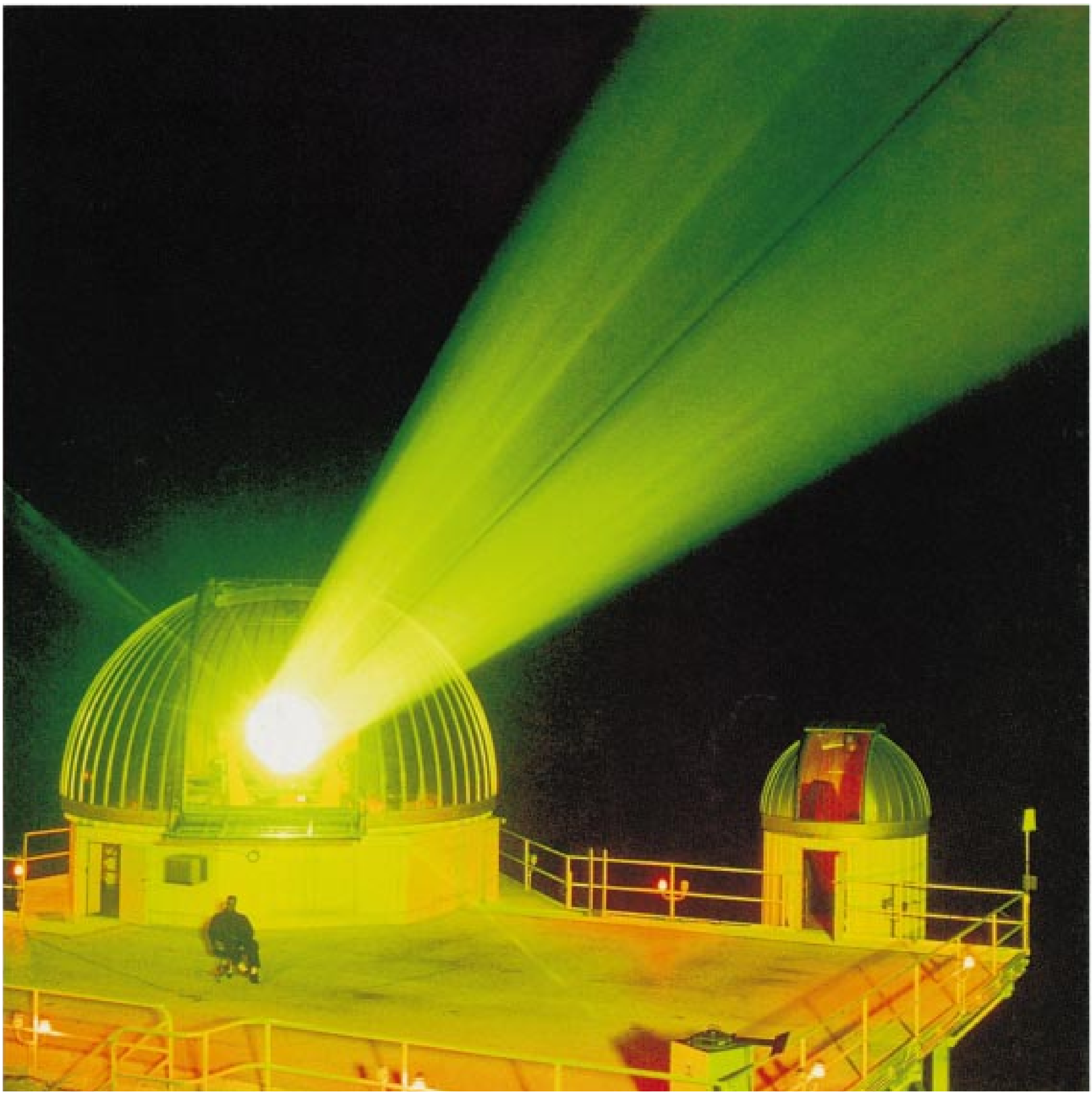
SCIENTIFIC AMERICAN

JUNE 1994
$3.95

Was there a race to the moon?

How the brain makes emotional memories.

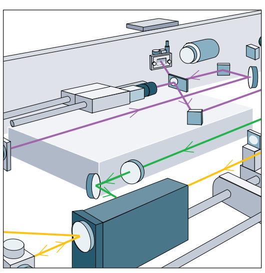
Genetic testing: boon or bane?

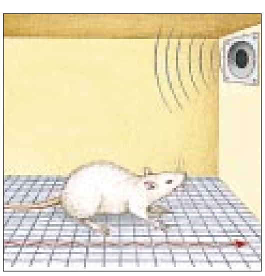


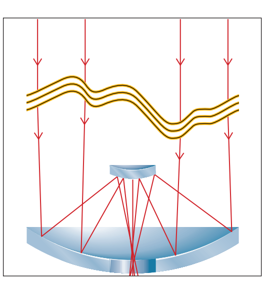
***Starfire laser beam** creates a guide star for adjusting a flexible telescope mirror.*

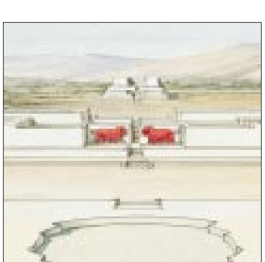
SCIENTIFIC AMERICAN

June 1994 Volume 270 Number 6

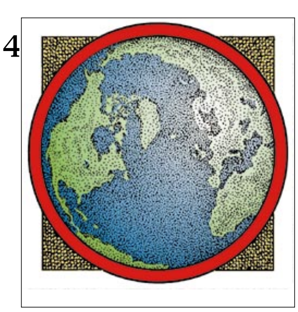


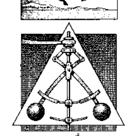


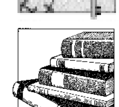


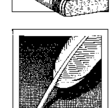


DEPARTMENTS









THE COVER photograph shows the powerful Starfire laser beam generated at the U.S. Air Force's Phillips Laboratory in New Mexico. The beam, when reflected in the upper atmosphere, creates an artificial guide star that is used to calibrate the Starfire telescope's flexible mirror to compensate for atmospheric turbulence. The man seated at the foot of the dome is a spotter who warns of approaching aircraft so that the beam can be shut down to protect the airplane's crew and instruments (see "Adaptive Optics," by John W. Hardy, page 60).

THE ILLUSTRATIONS

Cover photograph © 1994 by Roger Ressmeyer/Starlight Photo Agency, Inc.

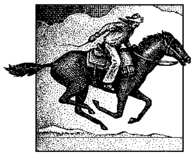
Page	Source
36–37	Courtesy of Glenn Swanson, *Quest* magazine (*left*), National Aeronautics and Space Administration (*right*)
38	NASA (*top*), Sovfoto/Eastfoto (*bottom*)
39	NASA (*top left*), UPI/Bettmann (*top center*), NASA (*top right*), Tass, Sovfoto/Eastfoto (*bottom*)
40	NASA (*top left*), AP/World Wide Photos (*top right*), Sovfoto/Eastfoto (*bottom*)
41	NASA (*top*), Sovfoto/Eastfoto (*bottom left*), A. Moklet Sov./Novosti Press Agency/Starlight Photo Agency, Inc. (*bottom center*), courtesy of Alain Dupas (*bottom right*)
42	NASA (*top*), courtesy of Glenn Swanson, *Quest* magazine (*bottom left*), courtesy of Alain Dupas (*bottom right*)
43	NASA (*top*), courtesy of Sotheby's (*bottom left*), Edwin Cameron (*bottom center*), Tom Stafford, Vance Brand/Starlight Photo Agency, Inc. (*bottom right*)
44–45	Ian Worpole
46	James Montanus, University of Rochester
47	Ian Worpole
48	Jack Harris/Visual Logic (*top left*), Ian Worpole (*top right* and *bottom*)
49	Ian Worpole
51	Roberto Osti (*drawings*), Andrew Leonard/APL Microscopic (*photographs*)
52	Roberto Osti (*top*), Ian Worpole (*bottom*)
53	Ian Worpole
55	Ian Worpole (*drawings*), Joseph E. LeDoux (*photographs*)
56–57	Roberto Osti
61	Roger Ressmeyer/Starlight Photo Agency, Inc.
62	Jared Schneidman/Jared Schneidman Design
63	Jared Schneidman/JSD (*drawings*), John W. Hardy (*photographs*)
64–65	Jared Schneidman/JSD
66–67	Steven N. Patricia (*top*), Gabor Kiss (*middle*)
69	Shelia Pozorski and Thomas Pozorski (*left*), Steven N. Patricia (*right*)
70–72	Shelia Pozorski and Thomas Pozorski
74–75	Mark Moffett/Minden Pictures
76	William F. Towne
77	Tomo Narashima
78	Tomo Narashima (*top*), Mark Moffett/Minden Pictures (*bottom*)
80	Thomas Seeley
83	Michael J. Balick
84	Michael J. Balick (*top*), Roberto Osti (*bottom*)
85	Gregory Shropshire, Ix Chel Tropical Research Foundation, Belize
87	Paul Alan Cox
88–89	Courtesy of Susan Lanzendorf, Jones Institute, Eastern Virginia Medical School (*left*), Hank Morgan (*right*)
90–91	Jared Schneidman/JSD (*top*), Gabor Kiss (*bottom*)
92	UPI/Bettmann
94	Jared Schneidman/JSD
95	Lester Sloan
96	Abraham Menashe
108–111	Kathy Konkle

SCIENTIFIC AMERICAN®

Established 1845

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

PRINTED IN U.S.A.



LETTERS TO THE EDITORS

Unfinished Business

In "Particle Metaphysics" [SCIENTIFIC AMERICAN, February], John Horgan argues that we particle physicists have bankrupted ourselves by our own successes. A "desert" of physics between the Large Electron-Positron Collider energies and the scale of grand unified theories means that our most beautiful theories are inaccessible to experiment, and thus our field is nearing a dead end. This is like saying that biology is a waste of time because the mystery of life is too difficult to comprehend.

Despite the data doldrums of the 1980s, the pages of the *Physical Review* are filled with experimental results in the physics of heavy quarks and leptons, tests of fundamental symmetries, searches for new phenomena and much more. Particle physics is as interesting and stimulating as it has ever been. Our successes have only added to that richness and to say otherwise reveals a shallow heart.

The best argument for the continued funding of particle physics experimentation is the one rooted in the true strengths of our field: its far-reaching beauty and profound implications. The experience of selling the Superconducting Super Collider to ourselves and to the country has left many of us cynical and unenthusiastic. But this is not the fault of the field—only of the times. Elementary particle physics will not die as long as we remember why we are pursuing it in the first place.

ALAN J. WEINSTEIN
Laboratory of High Energy Physics
California Institute of Technology

The Best Defense

In "The Future of American Defense" [SCIENTIFIC AMERICAN, February], Philip Morrison, Kosta Tsipis and Jerome Wiesner argue that collective security, such as coalition forces, can meet any future military challenges. That is simply not so, and the example of the Persian Gulf War, to which the authors point, demonstrates it. The U.S. took months to build up sufficient strength to attack Iraqi forces in Kuwait. The sea-lift capability of the U.S. is sadly lacking. The U.S. merchant fleet is practically nonexistent. The airlift capacity was stressed to the point that part of the Civil Reserve Air Fleet was required.

If the active forces are to be significantly reduced, then the reserve forces must be increased to retain qualified personnel for future conflicts. Additionally, the industrial base must be maintained and available to provide for a rapid buildup if needed.

The military still has valid "nonmilitary" missions around the world and at home. The basic rule for offensive operations is a three-to-one advantage in personnel and equipment. Perhaps a little more consideration is needed before the U.S. military shrinks away past the point of recovery.

(I am a major in the U.S. Army and a graduate of the U.S. Army Command and General Staff College. These views are strictly my own and do not reflect the official positions of the U.S. government, the Department of Defense or the Department of the Army.)

NIELS J. ZUSSBLATT
Chesterfield, Mo.

The U.S. does not have excessive airlift and sea-lift capability when it comes to addressing "brushfire" wars. Because we can only guess where we will confront aggression next, there should be an emphasis on weapons and equipment that make possible a powerful, conventional response in hours or days rather than weeks or months. It takes decades to introduce new weapons systems and considerable time to bring old ones out of mothballs; defense reductions will effectively be irreversible. We should resist the temptation to base our decision for our future defense on bean counting and wishful thinking.

CHRISTOPHER ROSEBERRY
Rowlett, Tex.

I agree with the authors that there should be some kind of drawdown from the years of the Reagan military buildup, but the plan proposed in the article should be sent back to the drawing board. Planning based on the assumption that the U.S. has only four potential adversaries (Iran, Iraq, North Korea and Libya) is an exercise with blinders. NATO has been paralyzed by the dilemma of whether to intervene in the Yugoslavian civil war. Some Pentagon planners thought that fighting in the mountainous terrain would require more combat personnel than had Operation Desert Storm. What wonderful glue holds Ukraine or Belarus together? How big a peacekeeping force would it require to sort out a civil war there patterned on Serbia versus Bosnia?

W. D. KELLY
Houston, Tex.

The authors reply:
It is confidence in our strategic-warning capabilities and in the prodigious capabilities of the U.S. Marine Corps, not bean counting or wishful thinking, that led us to our recommendations. In the Gulf, the U.S. was able to insert tripwire forces in Saudi Arabia promptly, as was urgently needed, and then to build up to win. We agree that sea lift and airlift should be maintained and that reserve forces should be augmented as we lower active strength. In addition, airrefueling tankers, now not needed for strategic missions, can support a U.S. air presence over many distant battlefields more cheaply than maintaining 12 carrier task forces.

Because few people foresee that the U.S. will be the aggressor anywhere in the world, we do not provide for sudden offensive operations requiring a three-to-one advantage. Finally, we do not believe the U.S. should be involved in every civil war conceivable, certainly not without our allies. What threatens Ukraine or Belarus most is not war but economic collapse, which we should help prevent with a policy requiring political leadership, even generosity, and not guns.

Letters selected for publication may be edited for length and clarity. Manuscripts will not be returned or acknowledged without a stamped, self-addressed envelope.

ERRATA

The credit for the illustration on page 28 of the March issue should read "Andrew Hanson/© Wolfram Research."

On page 58 of the April issue, the top left magnetic resonance image scan mistakenly lists the numbers identifying the other scans in reverse order. The slices should be numbered "1 2 3 4 5 6."

50 AND 100 YEARS AGO

JUNE 1944

"Television offers the soundest basis for world peace that has yet been presented. Peace must be created on the bulwark of understanding. International television will knit together the peoples of the world in bonds of mutual respect; its possibilities are vast, indeed.—*Norman D. Waters, President, American Television Society.*"

"Statistics show that, while much has been done to reduce industrial accidents, there is a long way to go. For example, from Pearl Harbor until January 1, 1944, 32,078 soldiers, sailors, and marines died as war casualties; 94,000 workers were killed in accidents. The number of workers injured will dwarf the total of war wounded: 45,595 of our armed men were wounded up to January 1, 1944, while 8,800,000 workers were injured."

"One of the most persistent enemies of safe flying—formation of ice on propellers of planes in flight—is now being overcome by a new electrically heated propeller 'skin' that enables the propeller surface to warm up like a sick-bed heating pad. The skin is made by two kinds of synthetic rubber, the outer surface being a thin coating that is tailor-made to conduct electricity instead of blocking its flow."

