarckian evolutionary models, mutation with intention and the inheritance of acquired characteristics.

The notion that the environment could reshape inheritance fell out of favor at the turn of the century because of the doctrine that the germ-line cells, which make eggs and sperm, are distinct and isolated from the somatic (body) cells. Cells in the immune system and other tissues might change their genes in response to the environment, but the germ cells would not. Yet, Pollard points out, that generalization is at odds with many biological observations. "What people forget," he says, "is that most organisms do not sequester their germ cells."

The sexual tissues of flowers, for example, arise from ordinary somatic cells. For unicellular organisms, the inapplicability of the doctrine is even more obvious. These exceptions raise the possibility that, in at least some species, genetic changes that promote survival might be immediately and preferential-

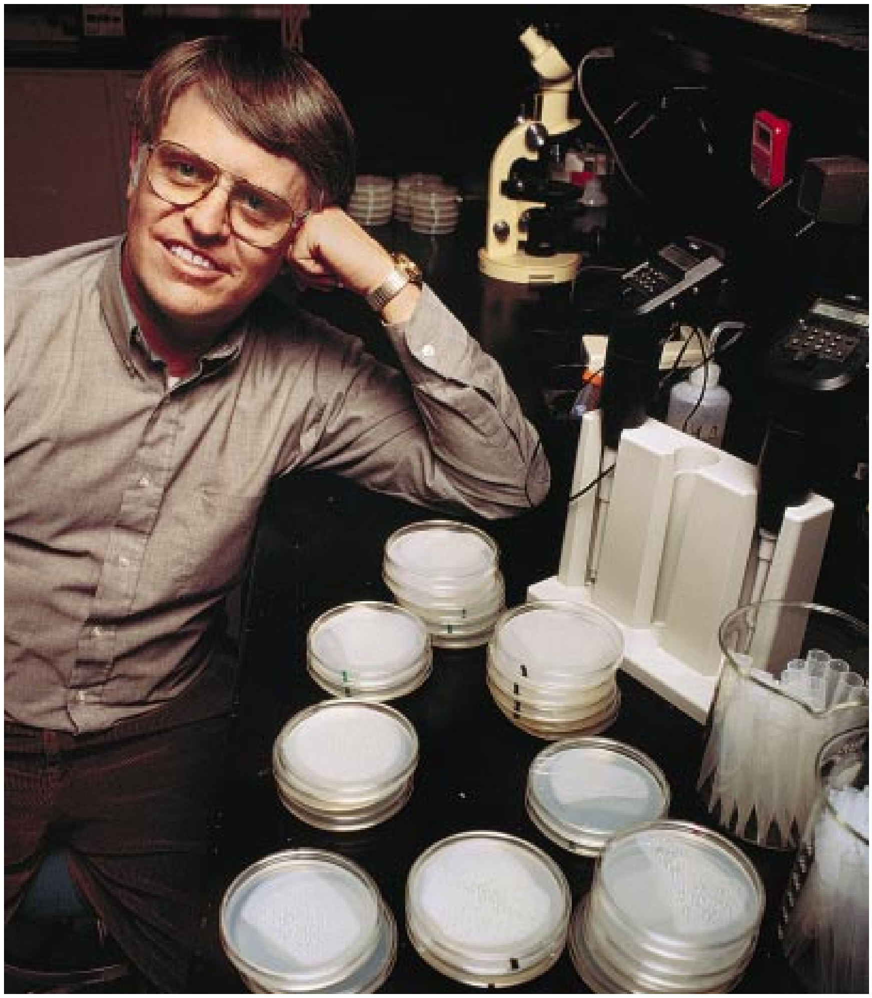



**IMPRINTING THEORIST Charles D. Laird of the University of Washington has proposed that chemical imprints placed on chromosomes by an organism's parents offer the best explanation for the odd genetics of fragile X syndrome.**

ly passed on to an organism's descendants. From a teleological perspective, the organisms would appear to have mutated to fit their environment better.

Much of the current fuss over directed mutations began in 1988 with a paper published by cancer researcher John Cairns, then at the Harvard School of Public Health. Curious about how organisms mutated, he decided to reexamine the genetic tenet that mutations favorable to survival are no more probable than unfavorable ones. Cairns grew bacteria on a medium where the sugar that they could metabolize was in short supply, but another sugar, lactose, was abundant. His results led him to believe that mutations reactivating the bacteria's defective gene for a lactose-digesting enzyme were indeed occurring more often than chance would dictate. The forces of selection seemed to be not merely weeding out unfit organisms but actively steering the mutations in a beneficial direction.

A few months later Barry G. Hall, now at the University of Rochester, presented even more convincing evidence for selection-induced mutations. He performed similar experiments in which hungry bacteria needed two separate mutations—neither of which seemed to confer any benefit alone—to use a new food source. Hall calculated that the odds of both mutations occurring without encouragement were astronomically poor, yet he, too, found that a surprising number of bacteria evolved to fit their surroundings. Several other investigators have also observed selection-induced mutations in bacteria; last summer Hall announced that he had also seen them in yeast.



MUTATION RESEARCHER Barry G. Hall of the University of Rochester thinks environmental stresses may sometimes direct evolution and not just spur it. One-celled organisms may preferentially mutate in beneficial ways.

"Now, I should say what we actually know about the mechanisms: nothing," Hall says. But, he continues, "we do have, I think, sufficient data to reject some of the models that have been put forth." Experiments, he says, have already invalidated Cairns's suggestion that reverse transcriptase was writing the information for beneficial mutations from abnormal RNA into the bacterial DNA.

A far more plausible mechanism is the transcriptional mutagenesis model. It suggests that the DNA in active genes might mutate exceptionally fast because it becomes a single strand—and therefore more vulnerable—during transcription. Unfortunately, Hall says, the evidence has ruled that out, too. Other ideas have been proposed: for example, Hall thinks there is still life in his suggestion that a few cells in a stressed population can enter a hypermutable state. Only the hypermutable cells that subsequently achieve a beneficial mutation would survive; the others would cease to be viable. "That would account for why cells with mutations at other sites aren't found, because those cells die," Hall explains.

Yet no matter how reasonable such hypothetical mechanisms may sound, many scientists are deeply skeptical of directed mutations. "I don't think the evidence is very good for it," remarks Richard E. Lenski of Michigan State University, who has co-authored several papers rebutting Cairns, Hall and others. He maintains that many of their experiments lacked sufficient controls and that estimating the number of mutations responsible for an observed population is tricky. Biochemical phenomena may well bias the mutation rates and create the illusion of directed mutations, as proponents of the idea claim, but Lenski has not yet seen results that persuade him such phenomena are having significant effects.

Yet Lenski sees value coming out of the work on directed mutations. "There does seem to be, in some cases, an increase in the mutation rate when cells are under stressful conditions," he says. "That isn't directed mutation, as has been claimed, but there does seem to be a physiological dependence to certain mutation rates." The fact that scientists are scrutinizing the potential molecular mechanisms and evolutionary consequences of such influences is a useful development, he observes.

### Editing RNA

Novel mutations and other quirks in the chemistry of DNA are not solely responsible for the layers of complexity being added to molecular genetics. Indeed, they cannot be considered apart from the accompanying discoveries that researchers have been making about what occurs during the transcription and translation of the genetic message into proteins. Contrary to early theories, the RNA molecules transcribed from the DNA of genes must often undergo extensive chemical processing before they are ready to serve as the messengers for protein synthesis.

In all organisms more complex than bacteria, for example, the instructions in a primary RNA transcript for making a protein are interrupted by long sequences that are essentially meaningless. During RNA processing, these "intron" sequences are chopped out of the molecule, and the meaningful "exons" are spliced together to make a shorter, more coherent molecule. In 1982 Thomas R. Cech of the University of Colorado and Sidney Altman of Yale University made the Nobel Prize–winning discovery that some intron sequences in RNA had a built-in enzyme that allowed them to perform these cut-and-splice operations themselves.

Perhaps the most extraordinary form of processing is RNA editing, in which crucial information that is not clearly specified in the DNA is added to RNA molecules. Properly speaking, RNA editing embraces several different phenomena occurring in organisms as diverse as mammals, amphibians, plants, protozoa and viruses. In all its variations, RNA editing involves the specific addition of bases to RNA molecules or the transformation of certain bases within them. More than one mechanism is probably at work. "The term 'RNA editing' in its general sense has been used loosely," notes Kenneth D. Stuart, who investigates the phenomenon at the Seattle Biomedical Research Institute.

The best understood instance of RNA editing—and possibly the most startling—came from studies of trypanosome parasites, which cause sleeping sickness and other illnesses. The mitochondrial DNA of trypanosomes has a unique structure: it consists of several dozen large loops called maxicircles and thousands of smaller ones called minicircles. All the loops interlink into a network that resembles a sloppily made piece of chain mail.

The oddness of this DNA did not end with its structure. The minicircles seemed to contain no useful genetic information at all. Biologists eventually determined that the maxicircles carried most of the genes found in the mitochondria of other organisms. Yet crucial genes, such as those that make transfer RNA, were missing.

Molecular biologists became even more baffled when they looked at the messenger RNAs: many of the molecules were longer than the DNA from which they had been transcribed. One strand of RNA was more than twice as long as its DNA counterpart. Uridine bases had been strategically added throughout those RNAs, making sense out of what at the DNA level had been nonsensical "cryptogenes," a term coined by Larry Simpson of the University of California at Los Angeles.

In 1990 Simpson, Beat Blum and Norbert Bakalara offered an explanation. Both the maxicircles and the minicircles produce small "guide RNAs" that find and correct omissions in the messenger RNA. A guide RNA and the messenger strand to which it anchors fit together like the halves of a zipper. Wherever the messenger RNA is missing a uridine, a misalignment occurs. At those mismatch sites, a uridine is spliced into the messenger sequence. "RNA editing actually creates the mature messenger RNA sequence," Stuart says.

Why such a system would have evolved is still a mystery, but one possibility is that it affords the parasites control over the expression of genes during different parts of their complex life cycles. How the guide RNAs direct the cut-and-splice activity of their editing is also still not completely understood. One model, first published by Cech and independently conceived by Blum and Simpson, holds that the guide RNAs may perform the operation themselves, donating uridines from their own strands in the process. Last April, Stuart and his team reported that they had found guide RNAs bound to the messenger molecules, a crucial piece of evidence in favor of their theory.

### The Not-So-Universal Code

After the splicing, editing or other processing is complete, a messenger RNA is ready to be translated into protein by the ribosomes. Even that step, however, turns out to be more complex than was once thought. To perform the translation, a ribosome must decipher the genetic code in the RNA. Each codon, or three-base sequence in the RNA, corresponds to one instruction, such as to add a particular amino acid or to stop lengthening the protein.