SCIENTIFIC AMERICAN

JUNE 1894

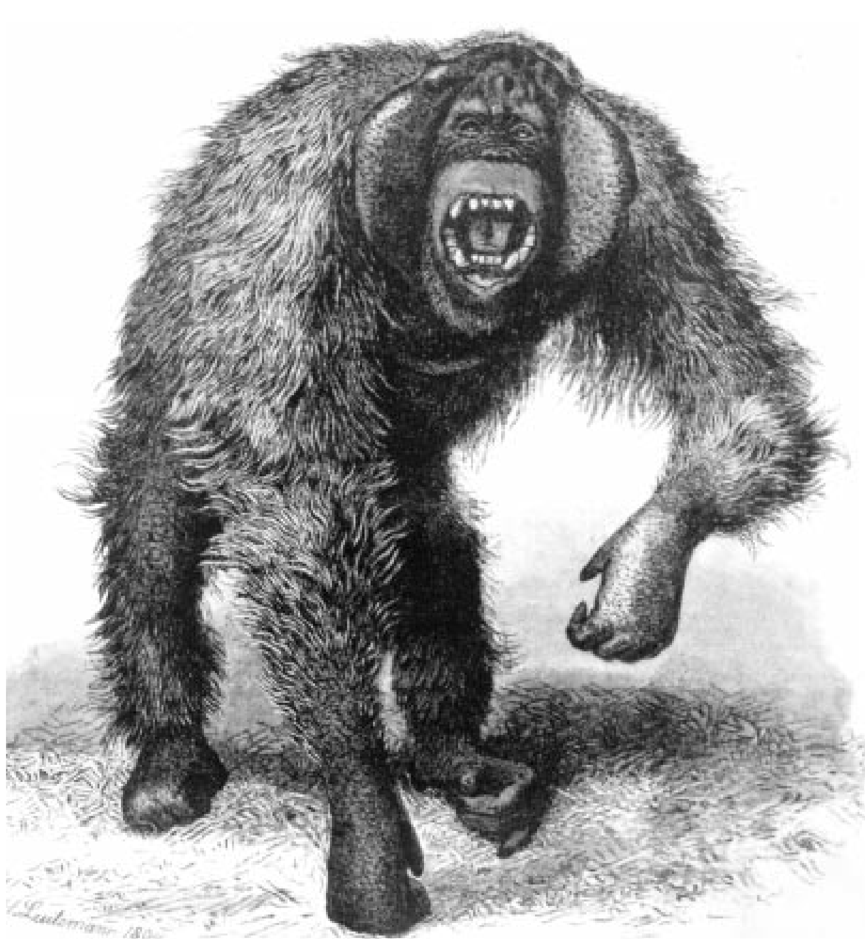
"The tendency of the present day is that the horse must go, must go metaphorically, for his days of labor seem nearly passed."

"The theory is advanced by S. E. Christian, in *Popular Astronomy,* that stellar scintillation is caused largely by inconceivable numbers of small meteoric bodies, which are constantly passing between the stars and our earth. Momentary oscillation of the stars by these bodies would certainly occur if these bodies were numerous enough, and recent investigation seems to point to the fact that they are."

"Mr. Francis Galton affirms that 'the patterns of the papillary ridges upon the bulbous palmar surfaces of the terminal phalanges of the fingers and thumbs are absolutely unchangeable throughout life, and show in different individuals an infinite variety of forms and peculiarities. The chance of two finger-prints being identical is less than one in sixty-four thousand millions. If, therefore, two finger-prints are compared and found to coincide exactly, it is practically certain that they are prints of the same finger of the same person; if they differ, they are made by different fingers.'—*Lancet.*"

"The *Medical Record* tells of a woman in Ohio who utilized the high temperature of her phthisical husband for eight weeks before his death, by using him as an incubator for hens' eggs. She took 50 eggs, and wrapping each one in cotton batting, laid them alongside the body of her husband in the bed, he being unable to resist or move a limb. After three weeks she was rewarded with forty-six lively young chickens."

"We publish to-day an engraving (for which we are indebted to the *Illustrirte Zeitung*) of the gigantic orang-outang in the Zoological Garden at Leipsic, Germany. This and two others that died last winter from the effects of the severe weather are the only full-grown orang-outangs that have ever reached Europe alive. The animal is not as tall as one would suppose from a first glance, for he measures only a little over 4 feet. The orang-outang shown has lost one of his upper eye teeth. Many scars on his hands and feet show that he has led an eventful life and received honorable wounds. His left thumb is bent and one of his toes is crippled. In captivity he eats soaked rice, milk, raw eggs, oranges, dates, and he is very fond of bananas and white bread."



The new orang-outang in the Leipsic Zoological Garden

Population Summit

Women's health and rights shape Cairo document

This fall in Cairo the United Nations will hold its once-a-decade conference on population. And if the third and final preparatory meeting held in April at U.N. headquarters is any indication, the plan the conferees will consider could differ radically from its predecessors. Women in the hundreds—and in the cloth and color of every culture—took over the halls of the U.N., shaping, with unprecedented force, the so-called plan of action that will emerge from the Cairo meeting in September. This document will provide a framework for the next 10 years of U.N. population programs. The Cairo meeting will presumably ratify it, and governments will pledge funding.

The Cairo text covers many of the same issues as did the 1974 Bucharest and 1984 Mexico City plans. The targets include stabilizing the world's population, currently 5.7 billion people, at 7.8 billion by 2050, instead of the projected 12.5 billion. Providing family-planning services to the 350 million couples who want but cannot obtain them continues to be a crucial goal as well.

But the draft plan of action also included phrases and words that never saw the light of day in previous U.N. population documents: reproductive rights, sexual health, female genital mutilation and gender equity. This new emphasis reflects the belief of women's health organizations and family-planning experts that to address issues of population, governments have to address the health of women and their economic and social well-being; coercive national family-planning programs or services that do not take a client's needs or culture into account are doomed to fail. "The field is getting much more sophisticated," notes Joan Dunlop of the International Women's Health Coalition.

Experts say the reason for the change at the U.N. lies in the novel role women and nongovernmental organizations (NGOs) are playing in the diplomatic process. "There are far fewer gray suits," comments Sally Ethelston of Population Action International. "What we are seeing is that the [1992 Earth Summit] opened the doors for NGOs. Particularly in the field of family planning, there is a recognition on the part of the delegates that the NGOs are most innovative. They are the ones that pioneered door-to-door delivery of contraceptives in Bangladesh." Some 900 NGOs were accredited to attend the final preparatory meeting; many delegations include NGO representatives.

The document, as it stood in early April, offered several fresh approaches. They included improving girls' access to education and addressing the contraceptive needs of adolescents as well as the responsibility of men for population growth, their sexual behavior and fertility. Because men stay fertile much longer than women do, the average man, by the end of his lifetime, could be responsible for more children than the average woman, according to Aaron Sachs of the Worldwatch Institute. For instance, "men in Kenya have more children than women do," Dunlop adds. "That is stating the obvious, but it is a very new thought."

But in their efforts to change dramatically the focus of the text, some NGOs have had to battle the tireless efforts of the Vatican to influence the summit. Certain NGO leaders assert that the Vatican's attacks on family planning and

DONNA DECESARE *Impact Visuals*

HEALTH SERVICES FOR WOMEN, such as this family-planning clinic in Egypt, are the focus of the document that will be finalized at the United Nation's International Conference on Population and Development in Cairo this September.

abortion seem especially fierce this time, possibly because the official support it enjoyed from presidents Ronald Reagan and George Bush no longer exists.

Prior to the New York meeting, Pope John Paul II issued a statement calling the International Conference on Population and Development a project to allow the "systematic death of the unborn." The Pope has also written to many national leaders urging them to combat some goals of the conference. At the session itself, the Vatican delegation, led by Monsignor Diarmuid Martin, requested that many references to women and all references to abortion and contraception be bracketed—that is, reserved from approval.

The Vatican's offensive has encountered deeply felt opposition. "One of the extraordinary breakthroughs has been the degree to which women have been outspoken about their distaste for and opposition to the Vatican," Dunlop explains. Some women from countries that are largely Catholic have denounced the Vatican's claim to represent their sex. Many of these women have presented data on the schisms apparent between the church's male leadership and its followers. In the U.S., for example, 87 percent of Catholics believe couples should make their own decisions about birth control, according to a Gallup poll; 84 percent believe abortion should be legal in all or some circumstances.

In a tactical session, Frances Kissling, director of the Washington, D.C.–based Catholics for a Free Choice, wearing a black dress that resembled a priest's robe, urged humor in dealing with the Vatican. Other NGOs have questioned the right of the Vatican to maintain permanent observer status at the U.N., given that Jews, Muslims, Buddhists, Episcopalians and other religious groups do not have the same privilege.

Nevertheless, the Vatican's success in bracketing many terms could ultimately mean that the final language of the plan of action is not as far-reaching as some family-planning experts and women's health advocates would like. If phrases addressing the need for safe abortions—even in countries where the practice is illegal—remain bracketed when they appear in Cairo, the conference may become focused on the abortion debate rather than on population issues. (A study presented at the preparatory meeting by the Alan Guttmacher Institute reported that every year about 2.8 million women have abortions and 550,000 are hospitalized for related complications in six of the Latin American countries where the practice is illegal: Brazil, Peru, Chile, Colombia, the Dominican Republic and Mexico.)

The ultimate outcome of the struggle between some NGOs and the Vatican will only become clear in September in Cairo. Much of the implementation of the plan will depend on how forthcoming governments are with money. The U.N. Population Fund anticipates that the broad-based plan will cost more than $13 billion a year by 2000—some $4 billion is currently spent every year.

In the meantime, the U.N. is a different place. Children sleep on chairs in the corners of conference rooms while their mothers lead discussions on the dangers of self-induced abortion or the informal economic sector. In hallways, men stand out because they seem rare and exotic against the backdrop of blue and gold saris, green and yellow headdresses and the rainbow textiles of Latin America. *—Marguerite Holloway*

Gathering String

Standing a safe distance outside a black hole, toss in a coin. As it nears the black hole's horizon—the point of no return—the coin will seem to fall ever more slowly until it hardly moves. Now suppose that the elementary particles making up the coin resemble not points but tiny bits of string. As they fall in, the strings grow continuously longer. They wind around until they encase the black hole in a giant spaghettilike entanglement.

Odd? An inevitable blend of black hole physics and string theory, says Leonard Susskind of Stanford University. The black hole warps the space-time around it so acutely that time stretches out as in a slow-motion movie—one microsecond for the coin seems to us to be several days or years. Even though the coin does fall into the black hole, we can only see it slow down and come to a stop at the horizon.

Moreover, a string, like the wings of a hummingbird, is always vibrating. Most of the time such movement is just a blur. But catch it in a slow-motion movie, and the vibrating object suddenly looks opaque—and larger. So, too, a string; it grows longer if we are able to see it slowed down. Further, a string vibrates in many different ways. Thus, as it falls toward the black hole, and its microseconds stretch out into minutes or days, it seems from our point of view to elongate endlessly.

This picture would be merely a curiosity if it did not promise to solve what Susskind calls "a puzzle as deep as the constancy of the speed of light was" at the turn of the last century. The puzzle is the information paradox. First posed in 1974 by Stephen W. Hawking of the University of Cambridge, the information paradox notes that objects such as encyclopedias or elephants can fall into a black hole, never to be seen again. What happens to the knowledge they carried, the details about the atoms they were made of? If, as Hawking believed, these are lost forever, then physics is in trouble. Whereas in practice information can be irretrievable, Gerard 't Hooft of Utrecht University has explained, quantum mechanics dictates that in principle the information should still be there in some form.

"Theoretical physicists have been very thoroughly confused for some time," says Edward Witten of the Institute for Advanced Study in Princeton, N.J. One suggested way out of the paradox is that as the coin falls toward the black hole's horizon, its information is somehow scrambled and sent back to us as radiation. Still, the horizon can hold an infinite amount of ordinary matter. Within its finite lifetime, how can the black hole possibly emit the infinite amounts of information the matter must have carried in?

This is where string theory holds out some hope. If strings make up matter, they will spread out and take up all the room at the horizon—allowing the black hole to absorb only a finite amount of material. Presumably information carried in could be encoded in radiation that the strings emit as they fan out.

So is the information paradox solved? "The scenario is plausible and attractive," Witten says, "but there is no smoking gun." String theory is very far from being complete; no one can as yet do all the calculations needed to verify this solution. As Susskind puts it, "Strings can't solve the problems of black holes until they solve their own first." Spaghetti may be on the plate of theorists well into the next century. *—Madhusree Mukerjee*

Sanity Check

Puzzling observations of things that go lump in the night

The farther astronomers peer into space, the more they come to appreciate the intricate structure of the universe at very large scales. In 1987 a group of observers inferred the presence of a vast accumulation of matter, nicknamed the "Great Attractor." Two years later another team discovered the "Great Wall," an aggregation of galaxies at least 500 million light-years across. New celestial surveys that take in larger chunks of the universe hint at still vaster gatherings of galaxies. Theorists find themselves hard-pressed to understand the origin of such enormous structures in a cosmos that, according to present knowledge, started out almost perfectly uniform. "The new surveys are very impressive," says Margaret J. Geller of the Harvard-Smithsonian Center for Astrophysics, "but the state of our ignorance is equally impressive."

Geller should know. Over the past decade, she and a number of colleagues—most notably John P. Huchra, also at the Center for Astrophysics—have produced information that has challenged the most ingenious theorizing. What the researchers do is measure the redshift (the stretching of light caused by the expansion of the universe) of thousands of galaxies. The redshift in turn indicates the galaxies' approximate distances from the earth.

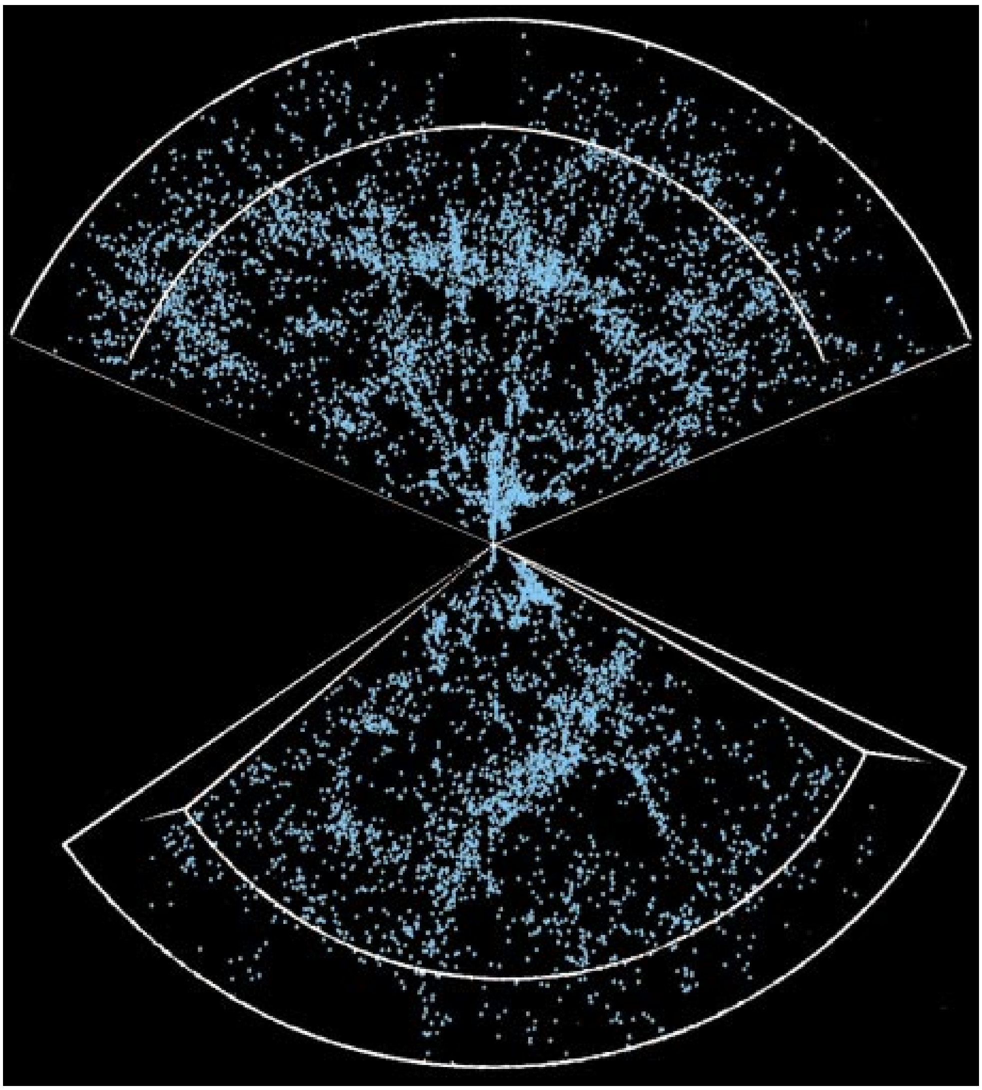
Those efforts have led to an increasingly comprehensive set of maps that show galaxies located along the bubblelike surfaces of enormous "voids." These comparatively empty regions measure as much as 150 million light-years in diameter (for comparison, the Milky Way is only about 100,000 light-years across). The Great Wall is more like a sheet of galaxies that outlines voids.