One might imagine that because the genetic code is so fundamental to life, it would be universal. Not so: the code used to decipher RNA from the nucleus is slightly different from that for RNA in mitochondria and chloroplasts. In mitochondrial RNA, for example, the codon adenine-guanine-adenine is an instruction to add the amino acid valine to a growing peptide chain, but in nuclear RNA it is a stop codon marking the end of the coding sequence.

Many experiments have shown that the codes are sometimes flexible. To make some proteins, a ribosome must alter its translation of certain codons within a single strand of messenger RNA. Investigators have shown that ribosomes making a protein in mammalian blood interpret certain stop codons—but not all—as instructions to add the amino acid selenocysteine.

Moreover, the meaningful codons in messenger RNA are not always side by side. Ribosomes must occasionally "frame shift," skipping forward or backward over one or more bases, to find a codon. In one case identified by Wai Mun Huang of the University of Utah Medical Center, ribosomes ignore a sequence of 50 bases in the messenger molecule. John F. Atkins of University College in Cork, Ireland, and his colleagues have proposed the name "RNA recoding" for this reinterpretation of messenger RNA by ribosomes.

RNA recoding and processing, genome imprinting of the chromosomes and all the new varieties of mutations are only a few of the phenomena now invigorating molecular genetics. The emerging picture of a dynamic genome in no way denigrates the more traditional model, which is itself a remarkable intellectual achievement. Geneticists sought and found rules that applied well to the inheritance of characteristics in organisms as diverse as bacteria, roses, giraffes and humans. "We should feel lucky that we can find any generalizations at all," Lederberg remarks. "It's unlikely that we're going to find any principles that are in absolutely universal application."

Indeed, it is the simplicity of traditional genetics that has endowed it with such power. Its generalizations do describe most genetic phenomena in most organisms most of the time. Without them, biology could never have progressed to its current understanding of life. The great challenges for molecular geneticists now and in the future will be to explore the implications of the exceptions and to find new sets of even more potent rules—if any exist.


FURTHER READING

MACROEVOLUTION AND RETROVIRAL ELEMENTS. John F. McDonald in *Bioscience*, Vol. 40, No. 3, pages 183–191; March 1990.

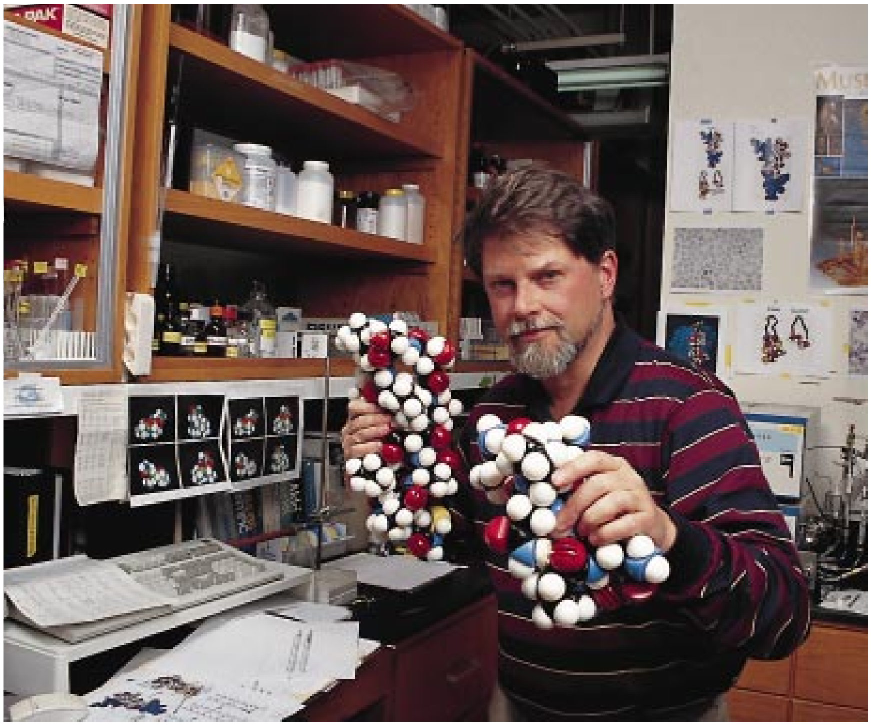
PARENTAL IMPRINTING OF GENES. Carmen Sapienza in *Scientific American*, Vol. 263, No. 4, pages 52–60; October 1990.

THE DYNAMIC GENOME: BARBARA MCCLINTOCK'S IDEAS IN THE CENTURY OF GENETICS. Edited by Nina Fedoroff and David Botstein. Cold Spring Harbor Laboratory Press, 1992.

TRANSPOSABLE ELEMENTS AND EVOLUTION. Special Issue of *Genetica*, Vol. 86, Nos. 1–3; 1992.

# Back to Basics

*Mapping malaria's genome may help produce a vaccine*



JAMES ARONOVSKY

***ARNOLD C. SATTERTHWAIT of the Scripps Research Institute models two synthetic peptides that one day might be used as part of a multistage vaccine for malaria.***

In humanity's age-old battle with malaria, the parasites retain the upper hand. As the mosquito-borne *Plasmodium* protozoa pass through the many stages of their life cycle, they adeptly change their structure to evade the immune system. Thus, for decades malaria has defied efforts to develop a vaccine while simultaneously evolving resistance to existing drugs. The disease is now raging out of control in Southeast Asia, South America and especially Africa. "In sub-Saharan Africa alone, between 1.5 and two million children are dying each year of malaria and its consequences," says Kirk Miller, director of the U.S. Agency for International Development's malaria vaccine research program, whose $9.6-million budget makes it the world's largest.

"What they've been looking for, for nearly 20 years, has been a quick kill—the 'magic bullet' that's going to provide a stable vaccine for malaria," says John B. Dame, an associate professor in the department of infectious diseases at the University of Florida. "I don't think that's going to work. It's a bit shortsighted not to invest a little more in basic understanding of the organism."

Obtaining that kind of knowledge is exactly what Dame and his colleagues at the University of Florida's Interdisciplinary Center for Biotechnology Research hope to achieve: they are sequencing all the genes of the malaria parasite. "You need to know the enemy in order to attack it effectively," says Sheldon M. Schuster, who directs the Malaria Genome Project.

The Florida researchers believe the genetic sequences and maps they generate, disseminated freely through journals and on-line data bases, will provide the intelligence needed to design drugs and vaccines that are effective against the various stages in the life cycle of the parasite. The team deliberately chose to sequence the most dangerous strain of malaria: *P. falciparum,* a more virulent but historically less common variety that is the only one of the four malaria species that can kill humans and that now accounts for 90 percent of the malarial infections in Africa. "The project will provide a large number of genes of interest; virtually anything you want from the parasite will be available," Dame says.

Dame estimates the malaria genome contains between 5,000 and 10,000 active genes. At its current pace, the team could sequence the ends (the outer 300 bases) of 5,000 genes in about two years. Because each successive gene is picked at random, the pace will slow as redundancies become more common. But the group is convinced its data will provide clues to intervening in the parasite's complex life cycle. "So far, of the 1,000 or so genes we have sequenced, 125 resemble something in genome data bases. That means about 90 percent of what we're seeing is either unique to *Plasmodium* or new to science," Schuster says.

Unfortunately, though, funding for the project—$450,000 over two years from the University of Florida—runs out at the end of this year. Finding a new sponsor is proving difficult. Flagging interest by drug companies has put more pressure on a dwindling supply of government money, forcing malaria researchers to focus on the bottom line. "Finding somebody in the industrial sector who will work on malaria is critical," says W. Ripley Ballou, who heads the department of immunology at Walter Reed Army Institute of Research.

"Progress hasn't been rapid enough to sustain private-sector investment," adds Carlos C. Campbell, chief of the malaria branch at the Centers for Disease Control. "The profit margin on any malaria vaccine is going to be like the profit margin on any malaria drug: low. Unfortunately, the greatest need is in the places in the world that have the least ability to pay."

Nevertheless, Dame is hopeful. "The challenge is that most of the funding agencies haven't recognized the value of a genome project," he explains. "But the paradigm shift in biology is beginning to sift down slowly to where the possibilities of this kind of approach are taken more seriously."

Although efforts by the Agency for International Development and Walter Reed emphasize more short-term goals, researchers at both are beginning to agree. Ballou admits that the genome project could speed the realization of "one of our hopes: going in and genetically manipulating the parasite so that it's no longer virulent but induces an immune response that is protective."

Even with their admittedly sketchy understanding of the protozoan, researchers at Walter Reed and elsewhere have been able to design several subunit vaccines. A subunit vaccine typically comprises three parts. The central antigen is taken from, or synthesized to look like, a vulnerable part of the parasite in order to produce antibodies and killer *T* cells that will disable the parasite. Attached to this section are a carrier protein, which stimulates the helper *T* cells produced by the recipient's immune system, and an adjuvant, which helps the immune system muster sufficient legions of antibodies.

The trick is to pick the right protein fragments, called peptides, to use as antigens and to get the immune system to recognize them as hostile. "The stumbling block that everyone in this field has run into is that these proteins are not very immunogenic, so although they look very promising in animal models, the response to them in humans is not very profound," Ballou says. "This led us very early on to search for appropriate adjuvants."

The pool of adjuvants from which to choose is small but growing slowly. Walter Reed is testing two new adjuvants with its vaccines: one is based on a detergent from the bark of the *Quillaja saponaria* tree and developed by Cambridge Biotech in Worcester, Mass.; the other is a fatty compound called monophosphoral lipid A, made by RIBI ImmunoChem in Hamilton, Mont.

Ballou thinks the RIBI adjuvant is particularly promising. Investigators at SmithKline Beecham have combined the lipid A compound with liposome delivery molecules and two different synthetic antigens to make two vaccines. The vaccines act against the sporozoite stage, which is transmitted by infected mosquitoes. "We're getting extremely good antibody responses, and there will be a challenge trial in the beginning of the year at Johns Hopkins to look for efficacy" of the first vaccine, Ballou says. The second, which combines antigens for malaria and hepatitis B to produce a dual vaccine, protected only two of eight volunteers. Ballou expects to test another version early next year.

Walter Reed also plans field trials of what may be the most promising malaria vaccine yet derived. Its active agent is a synthetic peptide developed several years ago by Manuel E. Patarroyo of the Hospital of San Juan de Dios Institute of Immunology in Bogotá, Colombia. Patarroyo claimed his vaccine protected 70 to 80 percent of the 30,000

## Preventing the Preventable

An alarming 1991 survey by the Centers for Disease Control found that less than half of school-age children in nine major U.S. cities had been fully vaccinated against infectious diseases by their second birthday. The results of that deficiency became tragically clear in 1989 and 1990, when an epidemic of measles swept through several large cities, killing more than 100 children. Three fourths of the 45,000 children affected were not vaccinated. "We have preventable diseases that are not being prevented," warns Vincent A. Fulginiti, dean of Tulane University School of Medicine.