The discovery of the Great Wall has raised two crucial questions: Are such formations typical of the universe as a whole, and does the universe contain even larger structures? In their search for an answer, researchers at the Center for Astrophysics teamed up with a number of astronomers working in Argentina, Chile and South Africa. Observatories in those locations can scrutinize southern parts of the sky that are invisible from the Whipple Observatory in Arizona, where most of the earlier mapping was done. Luis Nicolaci da Costa of the Brazilian National Observatory, a former graduate student at the Center for Astrophysics, headed the group that conducted the mapping of galaxies in the southern sky.

Nearly 3,600 galaxies appear in this latest survey. The distribution of galaxies in the southern sky shows a "gross similarity" to that seen in the north, Geller reports. For example, da Costa and his co-workers have uncovered a second feature much like the Great Wall, which is known—predictably—as the Southern Wall.

Yet statistical analysis reveals that "there are some differences in certain measures," according to Geller. Such differences are significant because they imply that parts of the universe contain structures even larger than the extent of the current north-south sky map. Otherwise, every section of the universe should, when viewed in terms of statistical averages, look like any other section. Da Costa and his fellow team members conclude that the nature of the "shells" of galaxies seen in the map varies over a scale of 300 million light-years or so. Even larger structures may be out there, simply too large to show up in the current study.

In the past few years, several groups of researchers have found that the universe displays another, unexpected kind of departure from uniformity. The Milky Way and all the galaxies around us seem to be rushing headlong in the direction of the constellation Leo; that motion appears superimposed on the



HARVARD-SMITHSONIAN CENTER FOR ASTROPHYSICS

COSMIC ROAD MAP shows the irregular distribution of roughly 11,000 bright galaxies* (blue dots); *the newly discovered Southern Wall runs diagonally across the lower slice of sky.

more general cosmic expansion associated with the big bang. In 1987 Alan M. Dressler of the Observatories of the Carnegie Institution of Washington and his six collaborators (known as the Seven Samurai) analyzed those motions and concluded that they result from the gravitational pull of some vast mass, which they called the Great Attractor.

Intrigued by that finding, Tod R. Lauer of the National Optical Astronomy Observatories in Tucson and Marc Postman of the Space Telescope Science Institute in Baltimore began what they call a "sanity check" to make sure the Great Attractor is real. The two researchers measured the motions of galaxies in a region 30 times the volume of space examined by Dressler's group. If the Great Attractor is just a discrete, local feature, Lauer explains, then it should show up as a zone of aberrant galaxy motions embedded within a larger group that shows no net motion.

Lauer and Postman studied the brightest elliptical galaxies in 119 galaxy clusters lying at distances of up to 500 million light-years from the earth in all directions. Previous work has shown that giant elliptical galaxies have a fairly consistent intrinsic luminosity, so their apparent brightness alone betrays their distance. The two researchers then measured each galaxy's redshift, which reveals its velocity, and compared it with the value expected for an object at that distance.

Over very large scales—a billion light-years or so—Lauer and Postman, like most of their colleagues, expected that the spread of matter through the cosmos would be very even. If so, the galaxies should appear, on average, at rest with respect to the cosmic microwave background, relic radiation from the time of the big bang that continues to fill the universe.

When he and Lauer looked at their results, Postman recalls, they were "surprised, to say the least": the entire group of galaxies appeared to be fleeing in the direction of the constellation Virgo at a speed of roughly 700 kilometers per second. The boggling implication is that some tremendous clump of matter located beyond the edge of the surveyed region is pulling at all the galaxies Postman and Lauer observed (including, of course, our own Milky Way). The Great Attractor, it seems, is only a small part of an even greater conglomeration of galaxies. "It's a very difficult measurement, and they've done a wonderful job," concludes P. James E. Peebles of Princeton University.

Such huge structures perplex the cosmologists who try to piece together the story of how the modern universe came to be. Data collected by the *Cosmic Background Explorer* satellite showed that the microwave radiation left over from the big bang (and, by extension, the matter that was embedded in that radiation) is very nearly featureless. Somehow gravity pulled together lumps and blobs of gas into galaxies, stars, planets and people. Given enough time, gravity could magnify extremely slight irregularities into distinct formations. But the latest crop of walls and attractors intensifies the mystery of how so much structure could have formed within the 15-billion-year age of the universe.

Many research teams around the world are racing to collect more observations in order to test the models and learn more about the processes that transformed the primordial blur into the modern, highly organized cosmos. Lauer and Postman plan to expand the volume of their survey fivefold. Postman also expresses great enthusiasm for a massive, multi-institution digital sky survey, headed by Donald G. York of the University of Chicago, which will collect data on one million galaxies, starting next year.

Cosmologists have frequently underestimated the baffling complexity of the universe, which is increasingly evident through modern telescopes. "I really don't think we understand how structure forms in the universe," says Geller in a cautionary tone. "It is a tough, tough problem, much harder than people realized when I was starting out. Answers are not just around the corner." —*Corey S. Powell*

Bright Spot

Here is another progress report from the "smaller, fewer, weirder" front in quantum physics. Researchers at AT&T Bell Laboratories have formed what may be the smallest and certainly the most evanescent laser ever. It consists of a gallium arsenide quantum wire in which electrons can move in only one dimension. The next step in the technology will meet the weirdness criterion.

The AT&T group, headed by Loren Pfeiffer, guessed that if energy were pumped into a one-dimensional space, or "wire," in semiconducting material, the electrons and holes would have little choice but to bind to one another and form particles called excitons. The excitons, which would be in an energetic ground state, would collapse and emit photons at a single wavelength. Pumped with energy from laser light, and more recently powered by a battery, the wire laser met the workers' expectations. As they varied the pumping power by two orders of magnitude, the material emitted stable, monochromatic red light.

Because of their size and stability, these lasers may be able to transmit more information with less interference than can their larger, three- and two-dimensional predecessors. They would also allow photonic technology to complement electronic technology on the quantum scale toward which computing and communications devices are shrinking. Striving for weirdness may prove eminently useful. "Now that we finally have a quantum wire laser," Pfeiffer says, "we can measure whether it has useful properties or not."

Indeed, making a quantum wire laser was a major challenge. The first step, using molecular-beam epitaxy (MBE), is to lay down a crystal film only a few atoms thick. Such a film, called a quantum well, is thinner than an electron's wavelength is long. Thus, the particle has only two dimensions in which to move. How can a second dimension be removed from such structure?

At the end of last year, Pfeiffer's group reported a solution to the problem. Drawing on elementary geometry, his team formed a one-dimensional electron conduit by growing quantum wells, each 70 angstroms wide, at right angles to one another. The T-shaped intersection of the films is in effect a continuous wire, 70 angstroms wide and some 600 microns long. "Our method may not be feasible for large-scale production," Pfeiffer says. "We were interested in making an ideal one-dimensional quantum wire so that we could study its laser properties first." He may have a point: MBE has also been rendered by others as megabuck evaporation.

What's next? Weirdness, of course, in the form of a zero-dimension, quantum dot laser. The group plans to grow a well across one end of a quantum wire. Three perpendicular quantum wells would then intersect at a single point. "One of my goals this year is to see the luminescence from a quantum dot structure," Pfeiffer says. For such a small feat, it would be a glowing achievement indeed. —*Kristin Leutwyler*

La Ronde

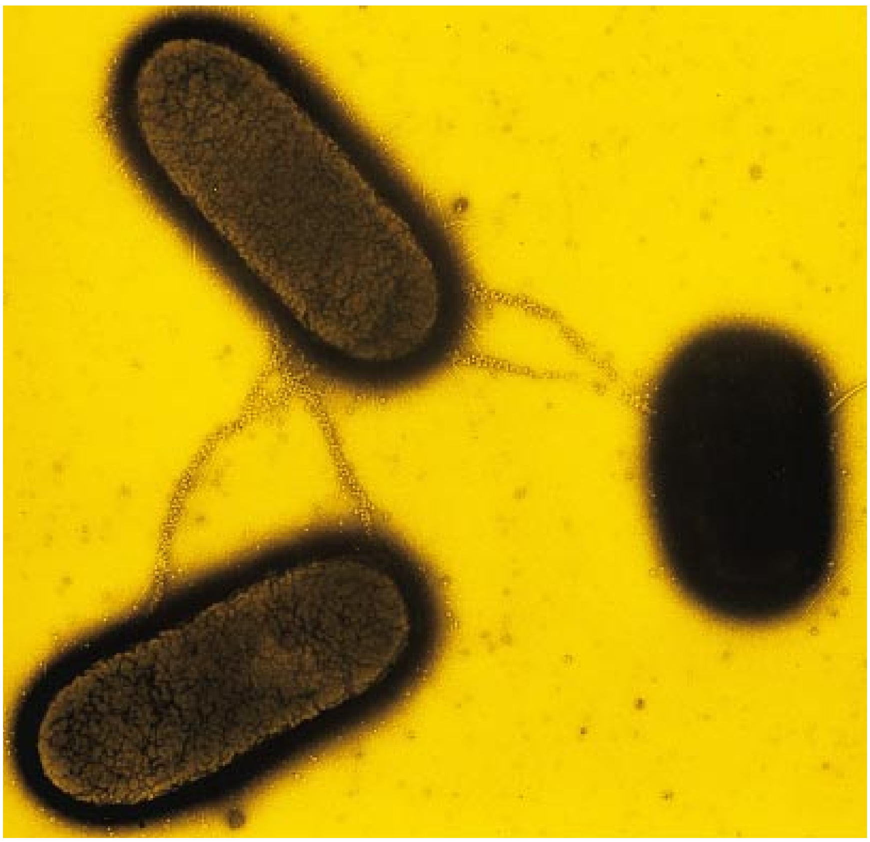
What goes around comes around for life's master molecule

Evidence is rapidly accumulating that a blizzard of genetic material blows freely through the microbial world—not only between bacteria of the same species but also between members of distantly related species and between bacteria and viruses. "In terms of the flux of DNA, the general impression is that it goes anywhere and everywhere," says Julian E. Davies, a microbiologist at the University of British Columbia. And although the genetic material of multicellular plants and animals tends to be better buttoned up, the exchange involves them, too.

Recent research has revealed how some of this promiscuity may come about. Since the 1920s bacteria have been known to exchange genetic material among their own kind. One method, conjugation, is the bacterial version of sex: genes are transferred from one bacterium to another through a special tube. In 1958 Joshua Lederberg shared a Nobel Prize for investigations that made use of bacterial conjugation.

In the 1980s conjugation began to attract more than just scholarly attention when researchers found clues that genes were spreading between species. In 1985 Patrick Trieu-Cuot proved that genes were indeed moving between distantly related bacteria by showing that neomycin- and kanamycin-resistance genes in three different species were virtually identical. Often bacterial genes are transmitted on plasmids: small, parasitic loops of DNA that are distinct from the bacterial chromosome. Some striking findings have come from Abigail A. Salyers of the University of Illinois. She has shown that when bacteria are exposed to the antibiotic tetracycline, they use a variety of methods, some still mysterious, to accelerate the exchange of genes for tetracycline resistance.

In the laboratory at least, environmental stress appears to enhance conjugation across species lines. German workers have found a possible explanation. Alfred Pühler and his colleagues at the University of Bielefeld showed that heat, acids, alkalis and alcohol all inhibit the action of enzymes in *Corynebacterium* that cut up foreign DNA. As a result, *Corynebacterium* subjected to such treatments became more accepting of DNA from *Escherichia coli.* Pühler notes that if environmental stress promotes gene exchange between bacterial species, genes deliberately engineered into microorganisms might spread more easily in nature than they do in the laboratory.



L. CARO *Science Photo Library/Photo Researchers, Inc.*

DNA PASSES through bridges linking individual bacteria in the process known as conjugation, shown here taking place between a "male" and two "females." Microbiologists have learned that conjugation also occurs between distantly related species.

Transformation is another mechanism that bacteria use to exchange DNA. It occurs when a bacterium absorbs naked DNA in the environment. The DNA may have been left lying around either by an experimenter or by some other organism, possibly one that has died. Because DNA is chemically not very stable outside of cells, transformation is probably less important in nature than is conjugation. Nevertheless, Guenther Stotzky of New York University and Marilyn Khanna, now at Cornell University Medical College, have shown that montmorillonite—a mineral better known as clay—can absorb and bind DNA in such a way that it is protected. The bound DNA can then be taken up by other bacteria.

The third major mechanism for DNA exchange in bacteria is transduction. It occurs when viruses that attack bacteria—known as bacteriophages—bring with them DNA they have acquired from their previous host. Because most bacteriophages have a restricted number of hosts, transduction probably does not routinely transmit genes between distantly related species of bacteria. Still, Gustaaf A. de Zoeten, chair of the botany and plant pathology department at Michigan State University, says, "Viruses are even worse than bacteria—they evolve by the exchange of whole functional genetic units." Fungi and plants are by no means immune to the pervasive DNA flux. The bacterium *Agrobacterium tumefaciens* has long been known to transfer plasmids to plants. And in 1989 Jack A. Heinemann of the University of Oregon proved that bacterial plasmids could be transmitted to yeast through a process very much like conjugation.

Experiments reported in *Science* in March by Ann E. Greene and Richard F. Allison of Michigan State indicate that plant viruses can combine the RNA that constitutes their genes with RNA copied from the genes of genetically engineered plants. Although the situation Greene and Allison investigated was artificial, plants engineered to contain useful viral genes may be commercially available within a few years. De Zoeten believes Greene and Allison's results mean more research is still necessary to establish the safety of such plants.

So far the evidence is slight for massive and long-lasting gene exchange between different species of multicellular animals or plants. But it would be unwise to assume that animals are completely out of the loop. In 1985 Joe V. Bannister and his colleagues at the University of Oxford found indications that genes from a species of fish had been transferred to bacteria. And genes that are introduced into humans by viruses probably have their origins in other animals.

What are the implications of interspecific gene transfer for evolution? Although the phenomenon is plainly a real one, little is yet known about how often it occurs. The standard neo-Darwinian picture in evolution, in which mutation is the main engine for introducing genetic novelty, has proved extremely powerful over the past half a century. But it seems evolution has some wrinkles that even Charles Darwin did not foresee. —*Tim Beardsley*

It's Just a Phase

Water, ice and steam might be the first examples that come to mind when describing various phases of matter. But to a physicist, any unusual configuration of particles or entities may also qualify as a new state. For example, electrons might organize themselves into a pattern called a charge-density wave. Another phase is the Luttinger liquid. Although not something one can drink, it represents a unique collective behavior of electrons in a conducting medium.

Under normal circumstances, electrons in conductors constitute a Fermi liquid. They form a sea of negative charge. In such a liquid, a single electron does not respond to the individual charges of other electrons present in the material. In effect, the Fermi liquid consists of noninteracting particles, which enables an electron to roam fairly freely through the substance. This picture explains in part how electrons in a conductor can transmit current.

During the 1960s, Joaquin M. Luttinger of Columbia University explored situations in which electrons are forced to interact with one another. For a simplified, one-dimensional case, he solved the equations that defined this state (a so-called many-body problem). There the matter mostly stayed until advances in technology and the discovery of high-temperature superconductivity led to an intense reexamination of the activity of electrons in solids.

Last year Charles L. Kane of the University of Pennsylvania and Matthew P. A. Fisher of the University of California at Santa Barbara and their colleagues squeezed a verifiable prediction from Luttinger's calculations. At the March meeting of the American Physical Society, Richard A. Webb of the University of Maryland presented the first experimental evidence.

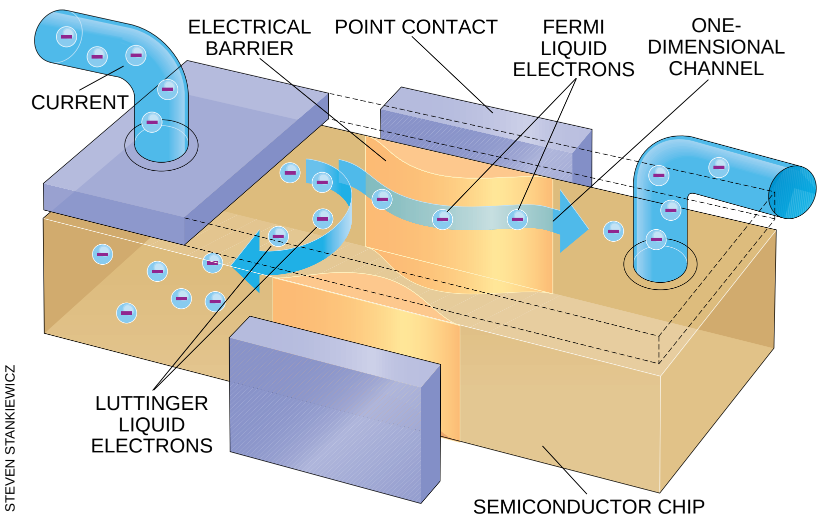
"The theory is rather specific in how you have to set the system up," Webb observes. The electrons must flow through a one-dimensional channel that can be obstructed in the middle. A point contact can create this blockage by acting as an electrical vise. Researchers simply apply a voltage to the point contact, which in essence pinches off the channel and thereby reduces the conductance.

As a Fermi liquid, electrons would occasionally tunnel through the obstruction; some conductance would always remain in the channel. Not so for a Luttinger liquid. At temperatures near absolute zero, each electron in this state would feel the individual charge forces from other electrons. This effect would serve to correlate their behavior. The correlation would manifest itself as a characteristic drop in the conductance through the point contact; eventually all the electrons would be reflected by the barrier.

"You would think the experiment is easy, but it's not," Webb says. "I worked on it on and off for two years." Collaborating with Frank P. Milliken and Corey P. Umbach of the IBM Thomas J. Watson Research Center, Webb finally created the Luttinger liquid in a semiconductor made of gallium arsenide. The theory stated that the particular signature of the liquid would appear only for ballistic electrons—that is, electrons that move in one direction without scattering. The source of the ballistic electrons was the fractional quantum Hall effect. This phenomenon refers to the sideways drift of electrons as they move through a sample exposed to an external magnetic field. Xiao-Gang Wen of the Massachusetts Institute of Technology had pointed out that under the correct conditions, these electrons move ballistically.

The Luttinger liquid is not likely to find applications. It destabilizes at temperatures higher than one degree above absolute zero. Its real value may be that investigators can now see how electrons truly interact with one another in a solid. Conventional methods of analyzing many-body problems demand a mixture of intuition and an approximation scheme called perturbation theory.

"The beauty of the Luttinger liquid is that the electron-electron interaction can be treated exactly," Kane explains. "It's an example of a many-body problem that you can really solve." Webb concurs: "It is one more tool in our bag to understand the physics of small structures." Now that's something you can drink to. —*Philip Yam*



LUTTINGER LIQUID forms in a channel etched into a semiconductor chip. In this state, electrons are reflected by an electrical barrier erected by a point contact. In contrast, electrons in a Fermi liquid can tunnel through the obstacle.

Shooting the Rapids

A new environmental agency navigates Potomac currents

Change in the natural world spans decades, even centuries. It follows that long-term monitoring is the only way to identify harmful trends. Yet human institutions operate on the basis of months, or years at best. Members of Congress run for reelection every two or six years. Many corporate managers live—and die—by quarterly results. Tenured professors scramble annually for research grants. How, then, can existing bodies identify environmental problems and assess the effectiveness of measures taken to mitigate them?

They cannot, argue the founders of the Committee for the National Institute for the Environment (CNIE). What is needed, they suggest, is their eponymous institution. The National Institute for the Environment would be a new federal agency that would sponsor research on critical environmental issues. Proponents say it could serve as an early-warning system for such ominous developments as global warming, stratospheric ozone depletion and the decline of biodiversity. Because the institute would be governed by an independent board, it would be relatively immune to political pressure.

The idea of such an organization was conceived more than five years ago by Henry F. Howe of the University of Illinois and Stephen P. Hubbell of Princeton University. Recently the CNIE appointed a high-profile president, Richard E. Benedick. Benedick, a former state department ambassador, was the principal force behind the 1987 Montreal protocol on chemicals that harm the ozone layer. He is currently a special adviser to the 1994 International Conference on Population and Development to be held in Cairo. The CNIE has so far secured the support of more than 6,000 scientists, numerous environmental organizations and at least one senior government official, Secretary of the Interior Bruce Babbitt.

There is opposition. Robert T. Watson, associate director for environment in the Office of Science and Technology Policy, says he "agrees completely" with the CNIE that current government research efforts are too short-term in focus and poorly coordinated. But he suggests instead redirecting some of the $4 billion to $6 billion the government already spends annually on environmental research. Watson maintains that a committee newly established under the National Science and Technology Council, the Committee on the Environment and Natural Resources, is a "virtual agency" that should achieve many of the CNIE's goals.

Others wonder how a new agency would be linked to existing institutions. Robert C. Szaro, a deputy research director of the U.S. Forest Service, complains that the CNIE seems to lack "any real recognition of what federal government scientists already do." He adds: "I don't think the CNIE supporters would have the exclusive role they think they would have" in ecological research. The National Research Council issued a report last year that considered the CNIE's plan but came down in favor of a department of the environment that would subsume several existing agencies. The Carnegie Commission on Science, Technology and Government has likewise supported creating an agency out of existing programs.

Benedick points out that an independent institute for the environment would have backers in Congress who could ensure continued funding even if a future administration were hostile. Furthermore, he says such an institute could bring in funds from industry.

To move from president of a committee to head of an institute, Benedick will have to persuade 218 representatives and 51 senators of the wisdom of the CNIE's plan. Success, if it comes, is unlikely to be in 1994: for now, the people's elected officials are far too busy navigating Whitewater, grappling with health care and courting a disgruntled, skittish electorate. —*Tim Beardsley*

SARAH LEEN *Matrix*

SCIENTISTS at the Dorset Research Center in Ontario test the acidity of Lake Muskoka, near the U.S. border. Sulfur dioxide from burning fossil fuels has acidified many lakes in the U.S. and Canada. Long-term monitoring is needed to track changes.

The Last Universal Mathematician

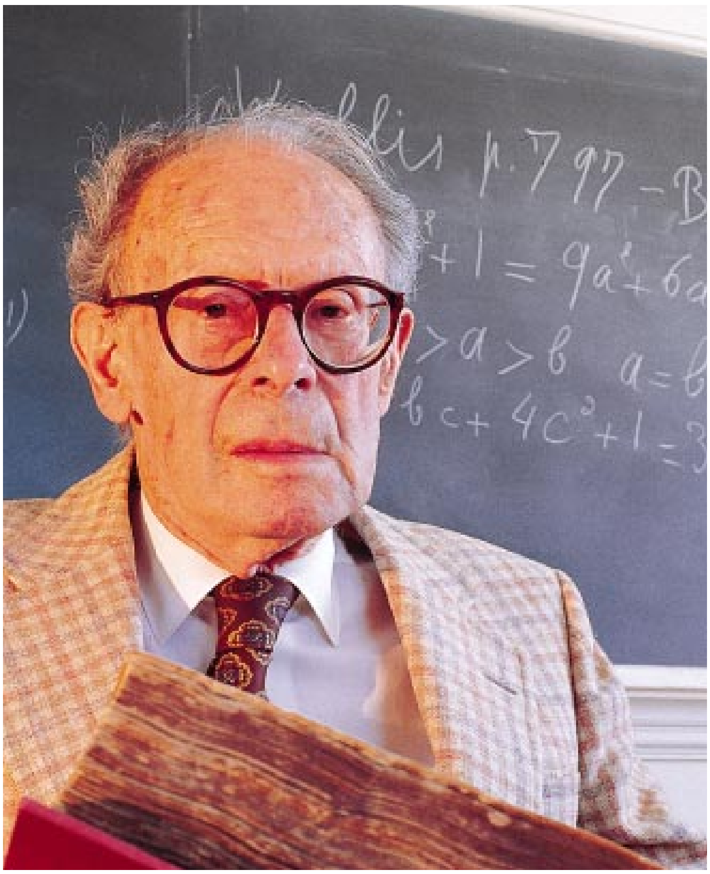
In 1939 a 33-year-old French mathematician proved that a profound conjecture about the behavior displayed by prime numbers as they meander toward infinity holds true for certain limited but crucial cases. The achievement, which is known as the proof of the Riemann hypothesis on the Zeta function for field functions, is a jewel of modern number theory. It is all the more remarkable because its author first scribbled it down in a French military prison.

This is only one in a series of extraordinary incidents in the life of André Weil, who eventually left his prison cell to become one of the 20th century's greatest mathematicians. Yet so isolated is mathematics from the rest of human culture that Weil, now a professor emeritus at the Institute for Advanced Study in Princeton, N.J., remains largely unrecognized outside his field. When Weil's autobiography, *The Apprenticeship of a Mathematician,* was published three years ago, not a single nonmathematical publication reviewed it. Weil's younger sister, Simone Weil, a philosopher and political activist, is more widely known in spite of the fact that she died more than 50 years ago.

Professional colleagues are therefore eager to praise Weil. They call him the last of the great "universal" mathematicians. They point out that he was a founder of Bourbaki, a legendary group that in the guise of a fictitious sage—Nicolas Bourbaki—wrote a series of monumental treatises that brought order and unity to mathematics. Weil himself navigated all the major tributaries of mathematics—notably, number theory, algebraic geometry and topology—erecting proofs and conjectures that, like levees, determined the future course of inquiry. One of these conjectures played a crucial role in the celebrated "proof" of Fermat's Last Theorem, perhaps the most famous unsolved problem in mathematics, announced last year by Andrew Wiles of Princeton University.

Weil's style has been as influential as his specific contributions. One number theorist likens him to a medieval monk doing work with "tremendous simplicity and purity and no unnecessary ornament." Weil "was always after what was essential," another agrees. Weil was reportedly feared for his sharp tongue as well as admired for his brilliance. One compatriot, comparing Weil to a violin whose strings have been stretched too tightly, recalls that "he suffered fools very badly." The colleague suggests Weil may have mellowed with age.

Indeed, Weil is 88 now, equipped with a hearing aid and plastic hip joints. And during an interview at the Institute for Advanced Study, he seems, at times, almost serene. Asked if he is bothered by the fact that so few people know of his work and even fewer can appreciate it, he gives a Gallic shrug. "Why should I be?" he replies. "In a way, that makes it more exciting."



JASON GOLTZ

ANDRÉ WEIL: "Always after what was essential."

Unlike some modern purists, Weil is also unconcerned by the growing collaboration between mathematics and physics (spurred in part by Edward Witten, a theoretical physicist whose office abuts Weil's). "I have lived through a period when physics was not important for mathematics," Weil comments. "Now we are coming back to a period where it is becoming important again, I think, and that is a perfectly healthy development."

Yet there are flashes of acerbity. When asked his opinion of Wiles's assault on Fermat's Last Theorem, Weil jokes at first that centuries hence historians will think he and the similarly named Wiles are the same person. Then his smile fades, and he adds, "I am willing to believe he has had some good ideas in trying to construct the proof, but the proof is not there. Also, to some extent, proving Fermat's theorem is like climbing Everest. If a man wants to climb Everest and falls short of it by 100 yards, he has not climbed Everest."

Explaining why his autobiography describes his life only through World War II, Weil offers another barbed response. "I had no story to tell about my life after that," he says. "Some of my colleagues have written so-called autobiographies, which I think are very boring. They consist entirely of saying, 'In the year such and such I was appointed to such and such an institution, and in such a year I proved this or that theorem.' "

Weil's life, at least its first half, was almost excessively eventful. He was born in Paris in 1906. Both his father, a physician, and his mother devoted themselves to all aspects of culture. By his early teens Weil had become "passionately addicted" to mathematics. He graduated from the University of Paris in 1928, after having delivered a Ph.D. thesis that solved a 25-year-old problem about elliptic curves posed by Henri Poincaré.

Weil had renounced philosophy as a fatuity years earlier, after he received a good grade on a philosophy test despite having read none of the relevant texts. "It seemed to me that a subject in which one could do so well while barely knowing what one was talking about was hardly worthy of respect," he wrote in his autobiography.

Not that he lacked other interests. His fascination with Indian culture—and in particular the Hindu epic the Bhagavad Gita—contributed to his decision to accept a teaching position in India in 1930. After two years, he became entangled in India's arcane academic politics and was fired, but not before meeting Gandhi. Weil sipped tea with the Indian leader as he was planning the revolt that toppled the British Raj.

On his return to France, Weil became a professor at the University of Strasbourg. In 1937 he married Eveline, who had a son by a previous marriage (she died in 1986). Two years later, as Germany grew increasingly belligerent, the French government ordered Weil to report for military service. Instead he fled to Finland, which at that point the Soviet Union had not invaded. Weil admits to some lingering ambivalence over his decision to avoid service. "My basic idea, which was correct, I think, was that as a soldier I would be entirely useless, and as a mathematician I could be of some use," he says. "Of course, that was in the days of Hitler, and I was entirely of the opinion that the world should not yield to Hitler, but I couldn't see myself taking part in that effort."