As head of the National Vaccine Advisory Committee, Fulginiti hopes to motivate the government into action. In March the committee plans to unveil its National Immunization Plan, commissioned by Congress in 1986, to ensure that by the year 2000 at least 90 percent of U.S. children will get their shots on time. The advisory group will likely find strong supporters in the Clinton administration. "I want all of us to get involved to get immunizations in our two-year-olds up to par," declares Surgeon General–designate M. Joycelyn Elders. "As a country, we should be ashamed of where we are."

Ironically, the committee assigns most of the blame for this failure of the health care system to complacency produced by the success of the U.S. immunization effort. In developing countries, where serious infectious disease is more common, immunization often commands more attention. The group concedes, however, that U.S. standards are higher, making "full immunization" a tougher goal. In seven of the nine cities surveyed by the CDC, for example, more than half of the children received all but the fourth dose of DTP (a mixed diphtheria, tetanus and pertussis vaccine) on schedule.

Still, many of the nine diseases for which vaccines are currently available—measles, mumps, rubella, diphtheria, tetanus, pertussis, polio, hepatitis B and *Hemophilus influenzae*—pose significant threats in the U.S. "The measles epidemic has now waned," Fulginiti says, "partially because of increased immunization and partially because the supply of susceptible children has been exhausted." But in recent years rubella and pertussis have also resurged.

"The main problem is missed opportunities, failure to track kids," says Walter A. Orenstein, director of the CDC's immunization branch. "It's not as if our kids exist in a vacuum. Ninety percent of the children in the U.S. start immunization. The problem is that they don't stay with it." Orenstein believes underimmunized children are risking more than just nine diseases. "Immunization is a basic preventive service," he observes. "If kids aren't getting vaccinated, they probably aren't receiving much other preventive health care either."

The committee wants to change that by establishing a computerized national immunization registry that would automatically be used to remind parents when their kids are due for boosters and to pinpoint underimmunized communities. The Robert Wood Johnson Foundation has awarded grants to 23 cities to set up local registries, as well as a $5-million grant to the CDC to work on integrating such registries into a national system. "We hope to elevate immunization records to the same status as a birth certificate or a passport," Fulginiti says. "With our mobile society, we have no uniform way of knowing what children received and when they received it."

The plan also calls for increased funding for new vaccine development and regulatory changes to make immunization more "user-friendly." The committee hopes to combine as many vaccines as possible into single shots, reducing the number of office visits necessary to complete the basic series. It also recommends adding vaccines against winter diarrhea, respiratory viruses, bronchial pneumonia and hepatitis A to the standard complement as soon as possible, although some of these vaccines are still being developed.

"It might take changes to insurance legislation to make sure children are covered by the private sector," Fulginiti adds. Half of all vaccines are given by private doctors, even though privately administered shots are twice as expensive (about $250 per child) and are covered by less than half of conventional health insurance plans.

The committee admits that many of its proposals will cost money—about $50 million, it estimates. But Fulginiti believes Congress will embrace at least some of the goals and fund them appropriately. He has a compelling sales pitch: "The biggest advantage of immunization is that for every dollar spent, we save at least $10 in the care it would otherwise take for those children who would be affected." *—W. Wayt Gibbs* volunteers who received it in field tests. Neither the vaccine nor the trials met Food and Drug Administration standards, however; according to Ballou, “all the trials had serious flaws—no one buys that 80 percent number.”

Recent tests in Colombia, these better controlled, showed a 30 to 60 percent protection rate, still significantly higher than that of any other malaria vaccine. Walter Reed contracted Multiple Peptide Systems in San Diego to reproduce the formula. Preliminary clinical trials, Ballou says, have shown the product to be safe.

Because Patarroyo’s vaccine works only against *falciparum* merozoites, a stage of the parasite that destroys the red blood cells of infected individuals, its failure might prove lethal to volunteers. Challenge trials must thus be conducted in the field using subjects already at risk. Such people are alarmingly easy to find. In Africa, Campbell notes, “we’re working in an area where by the age of six to eight months more than 70 percent of all infants will have experienced at least one malarial infection.”

Because malaria is so infectious and quick to evolve defenses against new drugs and vaccines, a combination punch will be needed to deal the disease a serious blow. “The ideal vaccine would be a multistage vaccine that attacks the parasite at different points in its life cycle,” says Arnold C. Satterthwait, a biochemist at the Scripps Research Institute in La Jolla, Calif.

Such a vaccine might induce an antibody isolated by David C. Kaslow of the National Institutes of Health. “This antibody blocks the development of malaria in the mosquito’s stomach; in other words, you prevent infection of mosquitoes,” Satterthwait explains. “It has been tested in vivo for transmission blocking; it’s 100 percent effective.”

As part of a multistage vaccine, the antigens that stimulate such antibodies could prevent any parasites that develop resistance from reproducing. To that end, Satterthwait and his colleagues Enrico A. Stura and Angray Kang are mapping the three-dimensional structure of the antibody and trying to synthesize an antigen to match.

An effective multistage vaccine for malaria remains a long way off. But if the new genetic information can be combined with new biochemical techniques and clinical experience, there is room for optimism. “Malaria as a public health problem remains catastrophic and is getting worse,” Campbell points out. “Some alternative technology is going to have to come along just to keep things under control, let alone to reduce the problem.” —*W. Wayt Gibbs*

## No Snake Oil Here

*But researchers are finding drugs in frogs, moths, sharks....*

Jay Moorin, the president of Magainin Pharmaceuticals, wants to put a little frog or a dash of shark into your medicine cabinet. In the company’s laboratories in Plymouth Meeting, Pa., substances extracted from a variety of animals, such as sand sharks and African clawed frogs, are showing promise as drugs for treating infections, healing wounds and killing some tumors. “We look for pharmaceutically active compounds in the host defense systems of animals, including man,” he says. Around the world, researchers are finding that many animals have disease-fighting secrets to share.

One of the first pharmaceutical markets in which the new animal-derived compounds may make themselves felt is antibiotics. Penicillins and most other antibiotics available today are byproducts of fungi or other unicellular organisms. Unfortunately, bacteria demonstrate maddening resilience: given sufficient time, mutant strains resistant to drugs almost inevitably evolve. To stay ahead of the resistance problem, the pharmaceutical industry is always looking for new antibiotics—preferably ones as different as possible from those that have lost their effectiveness.

The animal-derived products may fill that bill. Most of them are peptides, or small proteins, that act completely unlike traditional antibiotics and antifungal agents. Many can kill a wide array of pathogenic microorganisms. Best of all, because they extend the chemical defenses that animals have been using for millions of years, they may be more durable remedies.

Of these compounds, magainins (mah-GAY-ninz) are furthest along in industrial development. They were discovered in 1986 by Michael A. Zasloff, who was then a geneticist at the National Institutes of Health. In the course of his experiments, Zasloff would remove the ovaries from African clawed frogs, sew up the incisions and return the animals to the murky tanks in which they lived. Like many scientists before him, he wondered why the incisions always healed quickly and never became infected de-

JOHN McGRAIL

***ANTIBIOTIC RESEARCHER Michael A. Zasloff of Magainin Pharmaceuticals is turning defensive proteins from the skin of frogs into novel drugs.***

spite the abundant bacteria in the water. Unlike most, he tried to find out. He eventually isolated two molecules that had disinfectant properties in the skin of the frogs and named them after the Hebrew word for shield. Zasloff left the NIH a year later to found Magainin Pharmaceuticals.

By toying with the chemical structure of those original molecules, researchers have expanded the number of magainins to about 2,000. "Unlike current antibiotics, which essentially work on the inside of bacteria, magainins punch holes in the bacteria's cell membrane," Moorin reports. "The cell then lyses and dies." He adds that in laboratory studies, bacteria have not yet developed any resistance to magainins.

Magainin Pharmaceuticals has recently improved its technique for manufacturing the compounds. "The frogs of the world are safe," Moorin jokes. "There's no press in the back room." There never was: chemists had always synthesized the molecules, which was a slow and expensive process. Last November the company announced that it had learned how to use genetically altered bacteria to make magainins. "People had thought, 'It's an antibiotic, so maybe it will kill the host organism,' but that's not the case," Moorin remarks.

The magainin product closest to readiness is MSI-78, a broad-spectrum antimicrobial agent. According to Moorin, the product recently completed Phase I human clinical trials. If it passes all safety and efficacy tests on schedule, it might reach pharmacies in about three years. Initially, MSI-78 will be promoted as a disinfectant for skin ulcers, but Moorin expects eventually to position it as a treatment for accelerating the healing of wounds—a larger and more lucrative market. Magainin Pharmaceuticals is also working with Colgate-Palmolive, Moorin says, to create products for oral health care. "Magainins or a magaininlike substance could be added to toothpaste or mouthwash to control plaque," he suggests.

Magainin Pharmaceuticals hopes to get a piece of the cancer therapeutics market as well. "It's our theory that there are patches in certain tumors that look like bacterial cell membranes, and that's why magainins work on some cancers and not on others," Moorin explains. He claims that in tests on animals, the compounds have shown effectiveness against colon cancer, for example, but not against ovarian cancer.

Last December, Magainin Pharmaceuticals and Sandoz Pharmaceuticals struck a deal to investigate the anticancer potential of magainins. Under the terms of the 30-month agreement, Sandoz can develop any compounds it wants for treating cancer, and Magainin Pharmaceuticals gets to keep the antimicrobial market for itself.

Magainins are by no means the only animal-derived antibiotics, nor were they the first known. In the late 1970s Hans G. Boman of Stockholm University extracted peptide antimicrobials from *Cecropia* moths and named them cecropins. Patents for cecropins are now held by Bioscience Center, a subsidiary of Kabi Pharmacia.

During the early 1980s, a group of researchers at the University of California at Los Angeles, including Richard I. Lehrer, Tomas Ganz and Michael E. Selsted, gave the name defensins to a class of peptides sharing a common general structure. The molecules are produced by cells of the immune system that engulf invading pathogens; the defensins destroy the invaders by forming pores in their outer membranes.

A more remarkable discovery, however, is that defensins are also produced by some of the epithelial cells that line moist body surfaces in mammals. Five years ago, while studying the development of mice, Andrew J. Ouellette of the Shriners Burn Institute in Boston accidentally discovered defensins in their small intestines. Selsted, who is now at the University of California at Irvine, and Charles L. Bevins of Children's Hospital of Philadelphia have also independently found the same types of molecules in humans.