Unfortunately, the young professor typing abstract symbols hour after hour in the countryside aroused the suspicions of the Finns, who were fearful of a takeover by the Soviet Union. The Finnish police arrested Weil and—according to one account related to Weil subsequently—nearly executed him before learning that he was merely a French mathematician avoiding the draft. Weil's troubles did not end there. The Finns turned him over to the French authorities, who promptly convicted him of desertion and imprisoned him again.

Weil spent six months in jail, where he created his theorem on the Riemann hypothesis, before being released in exchange for agreeing to join the French army. His ability to make the most of his incarceration provided much amusement for colleagues in later years. Once when Weil made an uncharacteristic misstep during a lecture, the eminent mathematician Herman Weyl suggested that Weil return to prison so he could work out the problem.

After the Germans routed the French army, Weil fled to England. He eventually made his way with his wife and stepson to the U.S., where he began searching for a job. Weil was already sufficiently filled with self-regard that he was chagrined when the only institution that initially offered him a paid position was Lehigh University in Pennsylvania. On leaving Lehigh after two unhappy years in what they felt was an intellectual wasteland, he and his wife vowed never to utter its name again. Henceforth they would call it "the unmentionable place." In his autobiography, Weil uncharitably recalls Lehigh as a "second-rate engineering school attached to Bethlehem Steel."

In 1947, after a stint in Brazil, Weil moved to the University of Chicago, where he resumed his work on Bourbaki. The project had begun in the mid-1930s, when Weil and half a dozen French colleagues, concerned about what they felt was the lack of adequate texts on mathematics, vowed to write their own. They decided that rather than publishing under their own names, they would invent a pseudonymous figurehead: Nicolas Bourbaki, an eminent professor who hailed from the (also fictitious) eastern European nation of Poldavia.

At first, few people beyond their immediate circle guessed the true identity of Bourbaki. As the group churned out vast treatises on virtually every field in mathematics, however, doubts grew. In 1949 Ralph Boas proclaimed in an article in the *Encyclopaedia Britannica* yearbook that Bourbaki was a pseudonym and did not exist. Weil wrote a letter, in high dudgeon, denying the accusation. Bourbaki's members then began circulating rumors that Boas did not exist.

Although younger mathematicians have continued to perpetuate the legacy of Bourbaki, its influence has waned. Weil himself, who resigned from the group in the late 1950s, thinks "in some ways the influence has been good. In some ways it has not been good." Perhaps the most important contribution of Bourbaki was to carry out a famous proposal made by the great German mathematician David Hilbert in 1900 that mathematics be placed on a more secure foundation. "Hilbert just said so, and Bourbaki did it," Weil declares. Bourbaki's emphasis on abstraction and axiomatics was sometimes carried too far, but Weil emphasizes that it was not Bourbaki itself but its followers who perpetrated these crimes.

Weil dismisses the argument of some philosophers that a celebrated theorem proved by Kurt Gödel in the 1930s shows that attempts to systematize mathematics are ultimately futile. "It's a perfectly good mathematical proof," he says. "The philosophical importance is something else that does not interest me." So averse is Weil to philosophizing that he even claims to be an agnostic on the old question of whether mathematical truths are discovered or invented. In his autobiography, Weil describes "the state of lucid exaltation in which one thought succeeds another as if miraculously, and in which the unconscious (however one interprets that word) seems to play a role." Yet he denies that such inspiration might stem from an external or even divine source. Tapping his forehead, he exclaims, "I think it's *there!*"

In 1958 Weil came to the Institute for Advanced Study, where he kept probing for deep links between arithmetic, algebra, geometry and topology. These unification efforts spawned what has become arguably the most vital field of inquiry in modern mathematics. Although he officially retired from the institute in 1976, Weil still goes to his office almost every day. There he pursues an old passion, the history of mathematics. He is currently helping to edit the works of two previous French giants, Jacques Bernoulli and Pierre de Fermat.

The last universalist confesses he has difficulty following recent developments in mathematics: "Mathematics has passed me by, which is as it should be, of course." Although he thinks computers can be useful tools, he rejects the suggestion that they may become crucial for constructing proofs as mathematics becomes more complex. He contends that the use of computations in certain proofs—such as the famous four-color theorem—is only a temporary crutch. "I'm sure when something is proved by computers it will later be proved without computers."

On the other hand, Weil doubts whether any human can ever again have a grasp of all of mathematics. One problem, he says, may be that there are too many mathematicians, especially good ones. "When I was much younger, I thought there was a danger that mathematics would be stifled by the abundance of mediocre mathematics being produced. And now I am inclined to think that its greatest danger is that too much good mathematics is produced. Things are going too fast. Nobody can keep up with it all." —*John Horgan*

SCIENTIFIC AMERICAN

Was the Race to the Moon Real?

In 1961 President John F. Kennedy made the goal to be first on the moon a matter of national honor. But were the Soviets truly in the running?

by John M. Logsdon and Alain Dupas

Twenty-five years ago, on July 20, 1969, Neil A. Armstrong took the first footsteps on the surface of the moon. That event marked a political and technological victory for the U.S. in its cold war rivalry with the U.S.S.R. In the years that followed, the Soviet government insisted that the Soviet Union had never planned a lunar landing. Hence, it argued, the contest to send humans to the moon was a one-sided exercise. The reality is otherwise; recently declassified information from that era and testimony of key participants in the Soviet space program under Khrushchev and Brezhnev prove that the moon race was indeed real.

New evidence reveals that personal rivalries, shifting political alliances and bureaucratic inefficiencies bred failure and delays within the Soviet lunar-landing program. In contrast, the American effort received consistently strong political and public support. The National Aeronautics and Space Administration and its contractor teams also benefited from a pool of skilled and highly motivated workers and managers. Despite an early Soviet lead in human space exploration, these factors, along with more generous and effective allocation of resources, enabled the U.S. to win the competition to be first to the moon.

Soviet capability in space became clear to the world in October 1957, when the U.S.S.R. lofted *Sputnik 1,* the first artificial satellite. Two years later the Soviets launched a probe that returned close-up images of the lunar surface. And on April 12, 1961, cosmonaut Yuri A. Gagarin became the first human in space. Soviet officials cited each accomplishment as evidence that communism was a superior form of social and economic organization. The Soviet advantage in space rocketry underlined fears in the U.S. that a missile gap existed between it and its adversary, an issue that Kennedy belabored in the 1960 presidential campaign.

At first, the shape that a U.S.-Soviet space race might take was not clear. Indeed, if President Dwight D. Eisenhower had had his way, there might not have been one at all. Eisenhower rejected the idea that spectacular space achievements had anything to do with the fundamental strength of a country; he consistently refused to approve space programs justified on purely political grounds. In July 1958, however, he created NASA, an agency that brought together the resources to establish a U.S. civilian space program.

GIANT ROCKETS needed to transport humans to the moon were developed in both the U.S.S.R. and the U.S. The Soviet N-1 rocket (*opposite page*) failed in each of its four test launches before its development was canceled. The U.S. Saturn V (*left*), in contrast, proceeded roughly on schedule and successfully carried Americans to the moon in July 1969.

It was inevitable, perhaps, that NASA would argue that such a program should be ambitious.

Eisenhower's successor, President John F. Kennedy, perceived a much more direct link between space exploration and global leadership. Stimulated by the worldwide excitement generated by the Gagarin flight, Kennedy decided that the U.S. had to surpass the Soviets in human spaceflight.

On April 20, 1961, just eight days after the Gagarin flight, Kennedy asked Vice President Lyndon B. Johnson, "Is there any...space program that promises dramatic results in which we could win?" In particular, Kennedy inquired, "Do we have a chance of beating the Soviets by putting a laboratory in space, or by a trip around the moon, or by a rocket to land on the moon, or by a rocket to go to the moon and back with a man?" Johnson, whom Kennedy had named his primary adviser on space policy, promptly organized an intense two-week assessment of the feasibility of these and other alternatives. A series of memoranda trace the evolving response to Kennedy's questions.

One of the many people Johnson consulted was Wernher von Braun, leader of a team of rocket engineers whom the U.S. Army had spirited out of Germany during the last days of the Third Reich. In a memorandum dated April 29, von Braun told the vice president that "we do not have a good chance of beating the Soviets to a manned laboratory in space," but "we have a sporting chance of sending a three-man crew around the moon ahead of the Soviets," and "we have an excellent chance of beating the Soviets to the first landing of a crew on the moon."

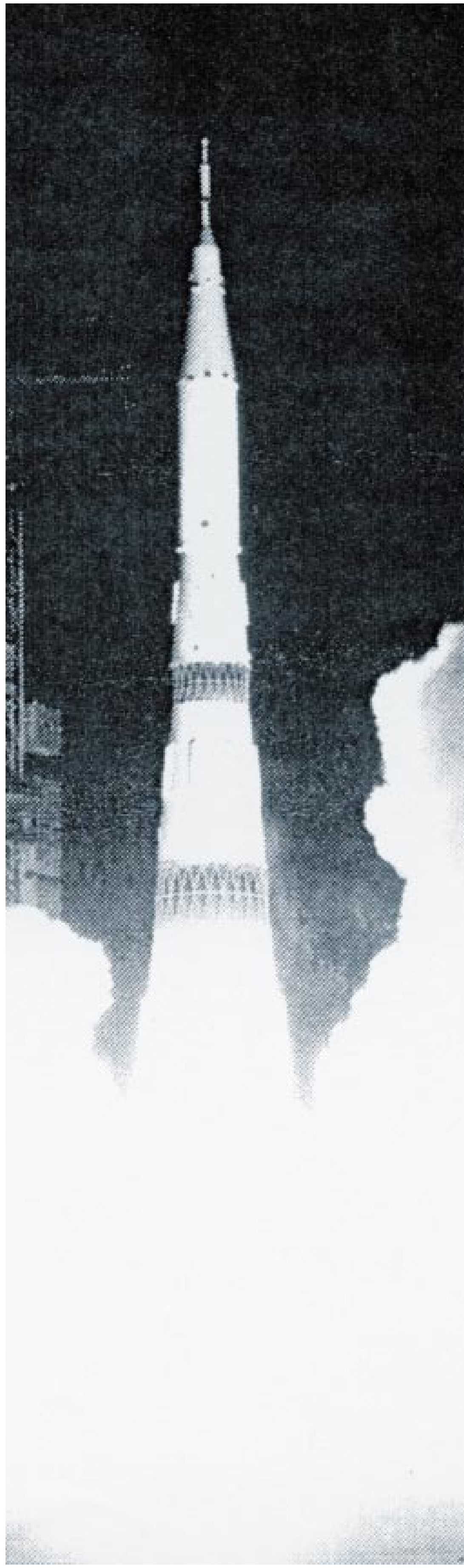
Von Braun judged that a lunar landing offered the U.S. the best opportunity to surpass the Soviets because "a performance jump by a factor 10 over their present rockets is necessary to accomplish this feat. While today we do not have such a rocket, it is unlikely that the Soviets have it." He suggested that "with an all-out crash effort, I think we could accomplish this objective in 1967/1968."

On May 8, 1961, Johnson presented Kennedy with a memorandum that reflected the results of his investigation. It was signed by James Webb, the NASA administrator, and Robert S. McNamara, the secretary of defense. Webb and McNamara recommended that the U.S. should set the objective of manned lunar exploration "before the end of this decade." They argued that "this nation needs to make a positive decision to pursue projects aimed at enhancing national prestige. Our attainments are a major element in the international competition between the Soviet system and our own." The two men cited lunar and planetary exploration as "part of the battle along the fluid front of the cold war."

Kennedy accepted these recommendations and presented them to a joint session of Congress on May 25. The president said, "I believe we should go to the moon.... No single space project in this period will be more exciting, or more impressive to mankind.... While we cannot guarantee that we shall one day be first, we can guarantee that any failure to make this effort will find us last." Kennedy vowed that Americans would set foot on the moon "before this decade is out."

The president's call to action struck a responsive chord in the U.S. populace. There was little public or political debate over the wisdom of the lunar commitment in the weeks following Kennedy's speech. Within months Congress increased NASA's budget by 89 percent; another 101 percent increase came the next year. Between 1961 and 1963 NASA's payroll swelled from 16,500 people to more than 28,000, and the number of contractors working on the space program grew from less than 60,000 to more than 200,000.

During the first year after Kennedy's announcement, a fierce technical debate erupted that threatened to delay progress in getting to the moon. The dispute centered on the most efficient strategy for sending people to the lunar surface and back to the earth. One possibility was to use several rockets to launch pieces of a lunar spacecraft separately into earth orbit, where they would be assembled and directed on to the moon. Jerome Weisner, the president's science adviser, and some elements within NASA initially inclined toward this "earth orbit rendezvous" plan.



JOHN M. LOGSDON and ALAIN DUPAS often work together in the analysis of the world's space programs. Logsdon is the director of the Space Policy Institute of George Washington University's Elliott School of International Affairs, where he is a professor of political science and international affairs. His research interests include space policy, the history of the U.S. space program and international science and technology policy. He is a member of the Aeronautics and Space Engineering Board of the National Research Council and sits on the board of directors of the National Space Society. Dupas is an international space policy strategist for CNES, the French space agency. He is also a partner at L.D. Associates, a strategic management consulting firm.

The Space Race between the U.S. and the U.S.S.R.

The competition between the U.S. and the Soviet Union in space grew out of the cold war conflict between the two nations. Early Soviet space achievements included the first satellite and the first human to orbit the earth. An aggressive, well-funded U.S. effort to place a human on the moon attempted to negate the propaganda value of these Soviet successes. By the mid-1960s the Soviets had initiated a secret, parallel program, setting the stage for a race to the moon.

1961–1962, UNITED STATES

Just four months after his inauguration, President John F. Kennedy vowed to land Americans on the moon by the end of the decade. That goal had been suggested by, among others, Wernher von Braun, the German-born rocket engineer. At the same time, the U.S. raced to catch up with the Soviets. Alan B. Shepard became the first American in space; nine months later John H. Glenn matched Gagarin's feat.

1957–1962, SOVIET UNION

The launch of *Sputnik 1,* the first artificial satellite, captured the world's attention. Subsequent flights lofted dogs into space, paving the way for humans to follow. On April 12, 1961, Yuri A. Gagarin circled the globe in the *Vostok 1* spacecraft, solidifying the Soviet lead in space. "Let the capitalist countries catch up with our country!" boasted Soviet premier Nikita S. Khrushchev.

Engineer readies Sputnik 1 *for flight (1957)*

Malyshka during preflight testing (1958)

McNamara was also intrigued by the potential military applications of earth-orbiting missions.

As they examined how best to meet Kennedy's goal of getting to the moon before the end of 1969, a growing number of engineers within NASA favored another approach, called lunar orbit rendezvous. In this scheme, the entire *Apollo* spacecraft would be sent into space in a single launch and would fly directly into orbit around the moon; a small landing craft would detach from the main spaceship and ferry the astronauts from lunar orbit to the moon's surface and then back to the mother ship, which would then return to earth.

Lunar orbit rendezvous dramatically lowered the overall weight of the *Apollo* spacecraft. Consequently, the Apollo mission could be carried out using a single Saturn V rocket. After fending off Weisner's objections, NASA officials approved lunar orbit rendezvous, realizing that it offered the greatest likelihood of reaching the lunar surface according to Kennedy's schedule. By the end of 1962 the U.S. was well on its way to the moon. Not so the Soviet Union.

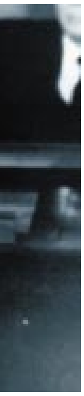
Until a few years ago, the Soviets officially claimed that the U.S. was the sole participant in the race to the moon. The very existence of the Soviet lunar program was a tightly held secret. As a result of *glasnost* and the collapse of the U.S.S.R., that situation has significantly changed. Several crucial players in the space program of the 1960s (most notably Vasily P. Mishin, who headed the Soviet human spaceflight effort from 1966 to 1974) have finally been allowed to place their recollections of the period in the public record. On August 18, 1989, the Soviet newspaper *Izvestia* printed a lengthy and unprecedentedly frank account of the nation's unsuccessful assault on the moon. And an increasing number of photographs and engineering descriptions of Soviet lunar hardware have become available to Western analysts and space observers. A recent study by Christian Lardier, a French space researcher, has been particularly valuable in bringing such information to light. The result is a much clearer picture of just how extensive the Soviet lunar program was.

In June 1961, at his first summit meeting with Soviet premier Nikita S. Khrushchev, Kennedy twice raised the possibility that the U.S. and the U.S.S.R. might travel to the moon together. Khrushchev was unresponsive, at least in part because Kennedy's lunar-landing announcement had caught the Soviet Union by surprise. The Soviet leadership was so confident in the country's space prowess that it had not anticipated that the U.S. might actually try to compete in that arena.

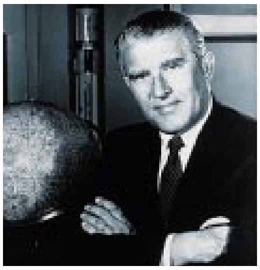
More than three years of political debate dragged on before the Kremlin decided, and then only tentatively, that the Soviet Union should also have a lunar-landing program. During that time, powerful and entrenched leaders of the Soviet design bureaus (industrial organizations in which the Soviet technical capabilities for space resided) struggled for priority and for resources related to possible lunar missions. Those conflicts presented a roadblock to establishing a single, coordinated plan of action for reaching the moon.

Sergei P. Korolev, the top space engineer, headed one of the design bureaus. He was, in many ways, the Russian equivalent of Wernher von Braun. Korolev had both designed the rocket used for all Soviet space launches to that point and had managed the programs responsible for developing most of the payloads lofted by those rockets. He was also an energetic and enthusiastic proponent of space travel. Such secrecy surrounded his work that Korolev was identified only as the "Chief Designer"; his name was not publicly revealed until after his death.

Unfortunately for the Soviet space effort, in the early 1960s Korolev became



John Kennedy announces lunar plans to Congress (May 25, 1961)



Wernher von Braun

Alan Shepard prepares for suborbital flight (May 5, 1961)

John Glenn enters the Mercury *capsule (February 20, 1962)*

Yuri Gagarin about to orbit the earth (April 12, 1961)

Gagarin (center) *celebrates his achievement with Nikita Khrushchev* (left) *and Leonid Brezhnev (May 1, 1961)*

embroiled in a personal and organizational conflict with Valentin P. Glushko, the head of the Gas Dynamics Laboratory and the primary designer of Soviet rocket engines. Disputes between the two dated to the 1930s, when Glushko was one of those whose testimony helped to send Korolev to a forced-labor camp. The two men clashed over the concept of the rocket engines for the next generation of Soviet space launchers. Korolev wanted to use high-energy liquid hydrogen as a fuel (the choice the U.S. made for the upper stages of Saturn V). Glushko was only interested in designing an engine fueled by storable but highly toxic hypergolic compounds, such as hydrazine and nitrogen tetraoxide, that ignited on contact.

The dispute grew so bitter that Glushko refused to work with Korolev in the creation of a new rocket. Instead Glushko allied his laboratory with another design bureau, headed by Vladimir N. Chelomei, to compete for the lunar assignment. Chelomei's group had developed military missiles but had no experience with rockets for outer space. On the other hand, one of Chelomei's deputies was Khrushchev's son, Sergei. That family link offered a great advantage in a system where such personal connections were often all-important. Chelomei had ambitions to expand his bureau's works into what had been Korolev's turf.

On major technical issues such as space exploration, the Soviet leadership relied on recommendations from the Soviet Academy of Sciences. Mstislav V. Keldysh, the president of the academy, was given the task of advising the government on the technical merits of competing proposals for future efforts in space. Keldysh and his associates took the path of least political resistance and did not fully support either Korolev or his competitors until after Khrushchev was removed from power.

From late 1961 on, Chelomei's design bureau devoted most of its attention not to landing on the moon but to sending cosmonauts on a flight around the moon without even going into lunar orbit. This mission was to use a UR-500 rocket (later known as Proton), derived from one of Chelomei's failed designs for an intercontinental ballistic missile (ICBM). Chelomei also promoted an overly ambitious plan for a reusable rocket airplane that could reach the moon and even the other planets.

In August 1964 the Chelomei design bureau received Kremlin approval to build both a spacecraft and the UR-500 rocket to send cosmonauts on a circumlunar mission by October 1967, the 50th anniversary of the Bolshevik Revolution. But Chelomei's apparent victory over Korolev was short-lived. The Politburo removed Khrushchev from power in October 1964.

The post-Khrushchev leadership quickly discovered that little progress had been made by the organization that had been receiving the lion's share of funding related to possible lunar missions. The Chelomei design bureau soon fell from favor, and its contract for the circumlunar program was canceled.

Korolev, meanwhile, had not been entirely shut out of the Soviet space program. After his successful efforts in using a converted ICBM to carry out the initial Soviet forays into space, he had been thinking about the design of a new heavy-lift space launcher, which he had designated N-1. In mid-1962 the Keldysh commission authorized the development of a version of the N-1 that could launch 75 tons into earth orbit, but the commission did not approve Korolev's plan to utilize the N-1 for a lunar mission structured around earth-orbit rendezvous.

The N-1 rocket was supposed to be ready for flight testing by 1965. Because he did not have access to the expertise of Glushko's Gas Dynamics Laboratory, Korolev had to find an alternative source of rocket engines. He turned to the design bureau led by Nikolai D. Kuznetsov, which had previously

1962–1967, UNITED STATES

After an intense dispute between Jerome Weisner, the presidential science adviser, and NASA managers, the agency in 1962 finalized its plan for the Apollo program to the moon. Under the guidance of NASA administrator James Webb, and with the strong backing of President Lyndon B. Johnson, the mission proceeded quickly. Meanwhile NASA continued to lag in feats such as a space walk, which the Soviets had accomplished three months earlier. NASA received a serious blow in 1967, when a cabin fire during a countdown rehearsal killed three *Apollo* astronauts.

James Webb* (left), *with Lyndon Johnson

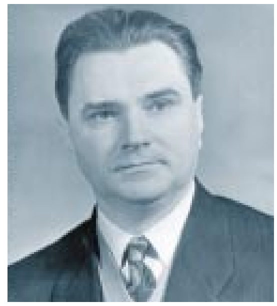
Jerome Weisner

1962–1967, SOVIET UNION

Personal conflicts hampered the Soviet lunar-landing program. Sergei P. Korolev conceived of a huge rocket, the N-1, that would transport cosmonauts to the moon. Korolev's plan was delayed by his clash with Valentin P. Glushko. After his death in 1966, Korolev was replaced by Vasily P. Mishin, who kept the beleaguered N-1 program alive. The Soviet space program also experienced technical setbacks, including a 1967 reentry mishap that killed the cosmonaut on the first flight of the new *Soyuz* spacecraft.

Sergei Korolev, "chief designer" of rockets* (right), *with Gagarin

Valentin Glushko, primary designer of Soviet rocket engines



worked on airplane engines. Kuznetsov's group had to begin its work on space propulsion systems basically from scratch. In the limited time available, Kuznetsov was able to develop only a conventionally fueled motor of rather little power. To achieve sufficient lifting power for a lunar mission, the N-1 ultimately needed 30 such engines in its first stage. (The American Saturn V had just five first-stage engines.)

After the fall of Khrushchev, the Soviet space program changed direction. Probably because it no longer feared angering Khrushchev, by December 1964 the Keldysh commission finally gave preliminary approval to a Korolev plan for placing cosmonauts on the moon. Korolev's revised lunar mission utilized a redesigned, more powerful N-1 rocket and the same lunar orbital rendezvous approach adopted for the Apollo mission. In May 1965 the Soviet government created the Ministry of General Machine Building to oversee the nation's space program; the ministry gave Korolev's lunar mission its highest priority. The official plan called for a first landing attempt in 1968, in the hope that the U.S.S.R. could still beat the U.S. to the moon.

Just as the Soviet effort was gaining momentum, disaster struck. In January 1966 Korolev died unexpectedly during simple surgery, robbing the Soviet space effort of its most effective and charismatic leader. Korolev's successor, Vasily Mishin, had neither Korolev's political standing nor his ability to lead. Continuing struggles with various government ministries and other design bureaus slowed progress. Chelomei continued to push an alternative lunar-landing scheme. To make matters worse, the revised N-1 launcher proved insufficiently powerful, so still more time was lost in another redesign.

Not until November 1966 did the Keldysh commission give a final go-ahead to the lunar-landing project. A joint government-party decree supporting the project was issued the following February, but still the Soviet government allocated only limited resources to it. By then the date for an initial lunar-landing attempt had slipped into the second half of 1969.

The U.S. was well aware of the Soviet decision to proceed with the N-1 but for several years remained unsure of the kind of mission for which it would be used. In 1964 U.S. intelligence satellites observed the construction of a launchpad for a large new booster and recorded the building of a second such pad in 1967. In a March 1967 national intelligence estimate (declassified in 1992), the Central Intelligence Agency suggested that "depending upon their view of the Apollo timetable, the Soviets may feel that there is some prospect of their getting to the moon first, and they may press their program in the hopes of being able to do so."

After 10 successful launches of the two-man *Gemini* spacecraft during 1965 and 1966, NASA seemed well prepared to move on to Apollo test flights leading to a lunar landing in 1968. Then, on January 27, 1967, the program received a tragic setback. An electrical fire broke out in the *Apollo 204* spacecraft (later renamed *Apollo 1*) during a countdown rehearsal on the launchpad. All three crewmen perished. Although critics lashed out at NASA, the agency never faltered. With limited congressional and White House intervention, NASA swiftly took the investigation into its own hands and identified and fixed the problems that had caused the fire. By the end of 1967 the space agency had set a new schedule for Apollo that called for an initial attempt at a landing by mid-1969, approximately the same target date as that of the Soviet program.

The U.S. and U.S.S.R. were also locked into a second contest: to see which country could first reach the vicinity of the moon. After the end of the Khrushchev era, the new

Edward S. White II takes the first American space walk (June 3, 1965)

Apollo 204 *cabin after the fatal fire (January 27, 1967)*

Vasily Mishin, Korolev's successor

UR-500 (Proton) launch

Soyuz *spacecraft*

Soviet leadership of Leonid I. Brezhnev and Alexei N. Kosygin asked Korolev to design a circumlunar mission similar to that of the now canceled Chelomei project. The Soviets still hoped to carry out such a flight in October 1967. After nearly a year of often acrimonious negotiations, Korolev and Chelomei in September 1965 agreed on a plan that would use the Chelomei UR-500 booster, supplemented by a Korolev upper stage being developed for the N-1 rocket and a two-cosmonaut version of the new *Soyuz* spacecraft being designed by the Korolev bureau.

Although the first few test flights of the UR-500 booster in 1966 were successful, there were a series of serious problems with subsequent launches. In addition, the first flight of the *Soyuz* spacecraft in April 1967 had a landing failure that killed the cosmonaut on board. Those setbacks made an October 1967 flight around the moon impossible. Even so, tests during 1967 and 1968 led to the successful Zond 5 mission of September 1968, in which the UR-500 launched a modified *Soyuz* spacecraft carrying living organisms, including several turtles, on a course that took it around the moon and then safely back to the earth. The flight of a Soviet cosmonaut around the moon seemed imminent.

At the time of the Zond 5 mission, the U.S. had no officially scheduled flight to the lunar vicinity until well into 1969. The reality was rather different, however. By mid-1968 development of the redesigned *Apollo* command-and-service module, which would carry astronauts into orbit around the moon and back to the earth, was on schedule for a first orbital test flight in October. But the separate lunar landing module, intended to place astronauts on the moon's surface, was months behind schedule. It seemed unlikely that the lunar module would be ready for an earth orbital test until February or March 1969.

George M. Low, deputy director of NASA's Manned Spacecraft Center in Houston, recognized that the delay in testing the lunar module presented a real possibility that the U.S. might not meet the end-of-the-decade deadline originally set by Kennedy. On August 9, 1968, Low therefore made a bold proposal: he suggested inserting an additional flight into the *Apollo* launch schedule, one in which a Saturn V would send the command-and-service module carrying a three-man crew into orbit around the moon.

Such a mission obviously carried substantial risks. It meant sending astronauts to the vicinity of the moon much earlier than had been planned, and it would be only the second flight of the *Apollo* spacecraft since its redesign after the 1967 fire. Moreover, the Saturn V had been launched only twice, and the second launch had uncovered several major problems. But Low's strategy would allow NASA to gain the experience of managing a mission at lunar distance many months earlier than had been planned. The additional flight would greatly increase the probability of meeting the Apollo schedule. It would also improve the likelihood that the U.S. would reach the vicinity of the moon before the U.S.S.R. did.

Low's plan gained rapid acceptance within NASA, encountering only temporary resistance from NASA administrator Webb and George Mueller, the head of NASA's Manned Spaceflight Program. In a little over a week the agency revised its entire Apollo schedule, creating a new mission just four months before it would lift off. The dramatic nature of that flight remained secret until after the October Apollo 7 mission, in which the command-and-service module performed flawlessly. On November 11, NASA's leaders formally sanctioned the Apollo 8 flight to the moon.

The Soviets, meanwhile, were struggling to keep up. In October 1968 a redesigned *Soyuz* spacecraft carrying one cosmonaut was successfully tested in

1967–1972, UNITED STATES

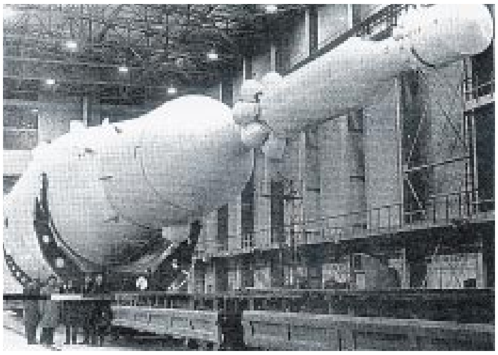
NASA recovered swiftly after the *Apollo* fire. But George M. Low, director of the Apollo program, worried about delays affecting the lunar lander. At his urging, NASA changed its launch schedule so that the first crew-carrying test flight of the Saturn V rocket (*Apollo 8*) went into orbit around the moon on December 24, 1968. Then on July 20, 1969, the *Apollo 11* lunar module made its historic touchdown on the surface, ending the race to the moon. Five more *Apollo* landings followed before the U.S. lunar program tapered off in 1972.