"The colon is typically chock-full of bacteria, but the small intestine is relatively sterile," Bevins notes. "Our working hypothesis is that these peptides help to maintain that sterile environment." Defensins and similar peptides may therefore be the body's "first line of defense" against bacteria, fungi and viruses. In keeping with that hypothesis, Bevins says he and his collaborators have found more antimicrobial peptides—similar to defensins but still distinct—in the epithelial lining of the lungs of a cow.

Selsted cautions that the pharmaceutical potential for defensins, magainins or any other similar peptides is likely to hinge in part on the delivery system. Because some peptides may have high toxicity, it may be necessary to cloak them in small, fatty bubbles called liposomes to regulate their release. He says his group is at present working on a project sponsored by the NIH to develop novel antifungal agents. It will focus on three sets of peptides: defensins; a family of similar molecules called beta-defensins; and indolacidin, a unique molecule his team isolated from the bovine immune system. "There's a lot of interest in this work in the private sector, and we're currently negotiating with a couple of companies about forming some sort of partnership," Selsted says.

Perhaps the most unusual addition to the list of animal-derived antimicrobials is squalamine, which Magainin Pharmaceuticals discovered last year. Squalamines, which circulate in the blood of sharks, are steroids, not peptides. "It's widely known that sharks rarely get infections and almost never get cancer," Moorin says. The first paper on squalamine is scheduled for publication in the *Proceedings of the National Academy of Sciences.* Moorin notes that in early tests, squalamine has shown exceptional usefulness as an antibiotic and antifungal agent. Preliminary plans for the compound are to develop it as an oral antifungal. Please, no man-eating-shark jokes. —*John Rennie*

## High on Gravity

*Increasing the g's grows crystals with few defects*

A large, blue-and-red centrifuge in a research laboratory at Clarkson University in upstate New York is putting a new spin on crystal growth. There and at other laboratories, materials scientists are finding to their surprise that solids sometimes form extremely well under high-gravity conditions. In fact, says William R. Wilcox, a director of crystal growth research at Clarkson, "the crystals were better than what grew in space."

Perfect crystal growth for semiconductors and other applications is one of the first reasons cited to justify the billions of dollars allocated for space stations and satellite laboratories. In the microgravity of near-earth orbit, the molten material, or melt, crystallizes with fewer bubbles, defects and dislocations than it does at the field strength that prevails on the earth. But sending furnaces into orbit is an expensive proposition. In addition, crystals grown on space shuttle missions, while of higher quality than those grown on the earth, have usually fallen below expectations.

Now it seems investigators may have set their gravitational sights low when they should have been aiming high. Normal gravity wreaks havoc by yanking on the colder, and therefore denser, regions of the melt more strongly than it does on the warmer regions. The result-

FOREST McMULLIN

***LIYA L. REGEL sits behind Clarkson University's new centrifuge, originally a boring mill designed to machine circular parts. The furnace used to grow crystals is mounted on the end of the centrifuge arm—a red, 10-foot-long I beam.***

ing convection causes defects in the solidified product. Yet counterintuitive evidence has been trickling in that strong gravity, or "high-g," forces can produce bigger and better crystals.

No one is sure exactly why high-g forces produced by centrifugation create superior crystals. For specimens grown between one and 10 g's, the centrifugal acceleration appears to play a critical role by stabilizing the convective flow of the melt as it solidifies.

Researchers have observed that uniform crystallization occurs at discrete g levels, which depend on the substance being solidified. Liya L. Regel, who directs a new international center that Clarkson set up to explore high-g crystal growth, calls these levels "magic g" values. Besides yielding superior crystals, centrifugation can alter the microstructures in alloys, improving their electrical properties. The process can also raise the transition temperature of high-temperature superconductors.

Evidence of improved crystallization at high g's has been around for more than 25 years. Paul J. Shlichta, who heads Crystal Research, a small company pursuing the technology in Olympia, Wash., noted that Allan E. Carlson, a solid-state physicist with the now defunct Clevite Corporation in Cleveland, described the importance of convection on crystal growth from solution at a conference in 1958. "The problem was that Carlson moved the session out to the lawn, and no one could hear him," Shlichta says.

In addition, much of the experimental work, which had been done in Germany, the Soviet Union and France, received little attention from the English-language press. In the mid-1980s Georg Müller of Friedrich Alexander University in Erlangen, Germany, found that centrifugation helps to limit the formation of striations in semiconductor crystals. Huguette Rodot of the CNRS laboratory in Meudon, France, and Regel, who at the time was working at the Space Research Institute in Moscow, discovered similar effects. A May 1991 workshop on materials processing at high gravity, held in Dubna, Russia, drew many investigators of high gravity phenomena together for the first time.

Another barrier that kept high-gravity growth from moving forward was access to appropriate centrifuges. Without a dedicated machine, Shlichta essentially panhandled his way to centrifuge time. While working in Moscow, Regel managed to squeeze a week out of the Gagarin Cosmonaut Training Center. The equipment, too, needed to be specially crafted. "There were a lot of destroyed ampoules and heaters because of the centrifugation," Regel says of initial experiments. The new centrifuge at Clarkson, which can generate about 15 g's, should prove the remedy for itchy experimental fingers. Although it was purchased used and needed modifications, the centrifuge will be the first in the world dedicated solely to materials processing. Regel expected the initial results to be in by mid-February.

Success in the field has begun to pique entrepreneurial curiosity. Companies involved in aerospace and semiconductor technology have inquired about the Clarkson work. For the moment, however, widespread commercial production of such crystals may not be entirely practical. Making crystals of, say, gallium arsenide for the manufacture of semiconductors could take an entire day, Regel explains. A melt of gallium arsenide would be loaded into a furnace hung at the end of the centrifuge arm. The machine would then spin the furnace for eight to 12 hours until the melt solidified. Furthermore, there is a limit to the size of the furnace that can be used.

Meanwhile Shlichta has been experimenting with a much faster method of growing crystals. He uses a small, high-speed centrifuge to generate ultrahigh-gravity forces—thousands to hundreds of thousands of g's—to produce crystals from solutions. The kilogravity forces "literally pull the solute out of the solvent," Shlichta explains. At about 41,000 g's, the centrifuge produces remarkably clear crystals of lead nitrate. This clarity gives Shlichta confidence that the procedure will find applications in optics.

Despite its promise, high-gravity crystal growth has melted the hearts of only a few funding agencies. The crystal-growth effect runs counter to conventional thinking. "People ask, 'Are you sure you didn't make a mistake in the experiment?' The results do not fit in with people's worldview, so they tend to ignore it," Wilcox explains. The research situation has been even less hospitable to Shlichta. "I bombed in getting funding," he laments. "The agencies did not see any immediate advantages" to kilogravity centrifugation.

That may change soon. The centrifuge at Clarkson and a second international meeting, scheduled to be held at the university in June, should stimulate more work in this field. Although the high-g researchers do not think centrifugal crystal growth will replace micrograms work—the crystals grown in each realm seem to have slightly different characteristics—they are bullish about the future. Going around in circles could actually prove to be a productive enterprise. *—Philip Yam*

# Headsets

*Television goggles are the vision of the future*

The folks who liberated music lovers from their living rooms and cars by coaxing high-quality stereo sound from tape players the size of a block of tofu now plan to free the couch potato. The latest gadgets from the consumer electronics industry reduce the boob tube to what looks like a pair of large, wraparound ski goggles but provide the viewer with eye-filling images that rival big-screen televisions.

In an apparent upset, a small U.S. firm will reach the marketplace first with a set of electronic eyeglasses that allow die-hard television lovers to walk while they gawk. The tiny entrepreneurial firm outpaced giant Sony, which has demonstrated a more sophisticated contraption that is still a year or so away from its commercial debut.

The television from Virtual Vision in Redmond, Wash., bears a striking resemblance to Arnold Schwarzenegger's sunglasses in *The Terminator.* The unit receives television broadcasts that are projected onto a small mirror in the bifocal area of one lens, enabling the viewer to see the real world simultaneously with both eyes. The wearer has to go through a cumbersome setup procedure before purchase to determine which eye is used most. People without a "dominant" eye cannot use the glasses.

The product was introduced at the Consumer Electronics Show in January and will be sold at stores such as The Sharper Image for $900 beginning in April. The set's designers, it turns out, did have Arnold in mind. Market research on the goggles determined that men walking around with television sets on their heads should not look like they are wearing Beanie-copters. "It shouldn't mess up their hair," says Brian H. Durwood, Virtual Vision's vice president of marketing. Durwood says market studies showed that most users would be "career-oriented males who are mechanically oriented and vain."

Patents for the Virtual Vision glasses were licensed to the company by Peter Purdy, a Seattle-area inventor who wanted a digital speedometer that could be projected onto the inside of his sunglasses for downhill skiing. Purdy worked with Thomas A. Furness, one of the early developers of virtual reality, the technology that permits a viewer to manipulate a video world created by a computer. But Sony's offering, which it calls Visortron, more closely resembles the head-mounted displays used in virtual reality. The Sony headset supplies an image that occupies the full visual field of both eyes, so it may one day be able to display three-dimensional images.

Visortron is the work of a design team at Sony's headquarters in Tokyo, which for six years has been pursuing the approach that served the corporation with products such as Walkman and Discman. The head-mounted television monitor comes with built-in headphones and weighs a little more than half a pound, far less than the bulky helmets currently used for computer-generated virtual-reality research. A small, portable videocassette recorder attached by a cord supplies the video and audio signals. "We have developed the unit now, since current technology allows us to create a high-quality picture," says Daiji Takahashi, who led the development effort.

Both companies claim that their products can generate an image that matches that of a large-screen television. The Virtual Vision television displays 96,000 pixels and weighs about five ounces. An image from a liquid-crystal display (LCD) is projected onto a reflective lens that produces an image equal to viewing a 60-inch picture from a distance of eight to 15 feet. A cable runs from the glasses to a belt pack where the battery, tuner and antenna reside.

**HOMO SPECTANS** ***can walk and chew gum while watching "The Simpsons."***

In Sony's system the television images are shown on two backlit, 0.7-inch liquid-crystal panels, one mounted in front of each eye. The color LCDs, which were borrowed from the viewfinders in the company's camcorders, have the highest resolution for their size, 103,000 pixels, equivalent to four-inch LCD screens. A pair of mirrors and simple injection-molded plastic lenses blend the left and right images into an apparent single screen that covers about 30 percent of the user's field of view, comparable to watching a 33-inch television from a distance of four feet.

Using the Sony gadget is not as simple as grabbing a remote control and pushing the "on" button. Eyeglasses have to be removed, the headband has to be adjusted and, once the device is mounted, the user must follow a careful procedure to position the LCDs and to focus them in front of each eye.