George Low

Earthrise over the moon, seen from Apollo 8 *(December 24, 1968)*

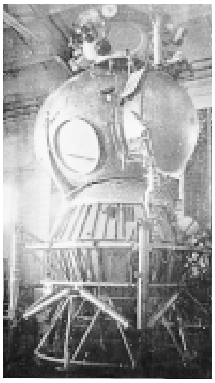
1967–1974, SOVIET UNION

The giant N-1 rocket never performed properly. On its second test launch, the N-1 exploded, wiping out its launch facilities. Glushko assumed control of N-1 development in 1974. He promptly canceled the program and dismantled the existing rockets. Pieces of the N-1 found ignominious duty as storage sheds. Many associated pieces of hardware, including a lunar lander and a semiflexible lunar space suit, were destroyed or placed into museums.



N-1 rocket being readied for testing

Lunar lander, designed to fit atop the N-1



earth orbit. The Zond 6 mission, which one month later sent a similar but unmanned spacecraft around the moon, did not fare so well. The spacecraft depressurized on reentry. If it had carried a crew, they would have died.

Nevertheless, the Soviets made preparations for launching a circumlunar Zond flight carrying two cosmonauts in early December. Both Mishin and the crew agreed to take the substantial risks involved, because by then they knew that the U.S. intended to send humans into orbit around the moon later that month. This launch presented the Soviets with perhaps their final opportunity to beat the Americans to the moon, but they did not take advantage of that chance. Just days before the scheduled takeoff, the Soviet leadership canceled the mission, presumably because they judged it too perilous.

During the final weeks of training for their mission, the *Apollo 8* crew members were well aware of when a Soviet circumlunar mission could be launched. In a conversation with one of us (Logsdon), Mission Commander Frank Borman recalls breathing a sigh of relief as the last possible date passed, and he realized that his own flight to the moon had not been preempted.

Apollo 8 entered lunar orbit on Christmas eve, 1968, all but ending the race to the moon. Furthermore, its accomplishments opened the way for the historic Apollo 11 mission seven months later, when Neil Armstrong planted the American flag in the lunar soil.

After the triumphs of Apollo 8 and Apollo 11, the Soviet lunar program faded into oblivion. But the Soviets did not give up on the moon immediately. Two more, unmanned Zond missions flew around the moon, one in 1969 and one in 1970. Shortly thereafter the Soviet leadership canceled the circumlunar program as it became clear that it had been totally overshadowed by Apollo.

The Soviet lunar-landing program suffered a more ironic fate. The first attempt to launch the N-1, in February 1969, failed one minute into flight. The second launch attempt on July 3, just 13 days before *Apollo 11* lifted off for the moon, ended in an explosion on the pad that destroyed much of the booster's ground facilities and halted the Soviet lunar-landing program for two years. N-1 launches in July 1971 and November 1972 also failed.

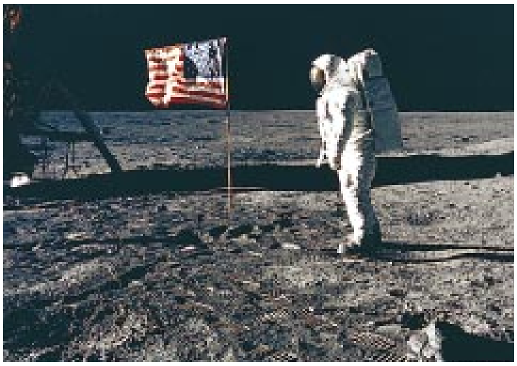
If they could not be first, the Korolev design bureau leaders reasoned, they could still be best. Led by Mishin, they reorganized the program around the concept of extended stays on the moon that would be longer than the brief visits made by the crews of the six Apollo missions. By early 1974 Mishin believed that he and his associates had identified the sources of earlier problems and were on the brink of success. But in May 1974, Mishin was replaced as head of the design bureau by Glushko, the man who more than a decade earlier had fought with Korolev over the choice of the N-1 propulsion system.

In one of his first acts, Glushko terminated the N-1 program and destroyed the 10 remaining N-1 boosters. Mishin argued that at least the two N-1s almost ready for launch should be tested, but to no avail. Rather than continue with the lunar program to which it had devoted substantial resources for more than a decade, Glushko and his superiors chose the almost pathological response of destroying most of the evidence of its existence. The Soviet human spaceflight program from the early 1970s on has concentrated entirely on long-duration flights in earth orbit.

Once astronauts had established an American presence on the moon, the U.S. lunar program also soon wound down. The sixth and last Apollo landing mission left the moon in December 1972. By then the lunar effort had clearly met the goals that Kennedy had set out in 1961.

Was the race to the moon worth win-

Apollo 11 *crew (May 1969)*



Neil Armstrong on the moon (July 20, 1969)

1975–PRESENT

In 1975 the U.S. and the U.S.S.R. conducted a rendezvous between a *Soyuz* and an *Apollo* spacecraft. That event set a precedent for the current plan to combine most U.S. and Russian human spaceflight activities, leading to an international space station by 2002. The station could open a new chapter in the collaborative exploration of space.

Astronaut and cosmonaut on board Apollo-Soyuz *(July 17, 1975)*

Soviet lunar space suit

N-1 pieces used as storage sheds

ning? In our judgment, that question can be answered only in light of the circumstances under which the competition occurred. The moon race was a cold war undertaking that should be evaluated primarily in foreign policy terms. On those grounds, it was an important victory. The Apollo program undoubtedly aided America's global quest for political and military leadership during the 1960s. The lunar landing constituted a persuasive demonstration of national will and technological capability for the U.S.

Likewise, the failure of the Soviet lunar program was more than a public relations defeat. In 1961, as the race to the moon began, many people in the U.S. (and around the world) thought Soviet centralized planning and management systems would allow the nation to pursue vigorously its long-range goals in space. The dissipation of the Soviet Union's lead in space during the 1960s tarnished the image of socialist competence and diminished Soviet standing in world affairs.

Throughout his brief presidency, Kennedy was ambivalent about the competitive aspects of the space race. In his inaugural address, he suggested to the Soviet Union that "we should explore the stars together." Shortly after being sworn in, he asked NASA and the state department to draw up proposals for enhanced U.S.-U.S.S.R. space cooperation. Those proposals arrived at the White House on the day of Gagarin's initial orbital flight, an event that convinced Kennedy that the U.S. had to assume leadership in space. Yet on September 20, 1963, in an address to the General Assembly of the United Nations, he still asked, "Why should man's first flight to the moon be a matter of national competition?"

Kennedy's dream of cooperation between the two space superpowers is at last on the verge of becoming a reality. On December 15 of last year Vice President Al Gore and NASA administrator Daniel S. Goldin signed agreements with their Russian counterparts for a series of joint space activities. That collaboration will culminate in an international space station, which will be built around U.S. and Russian capabilities but will include contributions from Europe, Japan and Canada. The station will begin operation soon after the turn of the century.

For 30 years, cold war rivalry was the lifeblood of both U.S. and Soviet programs of human spaceflight. If the adventure of space exploration is to continue into the 21st century, it will almost surely depend instead on widespread cooperation. The space station may serve as the harbinger of a new kind of foreign policy, one that brings the nations of the world together in the peaceful conquest of space.

FURTHER READING

THE DECISION TO GO TO THE MOON: PROJECT APOLLO AND THE NATIONAL INTEREST. J. Logsdon. MIT Press, 1970.

THE HEAVENS AND THE EARTH: A POLITICAL HISTORY OF THE SPACE AGE. Walter A. McDougall. Basic Books, 1985.

THE SOVIET MANNED SPACE PROGRAM. Phillip Clark. Orion Books, 1988.

APOLLO: THE RACE TO THE MOON. Charles A. Murray and Catherine Bly Cox. Simon & Schuster, 1989.

L'ASTRONAUTIQUE SOVIETIQUE. Christian Lardier. Armand Colin, Paris, 1992.

POURQUOI NOUS NE SOMMES PAS ALLÉS SUR LA LUNE? V.-P. Michine (with M. Pouliquen). Cépaduès Éditions, Toulouse, 1993.

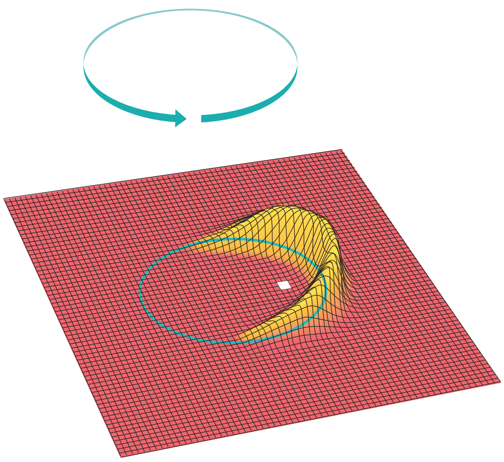
SPACEFLIGHT. Periodical published by British Interplanetary Society, 27–29 South Lambeth Road, London, England SW8 1SZ.

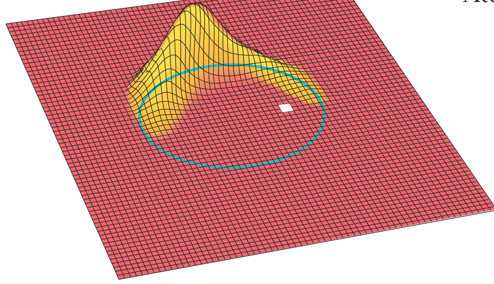
QUEST. A quarterly magazine on the history of spaceflight. P.O. Box 9331, Grand Rapids, MI 49509.

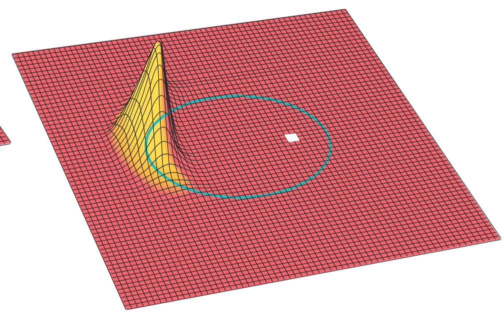
The Classical Limit of an Atom

By creating ultralarge atoms, physicists hope to study how the odd physics of the quantum world becomes the classical mechanics of everyday experience

by Michael Nauenberg, Carlos Stroud and John Yeazell







CLASSICAL ORBITAL MOTION can emerge from a quantum-mechanical object called a wave packet, which defines the probable location of an electron. The series of plots shows how the localized wave packet traces an elliptical orbit around the point where the nucleus resides (*white dots*). Note that the wave packet has begun to disperse after completing one revolution.

Throughout this century, physics has made use of two quite different descriptions of nature. The first is classical physics, which accounts for the motion of macroscopic objects, such as wheels and pulleys, planets and galaxies. It describes the continuous, usually predictable cause-and-effect relationships among colliding billiard balls or between the earth and orbiting satellites. The second description is quantum physics, which encompasses the microscopic world of atoms, molecules, nuclei and the fundamental particles. Here the behavior of particles is described by probabilistic laws that determine transitions between energy levels and govern tunneling through energy barriers. Because quantum mechanics is the fundamental theory of nature, it should also encompass classical physics. That is, applied to macroscopic phenomena, quantum mechanics should reach a limit at which it becomes equivalent to classical mechanics.

Yet until recently, the exact nature of this transition had not been fully elucidated. Now that goal is within reach. Atomic systems have been created that behave—for a short period—according to the laws of classical mechanics. Researchers fabricate such systems by exciting atoms so that they swell to about 10,000 times their original size. On such a scale the position of an electron can be localized fairly closely; at least its orbit no longer remains a hazy cloud that represents only a probable location. In fact, as the electron circles the nucleus, it traces an elliptical path, just as the planets orbit the sun.

The importance of understanding the classical limit of an atom takes on new meaning in the light of modern technology, which has blurred the distinctions between the macroscopic and microscopic worlds. The two domains had remained largely separate; a scientist would use classical mechanics to predict, say, the next lunar eclipse and then

MICHAEL NAUENBERG, CARLOS STROUD and JOHN YEAZELL combine theoretical and experimental expertise in exploring the classical limit of the atom. Nauenberg, who received his Ph.D. in physics from Cornell University, directs the Institute of Nonlinear Science at the University of California, Santa Cruz. Besides his focus on nonlinear physics, he also studies the history of Western science and mathematics during the 17th century. Stroud received his physics doctorate from Washington University. Currently a professor of optics and physics at the University of Rochester, Stroud divides his time formulating fundamental theories in quantum optics and then testing them in the laboratory. Yeazell received his Ph.D. in physics under Stroud's tutelage five years ago. As a fellow at the Max Planck Institute for Quantum Optics in Garching, Germany, he devotes his time to the study of quantum chaos—that is, quantum systems whose classical analogue acts chaotically. This fall he will join the faculty of Pennsylvania State University.

switch to quantum calculations to investigate radioactive decay. But engineers now routinely construct computer chips bearing transistors whose dimensions are smaller than one micron. Such devices are comparable in size to large molecules. At the same time, a new generation of microscopes can see and even manipulate single atoms. Finding the best way to exploit these technologies will be aided by the understanding we obtain from studies of the classical limit.

The profound differences between the quantum world and the classical domain emerged around the turn of the century. Experiments by such great scientists as Ernest Rutherford, the New Zealand–born physicist who worked at the University of Cambridge, established that the atom consists of a pointlike positive charge that holds negatively charged electrons. To early investigators, this arrangement mirrored the solar system. Indeed, the force that holds the electrons to the nucleus—called the Coulomb force—varies with the inverse square of the distance, as does the gravitational force.

This simple planetary model did not prove satisfactory. According to classical electromagnetic theory, any electric charge moving in a closed orbit must radiate energy. Thus, the electron in an elliptical orbit should quickly expend its energy and spiral into the nucleus. All matter would therefore be unstable. Furthermore, the radiation that an electron emitted as it plunged into the nucleus would have a continuous spectrum. But experiments indicated that electrons emit radiation in flashes, yielding a spectrum of discrete lines.

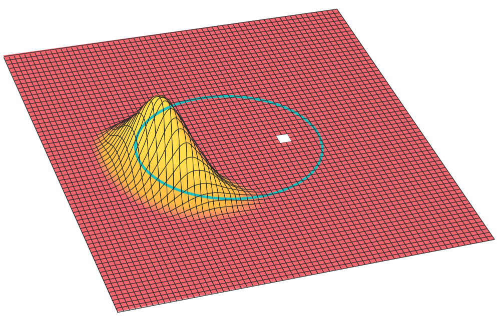
The Danish physicist Niels Bohr resolved some of these difficulties by augmenting the classical physics of the planetary model of the atom with a set of constraints. They were based on a theory about the nature of radiation first introduced by the German physicist Max Planck, who found that radiation is emitted in discrete units (the energy of which depends on the fundamental parameter now known as Planck's constant, h). Bohr retained the notion of classical orbits but assumed that only certain discrete values of energy and angular momentum were permitted. An integer, called a principal quantum number, characterized each energy state that an electron could occupy while associated with a nucleus. For example, the ground state was numbered one, the first excited state numbered two, and so on. Other quantum numbers describe a particle's angular momentum, which according to Bohr's theory would occur only in integer multiples of Planck's fundamental constant. Electrons could make transitions between orbits in the form of "quantum jumps." Each jump gave off a distinct frequency of light, which equaled the difference in energy between the two orbits divided by Planck's constant. The frequencies predicted in this way agreed completely with the observed discrete spectra of light emitted by hydrogen.