In the longer run, the biggest untapped potential of the new technology may be the ability of its dual screens to make projections of three-dimensional imagery, which has not been practical in consumer television markets. (Such programming would still have to be produced using special cameras to record separate stereoscopic left and right images.) Indeed, Sony engineers are already testing a 3-D system that can play on a conventional VCR. Instead of putting two completely separate tracks on the tape, the extra dimension is added by alternating the left and right video frames. Although this has the effect of halving the "refresh rate" of each LCD from 60 to 30 times a second, Takahashi says this technique fools the senses into seeing two separate images with the requisite parallax.

With only monocular vision, such "true" virtual reality is an impossibility with the Virtual Vision glasses. But being able to see other things while watching the television images has its advantages. Although jogging while taking in the news is possible, the company recommends its product for more sedentary activities, such as watering the grass or watching an instant replay while seated at a football game. Future versions might make their way into the workplace. With these glasses, a technician could read a projected diagram of how to assemble a part, or a surgeon might be able to glance at a chart without the need for head movements.

Sony plans to introduce Visortron in one to two years—and it is already facing more competition. William Johnson, a British inventor who previously devised a shoe that records a runner's time, announced last fall a television product he calls Goggle Vox.

In the meantime, Sony is proceeding with caution. Two medical departments at Japanese universities have been commissioned to study the effects of extended viewing. So far no problems have been encountered. Although marketing plans and pricing have not yet been set, Sony hopes that Visortron will make its debut on airlines for in-flight movie viewing. Then, the person sitting next to you wearing shades may really be watching *Terminator 6.*

*—Gary Stix with Tom Koppel, Tokyo*

## More Profitable to Give Than to Receive?

Every year billions of dollars in aid flow from rich nations to poor ones—and right back again. Grants and loans pay for machinery and equipment imported from the industrialized world, construction overseen by U.S., European or Japanese engineering firms, consulting work done by international aid organizations and, not least, interest payments on previous loans. Meanwhile many of the nations that receive these enormous amounts of aid remain poor.

Development economists will immediately point out that this situation (except for the intractable poverty) is exactly as it should be. After all, if a country's needs can be met by its own organizations and businesses, then it does not need help from abroad. Outside assistance should go to providing those goods and services that a country cannot provide for itself.

Furthermore, aid comes in the form of hard currency—yen, marks or dollars—that can be used only for buying goods from the industrialized world. Even if aid organizations wanted to give grants in local money, they would have to acquire it by trading their currency, and so the recipient's foreign exchange position would be precisely the same.

In theory, development aid works to the benefit of both the donor and recipient countries: the recipient gets a dam, steel mill or banking system, and the donor gets an immediate economic stimulus (from purchases of goods and services with the aid money) plus a new market if the development project pans out. Indeed, if the aid is in the form of a loan from the World Bank or some other multilateral organization, the original money spent earns interest and must eventually be repaid.

Historians can point to cases where the theory has worked wonders. During the 19th century, for example, the U.S. was a prime recipient of foreign loans. Railroads, canals, mines and factories were built with huge infusions of European capital, and U.S. trade with Europe enriched both sides. After World War II, Marshall Plan dollars rebuilt Europe, and the U.S. vaulted to preeminence in exports.

So why don't things work so well today? Candidate causes abound: corruption, incompetence and inappropriate projects top the lists of people at funding agencies. Recipients, meanwhile, often complain about "tying"—conditions on aid that dictate where the recipient can buy goods or whom it can hire to oversee projects.

Such ties raise the suspicion that the purpose of aid is largely to subsidize firms in the donor country. Legislators in the U.S. and elsewhere often justify foreign aid budgets on the basis of the economic benefits they will provide at home, and nations also use aid as a lever to open commercial markets.

Economist Elliot Berg of Development Alternatives in Bethesda, Md., considers such concerns ill founded. If the U.S. wants to subsidize U.S. firms, he says, it can do so by building bridges in Kentucky just as easily as by building them in Indonesia. Furthermore, tying allows donors to exercise control over the direction of a project and can help minimize corruption in the recipient country.

*"There is every incentive to draw up big projects with little economic value."*

Although most bilateral aid (grants and loans made directly from one nation to another) is tied, most multilateral aid (that dispensed by the World Bank and other international organizations) is not. Indeed, when competitive bids are solicited for projects funded by the World Bank, local businesses may win jobs even if their bids are 15 percent higher than those of more established international firms. This preference helps to develop management expertise in the recipient country, even if the local company still ends up purchasing most of its supplies on the international market.

Even without explicit tying, the requirements of donor organizations may still shape development decisions. When economic officials are determining the appropriate level of capital investment in their country, says Jim Boomgard of Development Alternatives, it makes little difference to them what particular forms that investment takes. If "this year the World Bank is into dams," he says, countries will propose building a dam. If microenterprise loans are the order of the day, they will set up rural credit cooperatives.

"There is every incentive to draw up big projects with little economic value," comments George Ayittey, a Ghanaian economist now at American University. He asserts that dams, generators and similar large installations have often served as "visible symbols of assistance" that gladden the hearts of donors rather than helping recipients. And in the five years or more that a project may take to come to fruition, "things happen: the government may be overthrown, a minister may be corrupt and demand 10 percent of all contracts, or prices may rise and $100 million becomes insufficient to complete the project, so it is abandoned."

In addition to the potential for distorting investment priorities, this kind of decision making has dangerous social consequences, according to Gabriel Roth, a former World Bank economist. Regardless of whether a development project succeeds or fails, he asserts, it will increase the power of government officials. The World Bank "caused aid to flow" for housing in Zambia, he recalls, "and to no one's surprise all the houses went to supporters of the governing party." Roth contends that multilateral aid organizations should focus less on simply transferring funds and more on helping governments to create laws and institutions that assist investment.

Government-to-government loans, Roth says, are particularly dangerous because they must be repaid regardless of whether the project they were intended to fund is successfully completed. As a result, aid organizations have little financial reason to lend money only for viable ideas. Last year an internal World Bank study reportedly found that the bank's proportion of "problem" projects had more than doubled between 1981 and 1991; it now stands just under 40 percent.

In contrast to the railroad bonds of the 19th century, it is not the investors in the industrialized world who bear the financial risk of a project gone wrong. Rather it is the citizens of the nation that received the aid.

*—Paul Wallich and Marguerite Holloway*

# Flight-Testing Fruit Flies

Insects are not as stupid as people might think. Studies have shown that several species can learn through association. This ability is crucial for many insects to survive in an environment filled with various chemical scents and visual cues. Parasitic wasps, for instance, locate their caterpillar hosts by learning odors that suggest the presence of the hosts. Such odors come from their hosts' feces and from the plants on which the caterpillars feed [see "How Parasitic Wasps Find Their Hosts," by James H. Tumlinson, W. Joe Lewis and Louise E. M. Vet; page 100].

You can see how insects can be taught to respond to an odor by raising some fruit flies on various food sources and testing them in a wind tunnel. The tunnel isolates the insects in a given area and provides the odor plume that the flies will use to locate food. The setup described here is similar to those used by researchers to investigate associative learning in wasps. Most of the material can be purchased at a hardware store.

Rearing the insects is the easiest part of these experiments. Anyone who has purchased fruit and let it sit on the kitchen counter has essentially "raised" the common fruit fly (*Drosophila*). To establish a number of *Drosophila* colonies, place a piece of fruit in the bottom of a clear jar. I used peeled bananas, but peaches, apples or oranges would also be good choices (use only one type of fruit). The jar should not be too small; otherwise you may end up with *Drosophila* that have developed poorly and cannot fly well.

Place the jar outside and watch the fruit for visiting adult insects. Many different insect species will come to the fruit, but they will not affect the project. If you have the patience, you can catch sight of the female *Drosophila* as she lays her eggs on the fruit. After two to three days, bring the jar inside. Place a double layer of cheesecloth over the jar opening and secure it with a rubber band. To keep the atmosphere inside the jar moist, it may be necessary to put a wet paper towel over the cheesecloth. Never leave standing water in the jar.

Depending on the room temperature, the eggs will hatch in about 21 hours. The larvae will not be visible for at least three to four days. On about the fifth day, they will crawl out of the fruit and up the walls of the jar. There they will pupate. The adults will emerge in about five days and will mate and lay their eggs in the fruit. In the experiments, I achieved the best results if the adult flies were at least four to five days old.

You will need to construct an aspirator to suck the insects out of the jar, because *Drosophila* are too small and fragile simply to pick up. To build the aspirator, you will need two copper pipes, about 1.5 inches long and about one-quarter inch in diameter, and 30 to 40 inches of flexible, plastic tubing, such as Tygon tubing. The pipes should plug tightly into the tubing; you may have to spend a few minutes at the hardware store mixing and matching. You will also need a clear plastic 35-millimeter-film canister or something similar.

Puncture two holes in the lid of the canister and insert a copper pipe through each hole. Stuff some cotton in one of the pipes; do not pack it in tightly. Cut the Tygon tubing in half and attach one onto each piece of copper pipe. Snap the cover back on the canister. To check for leaks, place the tube connected to the cotton-plugged pipe in your mouth and hold your finger over the end of the other piece of tubing. You should be able to feel a vacuum when you suck. If you do not, there must be a leak. Make sure that the cap is on tight and that no air is getting in around the copper-tubing connections. Leaks can be plugged with modeling clay.

Now you can use the aspirator to collect the fruit flies. Insert the tube without the cotton plug into the fruit fly jar and suck on the other tube. The flies will end up in the film canister. The cotton plug will keep the flies from going up the tube into your mouth. There is no harm, however, in swallowing them. Most of us have eaten a bug or two before anyway. If your aspirator does not work properly, check to see if the cotton is packed too snugly or if there are leaks. You may need to reduce the length of the tubing or to adjust the diameter. Sucking harder is not likely to be the answer.

To test the flies, I constructed a wind tunnel 49 inches long, 24 inches high and 30 inches wide. These dimensions were chosen because I used a pre-cut material called Melamine. It is a kind of veneered particleboard often used in shelving. Melamine comes in white, which makes the insects more visible. Its smooth, polished surface also minimizes wind turbulence.

Secure three pieces of Melamine to form the sides and bottom. Two one-inch by two-inch pieces of wood were used for the support across the ends. I purchased a piece of Plexiglas of approximately the same dimensions for the top of the wind tunnel. At one end I pulled two layers of outdoor screening tightly across the frame to straighten the wind path and reduce its speed. At the other end I secured cheesecloth, leaving the bottom free so that I could easily reach into the tunnel to add the insects. (Make sure the cheesecloth is long enough to tuck under the bottom.)

A large standard box fan created the wind. I suggest setting it on its lowest speed and positioning it approximately four feet from the screened end. There should be just a slight breeze inside the tunnel. Remember, you are flying fruit flies, not jumbo jets.

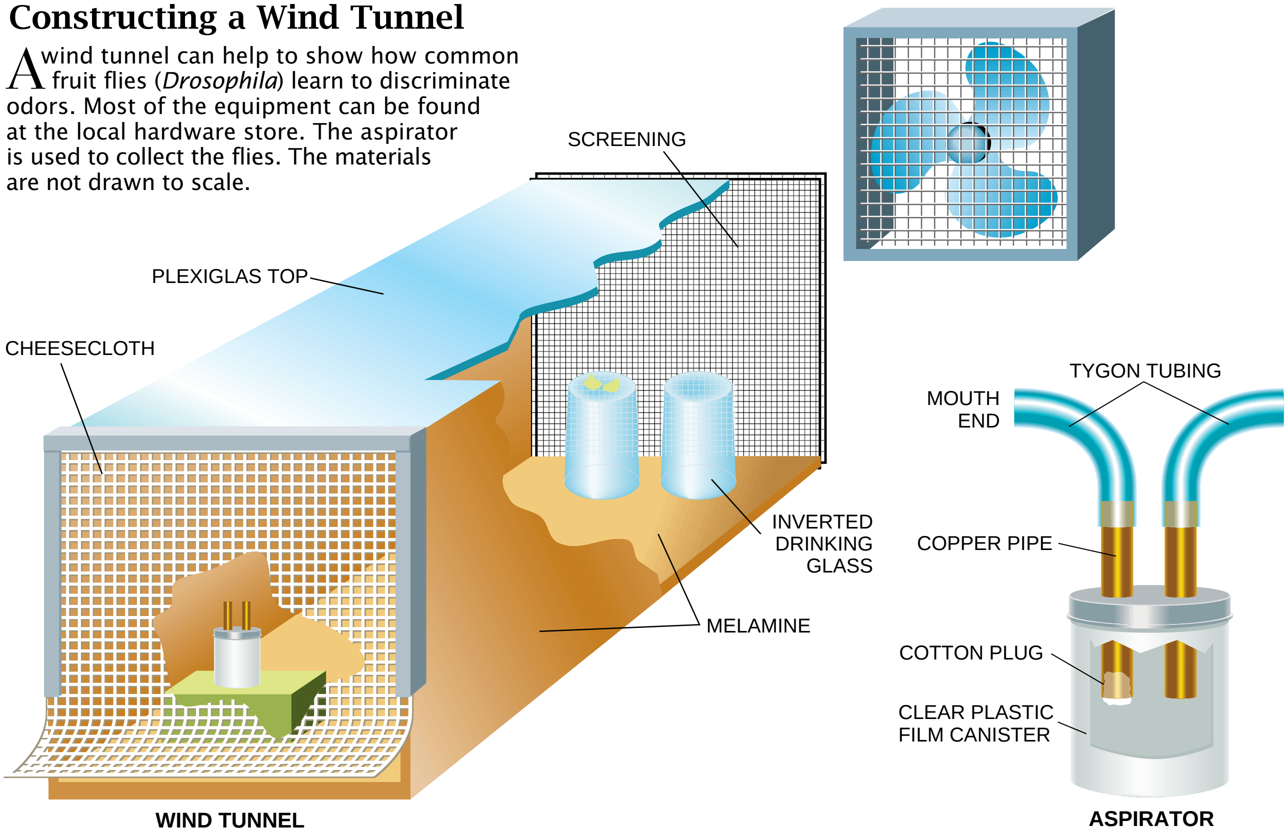
For accurate results, an even and bright luminescence over the entire length of the tunnel is necessary. One long fluorescent light strip or two rather large and bright floor lamps placed on opposite sides and on opposite ends of the tunnel will do. Background odors may also present complications. I noticed that if I tried to fly the insects after cooking an odorous dinner, the flies did not seem to be as efficient at finding the fruit odors as they did when I went out to eat. Either they did not like my cooking or they were distracted by the extraneous odors. (I hope it was the latter.)

The fruit needs to be sitting on some kind of platform about 10 inches above the tunnel floor and about three to six

JANET M. STRONG-GUNDERSON currently works in the Environmental Sciences Division of Oak Ridge National Laboratory in Tennessee. Her diverse projects in ecology and microbiology examine the processes by which microorganisms break down hazardous chemical contaminants in the soil and in the water. She received her Ph.D. in zoology from Miami University in Oxford, Ohio.

## Constructing a Wind Tunnel

A wind tunnel can help to show how common fruit flies (*Drosophila*) learn to discriminate odors. Most of the equipment can be found at the local hardware store. The aspirator is used to collect the flies. The materials are not drawn to scale.




inches from the screen. I found that inverted narrow drinking glasses worked nicely. Place two of them in the tunnel, each about two inches from the sides.

To see if your wind tunnel is working properly, you will need to "test fly" a few of your insects. After capturing the flies in the film canister, give them a few minutes to settle down. Detach both pieces of tubing from the canister. Set the canister upright in the tunnel on a platform elevated about six inches above the tunnel floor. Touch your finger to the fruit and then dab your finger on the copper pipe that does not have the cotton in it. The insects will crawl up this pipe. This process helps them orient to the odors. Observe how the flies "taste" the material you have dabbed on the pipe. If the insects cannot crawl up the pipe, the pipe is probably inserted too deeply into the canister. Pull it out so that the bottom of the pipe is almost even with the canister cover.

Once the insects have crawled out of the canister, they will usually walk around the top of the pipe and maybe down the sides. When an individual does "decide" to fly to the fruit odor, you will see that its flight pattern is initially a casting from side to side. As it moves closer to the source of the odor, it narrows in on the fruit and finally lands. Difficulties in landing may be a consequence of turbulences on the downwind side of the fruit. Try using a narrower drinking glass.

Once you are satisfied that your flies can function in the wind tunnel properly, you can begin the experiments. Place a piece of the fruit on which the flies had been raised (I used banana) on one of the drinking glasses and a piece of foam or some other inert material on the other. The foam should be about the same size and color as the fruit. To which do the insects fly? Then reverse the positions of the foam and fruit. Is there any difference? Flies might prefer one side of the tunnel over the other because of factors such as lighting, wind speed or turbulence; you may need to adjust the lighting, the placement of the fan or the size of the fruit platform.

Next, repeat the experiment but replace the foam with a different piece of fruit—say, peach. In some instances, I observed that if a banana-raised insect began to fly into the banana-odor plume but somehow "lost" the odor and then detected the peach plume, it did not land immediately on the peach but continued to search, casting back and forth in a larger arch. Often it was able to relocate the banana odor and then landed on the correct fruit. I also observed that when a banana-raised fly "accidentally" landed on the peach it usually did not stay long. It began searching again.

For a third set of choices, I offered a peach and an apple to a banana-raised fly. In this case, the number of flies that landed on each fruit were all about equal. The searching process seemed to take longer than when one of the choices was a familiar one. You could also see how the flies respond to "neutral" odors such as vanilla or almond.

Successful results will depend on the care taken during construction of the wind tunnel, the operating conditions such as wind speed and luminescence as well as the health of the insects. Perhaps the most frustrating problem you might encounter is when the insects do not fly at all. There are several possible reasons. The size of the tunnel plays a role; I tested larger and smaller sizes, and neither extreme works well. The flies might be too young; try waiting a few days. Like most scientific experiments, patience and a little imagination will pay off.

FURTHER READING

THE INSTINCT TO LEARN. J. L. Gould and C. Grant-Gould in *Science 81*, Vol. 12, No. 4, pages 44–50; May 1981.

INDUCTION OF HOST PREFERENCE IN DROSOPHILA MELANOGASTER. J. Jaenike in *Oecologia*, Vol. 58, No. 3, pages 320–325; June 29, 1983.

THE BIOLOGY OF LEARNING. P. Marler and H. S. Terrace. Springer-Verlag, 1984.

# BOOK REVIEW by Robert M. May

## The End of Biological History?

THE DIVERSITY OF LIFE, by Edward O. Wilson. Belknap Press of Harvard University Press, 1992 ($29.95).

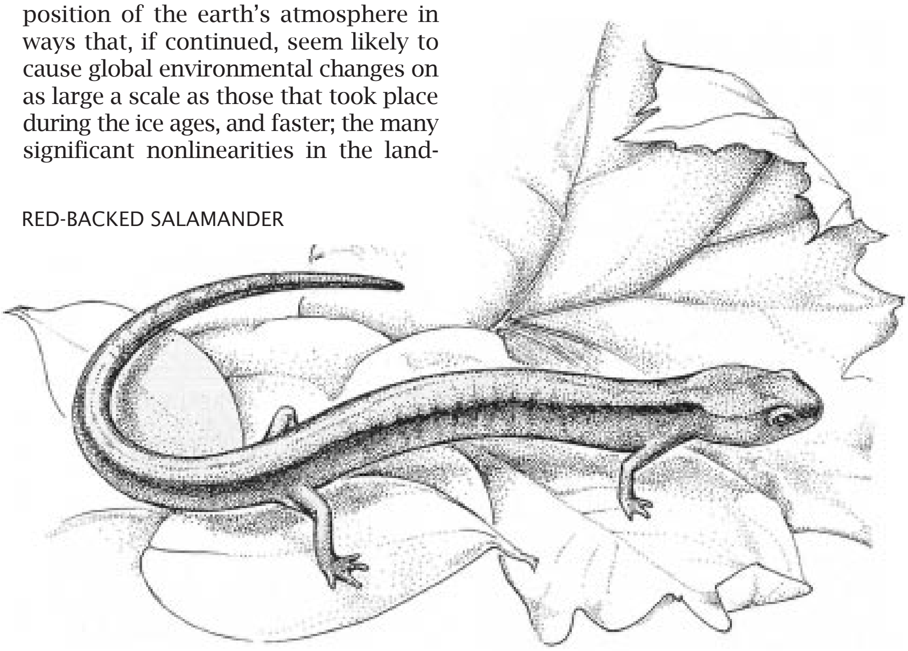
People have always felt that the times in which they lived were special: the best of times, the worst of times, and sometimes both together. The feeling tends to be particularly strong at points of calendric resonance, such as the end of centuries, much less the end of millennia. But our own times *are* special, by many objective measures, in that the scale and scope of human activities have, for the first time, grown to rival the natural processes that built the biosphere and that maintain it as a place where life can flourish.

Many facts testify to this statement. It is estimated that somewhere between 20 and 40 percent of the earth's primary productivity, from plant photosynthesis on land and in the sea, is now appropriated for human use. The global amounts of biologically available nitrogen and phosphorus associated with fertilizers and other chemicals used in agriculture rival the amounts mobilized by natural processes. Macroscopic holes in the ozone layer have been created by chlorofluorocarbons. Carbon dioxide from the burning of fossil fuels and tropical forests, along with other greenhouse gases, is changing the composition of the earth's atmosphere in ways that, if continued, seem likely to cause global environmental changes on as large a scale as those that took place during the ice ages, and faster; the many significant nonlinearities in the land-ocean-atmosphere system make precise forecasts impossible at this point. And it seems probable that today's species of plants and animals stand on the curling tip of a wave of extinction more rapid and extensive than any in the fossil record.

This last fact—the impending impoverishment of life on the earth—is the tocsin that sounds throughout Edward O. Wilson's latest book. He is largely responsible for giving currency to the word "biodiversity." In its broadest terms, biodiversity stands for the totality of living things, embracing all species of organisms, big and little, on land and in the sea. More precisely, biodiversity usually refers to the total number of species present in a defined region. The term is, however, sometimes expanded downward to embrace the genetic diversity within populations and species, and upward to embrace diversity at the level of communities or ecosystems. I think it is reasonable to consider species (and their constituent populations) as the units in which biodiversity is measured, partly because they are the natural units of evolution and partly because this definition is practical and relates directly to concerns about conservation.

In 1988 Wilson edited an influential book entitled simply *Biodiversity,* which recorded a symposium organized by the U.S. National Academy of Sciences.

RED-BACKED SALAMANDER



*The Diversity of Life* covers a lot of the same ground (along with subsequent advances), but it is a much more personal statement.

The opening chapter is prose poetry:

> In the Amazon Basin the greatest violence sometimes begins as a flicker of light beyond the horizon. There in the perfect bowl of the night sky, untouched by light from any human source, a thunderstorm sends its premonitory signal and begins a slow journey to the observer, who thinks: the world is about to change. And so it was one night at the edge of [a] rain forest north of Manaus, where I sat in the dark, working my mind through the labyrinths of field biology and ambition, tired, bored, and ready for any chance distraction.... Enclosed in darkness so complete I could not see beyond my outstretched hand, I was forced to think of the rain forest as though I were seated in my library at home, with the lights turned low. The forest at night is an experience in sensory deprivation most of the time, black and silent as the midnight zone of a cave. Life is out there in expected abundance. The jungle teems, but in a manner mostly beyond the reach of the human senses.... When I rose at dawn the next morning, Fazenda Dimona had not changed in any obvious way from the day before. The same high trees stood like a fortress along the forest's edge; the same profusion of birds and insects foraged through the canopy and understory in precise individual timetables. All this seemed timeless, immutable, and its very strength posed the question: how much force does it take to break the crucible of evolution?

The final chapter offers nothing less than a new religion, the Environmental Ethic. Unlike the vast array of existing religions, essentially all of which see humans as special, deferring only to some god or gods (commonly males of choleric disposition), Wilson's Environmental Ethic sees us as one among many species, with our only special responsibility being to respect and conserve the biological riches we have inherited:

> The evidence of swift environmental change calls for an ethic uncoupled from other systems of belief. Those committed by religion to believe that life was put

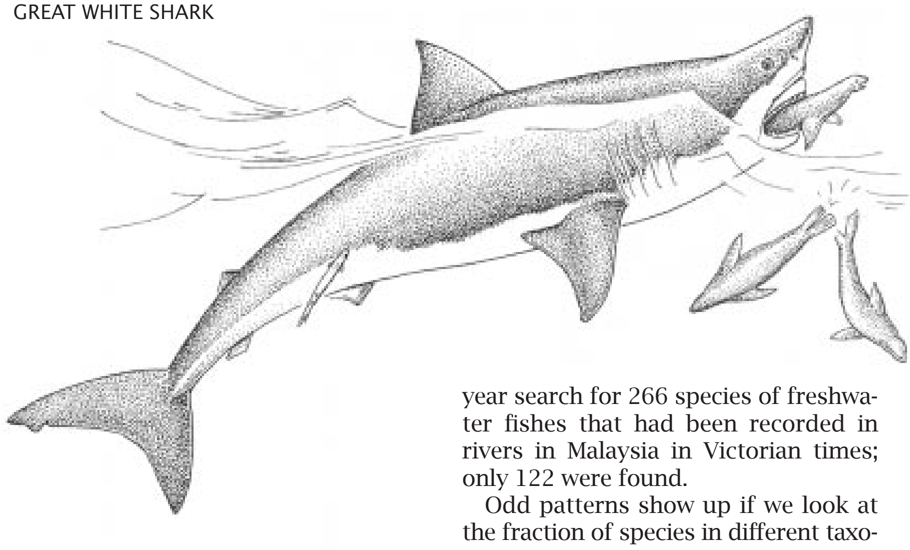
> on earth in one divine stroke will recognize that we are destroying the Creation, and those who perceive biodiversity to be the product of blind evolution will agree. Across the other great philosophical divide, it does not matter whether species have independent rights or, conversely, that moral reasoning is uniquely a human concern. Defenders of both premises seem destined to gravitate toward the same position on conservation.... An enduring environmental ethic will aim to preserve not only the health and freedom of our species, but access to the world in which the human spirit was born.

In between these opening and closing chapters, Wilson discusses the evolutionary processes that create species and the five major cataclysms that have diminished global diversity over the past 600 million years, each requiring 10 to 100 million years of evolutionary repair. He also surveys our emerging understanding of the factors that structure communities of plants and animals and that govern the numbers and biogeographic distribution of species among real or virtual archipelagoes of islands. Our current and lamentable ignorance as to how many species there actually are on the earth is summarized. Against this background Wilson assesses the likely impacts of human population growth and associated patterns of resource utilization on biological diversity, in the tropics and elsewhere. He sketches innovative approaches to evaluating the economic value of biological diversity, as the potential capital stock for pharmaceutical and agricultural industries and as the gene store that will provide the raw stuff for the biotechnological revolution. The penultimate chapter draws all this together in an "action plan" for biodiversity, to which I will return.

Although much of this material will be familiar to many of the people who read the book (particularly those likely to be reading this review), there are many fresh perspectives and unfamiliar facts.

We are increasingly used to seeing dramatic statements that *x* species are being extinguished every day. But relatively few understand how these estimates are derived. Wilson gives a good account of our lack of any real understanding of current extinction rates. And he explains how, lacking better, we can make crude estimates by combining knowledge about habitat destruction with established "species area" relations (which roughly say that if you reduce the area to one tenth, you halve the eventual number of species in it).

GREAT WHITE SHARK

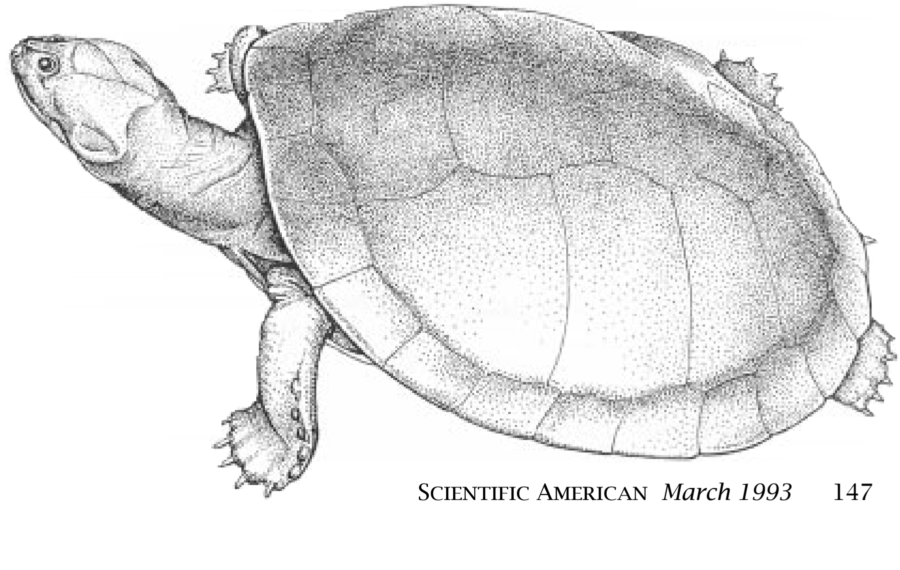


But he also gives an excellent survey of specific and detailed studies. For instance, only about 1 percent of the world's 9,000 bird species have been certified extinct since the 17th century, and the current *Red Data Book* of the International Union for the Conservation of Nature (IUCN) in Switzerland lists only about 11 percent as threatened with extinction. But Jared M. Diamond of the University of California at Los Angeles found in a survey of 164 bird species recorded from the Solomon Islands that 12 had not been seen for 50 years or more, nor could he find them. But only one of these 12 had satisfied the establishment, "middle class" criteria needed to get it a tombstone of certified extinction. Diamond believes that many tropical birds in poorly studied regions are extinct or severely threatened but have not made it into the *Red Data Book;* he estimates that a "Green Book" of unthreatened bird species would include less than half. Wilson surveys other studies pointing in this direction, including a three-year search for 266 species of freshwater fishes that had been recorded in rivers in Malaysia in Victorian times; only 122 were found.

Odd patterns show up if we look at the fraction of species in different taxonomic groups that are listed as "endangered" or "vulnerable" in the IUCN lists. Thus, working with Robin Pellew and other people at the World Conservation Monitoring Centre in Cambridge, England, Fraser Smith, a graduate student at the University of Oxford, has noted that the percentages listed as threatened in the following animal groups are: mammals, 11 percent; birds, 11 percent; reptiles, 3 percent; amphibians, 2 percent; fishes, 2 percent; mollusks, 0.4 percent; crustaceans, 3 percent; and insects, 0.07 percent(!). For trees, we have 32 percent of the 70 species of gymnosperms but only 9 percent of the roughly quarter of a million species of angiosperms (8.5 percent of monocots and 9.3 percent of dicots). The most immediate conclusion is that the data speak mainly to how much attention you get, not how threatened with extinction you are.

Wilson emphasizes that extinction threats are not confined to the tropics. He cites the numbers of insect and other invertebrate species classified as endangered or vulnerable to extinction in

AMAZON RIVER TURTLE

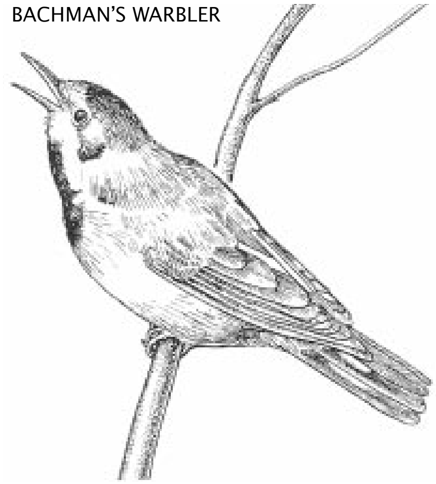


Germany (34 percent), Austria (22 percent) and England (17 percent). This is a bit misleading, being based on a report that gives percentages of invertebrate populations that are in decline, rather than formally classed in a threatened category. Even so, the true figures are alarming enough. In the U.K., 1 percent of all insect species have become extinct this century, and 7 percent are currently classified as endangered (3 percent) or vulnerable (4 percent); the 7 percent figure also applies to invertebrates other than insects. There are some 5,700 designated Sites of Special Scientific Interest (SSSI) in the U.K. These, unfortunately, do not receive absolute statutory protection, and estimates of the rate at which they are being destroyed range from 1 percent per year (the official number in the government's White Paper on the Environment, published in 1990), to 5 percent. Thus, SSSIs are being destroyed at roughly the same rate as the tropical rain forest is being burned. Some First World pieties may be shaken by these statistics.

Wilson also offers an illuminating tour of about 20 of the world's "hot spots" of biodiversity. These are regions containing many species, a large proportion of which are found nowhere else ("endemics") and are in danger of extinction from human activities. These hot spots are based partly on information about plants and partly on an influential analysis of patterns of bird-species endemism by Colin Bibby of the International Council for Bird Preservation (ICBP) and others (see *Putting Biodiversity on the Map: Priority Areas for Global Conservation,* ICBP, Cambridge, 1992). As Wilson emphasizes, such studies of hot spots are far from complete, and they tend to be dominated by favored taxonomic groups. Wilson's reservations are borne out by preliminary studies by John H. Lawton and John Prendagast of Imperial College, London, along with people at the U.K. Biological Records Centre in Monks Wood, who have analyzed the extensive data recording presence and absence of British plants and animals, on 10-square-kilometer grids over the entire country. They find that the hot spots of diversity for one group do not necessarily have great overlap with those for a different group. They also find that endangered, vulnerable or rare species in general tend to be found outside hot spots, which reinforces Wilson's caveats. We need to know more about the ecology and biogeography of commonness and rarity.

Wilson seeks to end on a constructive note by prescribing a five-point plan for action: survey the world's fauna and flora; create wealth out of biological products (medicines, foods and other materials); promote sustainable development; save what remains of biological diversity; and restore wildlands in regions that have been exhausted and abandoned.

BACHMAN'S WARBLER



I have recently expressed my own enthusiasms and suggestions for the first of these tasks [see "How Many Species Inhabit the Earth?," SCIENTIFIC AMERICAN, October 1992]. Its feasibility depends in part on how many species you think there are on the earth (my guess is around five to eight million; Wilson countenances much higher estimates) and in part on imaginative approaches to the job ("parataxonomists" backed up by information technology). The second and—to a vastly greater extent—the third of Wilson's tasks require significant changes in cultural norms, in order that people accept new products and, more difficult, new ways of life. How do we bring about these changes? What kinds of economic and educational efforts are likely to be effective? These are questions that deserve more attention.

Under the fourth heading, Wilson gives an excellent survey of current activities aimed at ex situ preservation of endangered animals and plants in zoos, botanical gardens and the like. There are some 1,300 botanical gardens and arboretums in the world, many harboring species that are endangered or even extinct in the wild. Animals are much more difficult than plants or microorganisms to maintain ex situ. Zoos and other institutions currently have breeding populations of more than half a million individuals belonging to more than 3,000 species of mammals, birds and, to a lesser degree, amphibians and reptiles. These collections represent 13 percent of all known species of land-dwelling vertebrates. As more and more animals are put under pressure by fragmentation or outright destruction of their habitats or by other factors, there is going to be increasing difficulty in keeping this ark afloat. And—like the Victorian Noah's ark toys—this takes no account of endangered insects and other invertebrates. Wilson sensibly concludes that although ex situ methods have their important uses, real hope for the future must rest on the preservation of natural ecosystems. Currently 4.5 percent of the earth's land surface is formally designated as park or reserve, but many of these need better protection than they are getting, and we need more of them. Some such addition may be achieved by restoration, as suggested in Wilson's fifth point and exemplified by current plans to "grow" a park in Costa Rica's Guanacaste Province, by buying low-grade farm and timberland next to Santa Rosa National Park and letting the park spread into it.

Given all the realities, I would have welcomed a bit more on how to make agonizing choices, as we increasingly will have to do. How do we quantify "taxonomic uniqueness" or degree of "independent evolutionary history"? We ultimately will need such a calculus of biodiversity, to help us apportion values and efforts between, say, the New Zealand tuatara (almost a monogeneric subclass by itself) and any other handful of less taxonomically unique reptiles—or between 10 beetle species and one small amphibian.

Wilson says relatively little about the continuing growth of human populations. But this is the engine that drives everything. Patterns of accelerating resource use, and their variation among regions, are important but secondary: problems of wasteful consumption can be solved if population growth is halted, but such solutions are essentially irrelevant if populations continue to proliferate. Every day the planet sees a net increase (births less deaths) of about one quarter of a million people. Such numbers defy intuitive appreciation. Yet many religious leaders seem to welcome these trends, seemingly motivated by calculations about their market share. And governments, most notably that of the U.S., keep the issue off the international agenda; witness the Earth Summit meeting in Rio de Janeiro. Until this changes, I see little hope.

*ROBERT M. MAY is Royal Society Research Professor at the University of Oxford and Imperial College, London.*

ESSAY by Otto E. Landman

# Inheritance of Acquired Characteristics

Can environmentally induced or acquired changes in organisms be transmitted to future generations? Does the inheritance of acquired characteristics—if it occurs at all—play a significant role in evolution? These questions were the subject of heated scientific and political controversy until as recently as the 1960s, when the decisive successes of classical genetics submerged this debate. If asked, most biologists today would say that inheritance of acquired characteristics never occurs. Yet there are actually numerous well-documented examples of the phenomenon, and I believe it has played a major role in speeding up evolution.

Jean Baptiste de Lamarck, the 18th-century French evolutionist whose name is frequently linked to the subject of inheritance of acquired characteristics, believed that an animal's use or disuse of an organ affected that organ's development in the animal's offspring. Lamarck explained that the giraffe's long neck, for example, resulted from its ancestors' having stretched their necks to reach distant leaves. Genetics and molecular biology have since shown convincingly that adaptations of a body part cannot trigger changes in the sperm and egg cells that transmit traits.

Some environmentally induced changes are heritable, however. In growing bacteria, for instance, the rigid cell walls are simultaneously synthesized and cut by enzymes. If an experimenter completely removes the cell wall from a bacterium, the balance between wall synthesis and destruction is shifted, and the bacterium continues to grow and multiply indefinitely without walls. Bacterial nakedness is clearly an acquired characteristic that is inherited.

By accident, the protozoan *Oxytricha* sometimes produces "double monsters" that consist of two individuals fused together like Siamese twins. When a double monster is cut in two lengthwise, the result is two "singlets" that reproduce to give normal offspring. If a double monster is cut in two crosswise, however, each half gives rise to a double monster, which gives rise to more double monsters. A heritable characteristic can be acquired through a single cut.

The inherited absence of cell walls in bacteria and the accidentally acquired monster condition in *Oxytricha* result from stabilized changes in the expression of gene activity—without any attendant changes in the genes themselves.

A fundamentally different kind of environmentally induced heritable change occurs when specific sets of genes are either eliminated from or added to an organism. For example, virus sigma causes fruit flies to be sensitive to carbon dioxide. Infected flies pass on the virus to their offspring. But if the flies are kept warm while they are producing eggs, the virus is eliminated, and the offspring are resistant to carbon dioxide.

Sometimes the distinction between an organism and its evolutionary fellow traveler is not so obvious. In the protozoan *Euglena,* as in all green plants, self-replicating organelles called chloroplasts carry out photosynthesis. If *Euglena* are treated for six days with the antibiotic streptomycin, they lose their chloroplasts. *Euglena* thus "cured" of chloroplasts can survive without photosynthesis and transmit the condition to their offspring.

The opposite process—the acquisition of foreign self-replicating organelles—can also occur. Plasmids are small circles of DNA commonly found in bacteria. If an *Escherichia coli* bacterium carrying a plasmid called fertility factor F is added to *E. coli* lacking the factor, F spreads rapidly through the culture and is inherited from then on. Viruses, too, can be passed to offspring and may become a permanent part of the inheritance of an organism. For example, it has been estimated that mouse chromosomes contain about 25,000 genes from retroviruses that were acquired through the ages by means of infection.

There are many other examples. In fact, most biologists now believe that complex cells first acquired the ability to carry out photosynthesis when one of them fused with a bacterium possessing that trait. The descendants of such bacteria are the chloroplasts found today in all plants.

If we include in our concept of acquired characteristics the foreign genes introduced through viral infection, plasmids and bacteria, it becomes evident that these acquisitions have played a major role in evolution. Because the genes of bacteria or plasmids have had many millions of years to develop coordinated systems, their gene banks may confer on a new host fully developed capabilities—such as photosynthesis—that would have taken eons to evolve anew through random mutations coupled with natural selection.

I suggest that the genes of organisms can be divided into two groups. Most are inherited "vertically" from ancestors. Some, however, were acquired "horizontally" at different times from viruses, plasmids, bacteria or other agents. Biologists are starting to realize that genes in nature are transmitted horizontally even between organisms considered to be unrelated: from bacteria to plants, for example, and from bacteria to yeast. Evidently, nature has anticipated the artifices of genetic engineers.

Despite the extensive experimental evidence demonstrating inheritance of acquired traits, in a recent survey of 30 current college genetics texts I found no mention that the phenomenon actually occurs. To quote one of these: "Lamarck's hypothesis of the inheritance of acquired modifications has been discarded because no molecular mechanism exists or can be imagined that would make such inheritance possible."

The skeptical attitude toward inheritance of acquired characteristics reflected in this quote crystallized in the atmosphere of the highly politicized debates of Western geneticists with the Soviet plant breeder Trofim D. Lysenko. Lysenko, who controlled agricultural and genetic research in the Soviet Union during the 1940s and 1950s, clung to Lamarck's ideas about inheritance on ideological grounds. He insisted that useful traits could be imparted to plants by "training" them; he maintained that genes did not exist. He was wrong, and Soviet agriculture paid dearly. But biologists have thrown an important baby out with the Lysenkoist bathwater. Great advances in genetics and molecular biology since then have shown that the inheritance of acquired characteristics coexists comfortably with molecular genetics. It is now time to recognize that an understanding of the role of acquired characteristics opens a broader perspective on genetics and evolution.

---

*OTTO E. LANDMAN is professor of biology emeritus at Georgetown University. He is the author of a review on the inheritance of acquired characteristics in the 1991* Annual Review of Genetics.