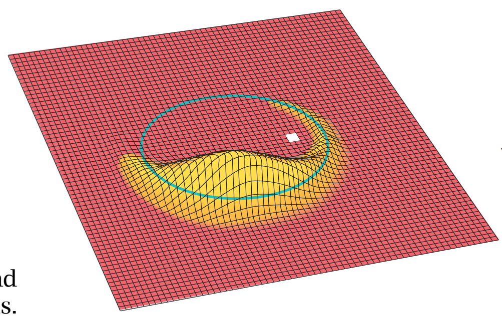
Bohr also postulated a rule that identified the classical limit of his quantum theory. This rule is named the correspondence principle. It states that for large quantum numbers, quantum theory should merge into classical mechanics. This limit corresponds to physical situations in which the classical action is much larger than Planck's constant. Therefore, it has become customary to refer to the classical limit as the scale at which Planck's constant vanishes. Bohr's correspondence principle has remained as a basic guideline for the classical limit of quantum mechanics, but as we shall see, this principle, while necessary, is not sufficient to obtain classical behavior.

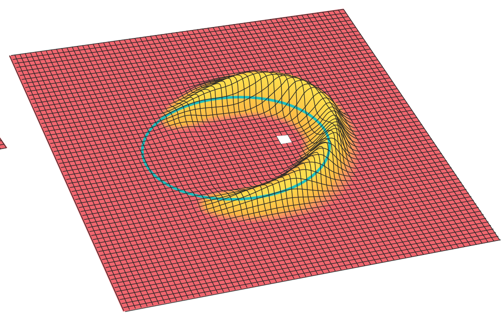
The Bohr-Rutherford model successfully explained the characteristics of hydrogen. But it produced difficulties and inconsistencies when applied to the behavior of more complicated atoms and to the properties of molecules. The German physicist Werner Heisenberg surmised that to make further progress, a quantum theory of atoms should be based only on directly observable quantities, such as the well-known spectral lines mentioned above. He believed certain concepts of classical physics, such as the electronic orbits that Rutherford and Bohr used, had to be completely discarded. He wrote to his Austrian colleague Wolfgang Pauli that these orbits do not have the slightest physical significance. Indeed, his matrix formulation of quantum mechanics did away with electron orbits entirely. Heisenberg accounted for the frequency and magnitude of the discrete spectral lines in terms of Planck's constant and other fundamental values in nature. Independently, the Austrian physicist Erwin Schrödinger derived an alternative but equivalent formulation. Following ideas of the French physicist Louis de Broglie, he represented physical systems with a wave equation. Solutions to this equation assigned probabilities to the possible outcomes of a system's evolution.

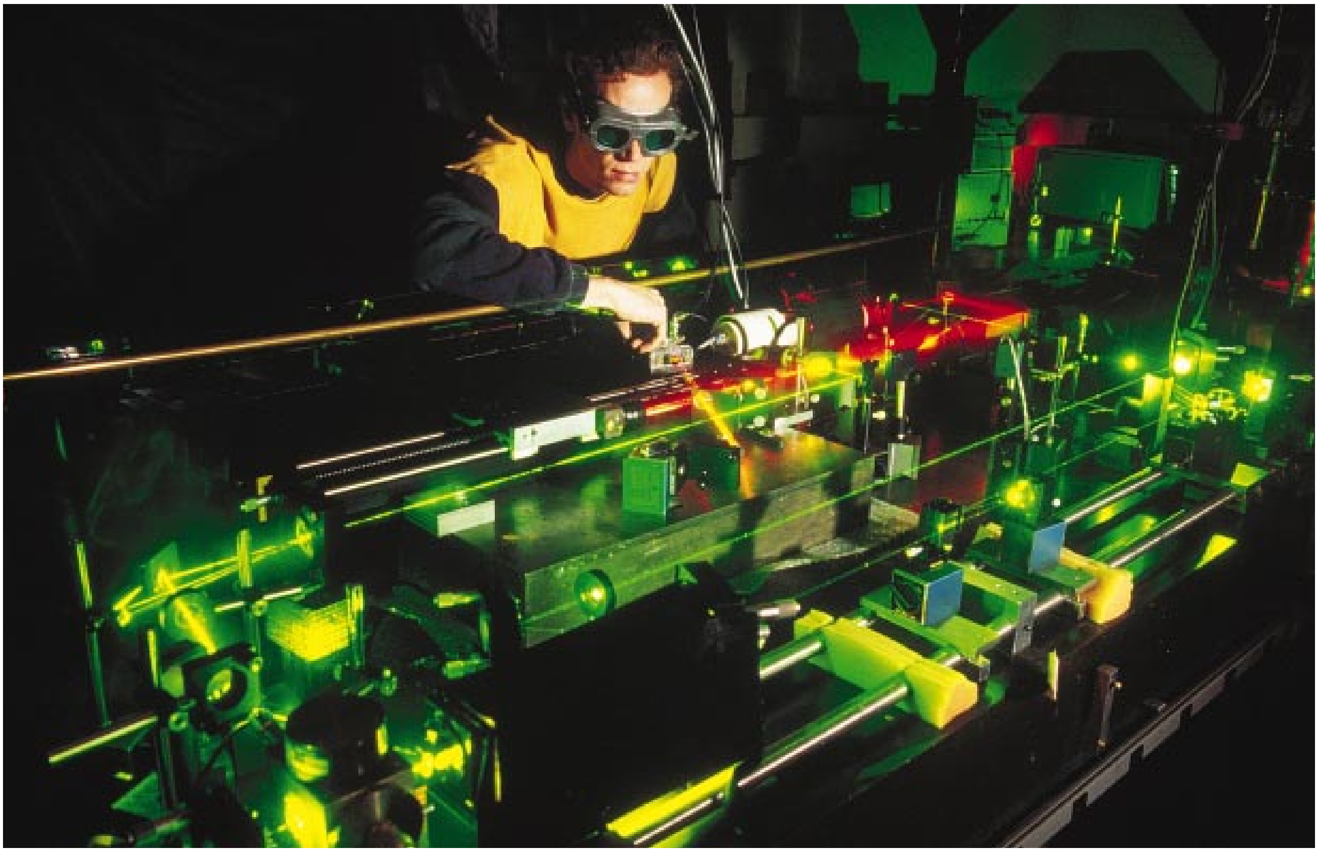
Whereas Heisenberg felt that classical orbits had no place in quantum theory and should be abandoned, Schrödinger was of a different mind. From the start he was concerned with the relation of the microscopic to the macroscopic world. Classical dynamics, he believed, should emerge from his wave equation. As a first step, Schrödinger investigated a very simple kind of system, called the harmonic oscillator. This system is not exactly that of an orbiting body; it corresponds to the up-and-down motion of a block hanging from the end of a spring. The harmonic oscillator shares a crucial feature of an orbit around a Coulomb or gravitational potential: periodicity. Such an orbiting body repeats its motion once each cycle—the period of the earth's orbit is just a year. The suspended block also has a cycle: it completes one up-down action over some unit of time.

Schrödinger managed to extract clas-









sical behavior from his theory for a harmonic oscillator. He did so by constructing a solution for his equation that was a sum of solutions that had discrete energy values. Graphically, these solutions resemble sinusoidal waves of different frequencies. Superposing such waves produced a "Gaussian wave packet," which looks like a bell-shaped curve. The remarkable property of this wave packet was that it remained localized around a center that executed classical, periodic behavior. Schrödinger, however, failed to derive similar classical motion for more complex cases, such as the movement of an electron in the hydrogen atom.

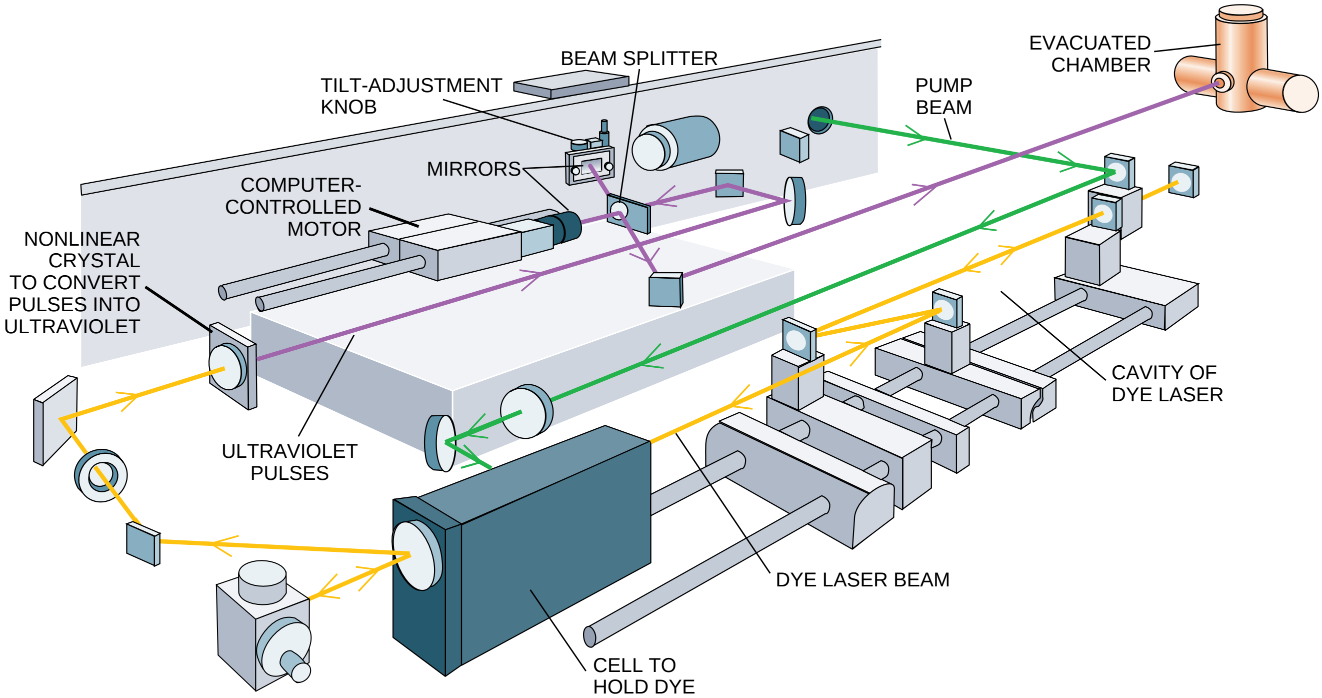
On the face of it, formulating a classical wave-packet description for an electron associated with an atom would not seem to be difficult. One would similarly choose appropriate atomic energy states, find their wave solutions and then superpose them. The problem lies in the way energy states are actually separated. A theorem developed by the French mathematician Jean-Baptiste Fourier indicates that only energy levels that are equally spaced with respect to one another can be combined to form a coherent state that moves periodically. But in an atom, the adjacent energy states are not equally spaced. For example, the energy separating the ground state from the first excited state is extremely large compared with the energy gaps at high quantum numbers: the first gap is one million times greater than that separating energy states whose quantum numbers are 100 and 101. A wave packet made up of a superposition of states near the ground state therefore disperses shortly after its creation. Obviously, a classical atom cannot be constructed from such states.

As Bohr noted, the key to achieving classical correspondence is to work with high-energy states, which have large quantum numbers. The energy separating these adjacent states is proportional to the inverse cube of the principal quantum number. That means, for large quantum numbers, the energy spacings between adjacent states are almost equal. In this limit, the spatial localization should persist for some time, permitting the center of the wave packet to evolve in a classical manner. Thus, the bigger the quantum numbers used, the easier it should be to produce a relatively stable, classical atom.

Until recently, no experimental device existed by which researchers could test the proposition by creating a superposition of excited atomic states in the laboratory. The development of lasers that can deliver short, powerful pulses of light proved to be the answer. By means of such devices, researchers formed the first localized wave packets in atoms during the late 1980s. Among the successful groups were ours at the University of Rochester, Ben van Linden van den Heuvell and his colleagues at the FOM Institute of Atomic and Molecular Physics in Amsterdam and Paul Ewart and his co-workers at the University of Oxford.

In a typical experiment, a brief, ultraviolet pulse of laser light lasting a mere 20 picoseconds (20 trillionths of a second) intersects a beam of potassium atoms in an evacuated chamber. Potassium is used because it readily absorbs the energy from our lasers, and, like hydrogen, it has one electron available for bonding. Each pulse excites an electron from a single ground state to many very high states. The result is a wave packet localized at a distance of about one micron from the nucleus.

Laser pulses of picosecond duration are essential because short bursts have a broad spectrum of frequencies. The spectral width of such a coherent pulse is proportional to the reciprocal of its duration, so that a pulse with a spectrum wide enough to overlap many levels must be extremely short. Traditional spectroscopy relies on long pulses, which contain a narrow band of frequencies and so excite only one or a few states. In our experiments, the average quantum number excited was 85, and about five states were superposed.



REACHING THE CLASSICAL LIMIT demands the excitation of atoms by brief pulses of laser light. A green laser beam emerges from behind the right side of the partition. It "pumps" a dye laser, which then produces yellow pulses (it appears faint green in the photograph on the opposite page). The nonlinear crystal converts the yellow light into ultraviolet (*invisible in photograph*). A beam splitter separates each ultraviolet pulse into two parts that move along different paths. A computer-controlled motor can alter the length of one path by shifting a mirror. Such adjustments allow one pulse to lag behind the other: a 0.3-millimeter increase produces a one-picosecond delay. The beams are recombined and directed at atoms in an evacuated chamber. The first pulse excites the atoms; the second pulse probes the result. The red and orange beams, used to maintain mirror alignments, and some components have been omitted from the diagram for clarity.

We probed the characteristics of our wave packet by measuring how it absorbs energy from a second laser pulse, fired shortly after the first one. At the perigee of its orbit, the wave packet absorbs the most energy. In fact, the amount of energy absorbed is sufficient to tear the electron away from the atom. Thus, to map out the electron's orbit, we simply measured the number of atoms that were ionized as we varied the delay between the two laser pulses. The ionization signals correspond to the expected oscillation of the wave packet as it periodically moves through the perigee of its orbit.

This method excites orbits of a fairly well defined energy and angular momentum. It does not select the orientation of the orbits. Instead the wave-packet state resides in the form of a statistical ensemble of classical orbits. Each member of the ensemble possesses the same radius and eccentricity but occupies all possible orientations in space. This superposition is well localized only in the radial dimension—that is, at a particular time, its distance from the nucleus is about as well determined as Heisenberg's uncertainty principle permits. Hence, investigators have christened this object a radial wave packet.

The motion of the radial wave packet contains many classical elements. The wave packet evolves from the nucleus toward the edge of a classical orbit and then returns. The period of this oscillation is just that of an electron following a classical elliptical orbit about the nucleus. Moreover, the wave packet moves most slowly at the apogee of its circuit and most quickly at the perigee, just as do comets and other orbiting bodies in their paths around the sun.

In forming a radial wave packet, we created a state that exhibits strong classical characteristics. Our goal, however, was to form a classical atom. In that regard, the radial wave packet has a shortcoming. Despite the classical orbital period of its oscillations, the packet follows a planetary trajectory only in a statistical sense. An electron in such a wave packet traces orbits that are oriented at all angles in space. In effect, the particles move about in a spherical shell wrapped around the nucleus [*see upper left illustration on next page*]. Obviously, this picture is not equivalent to that of a planetary system, in which the major axis of the ellipse describing the motion of a planet is (approximately) fixed in space. Furthermore, the wave packet spreads as it propagates radially, a behavior comparable in classical physics to a planet breaking up as it moves in its orbit.

Jean Claude Gay, Dominique Delande and Antoine Bommier of the École Normale Supérieure in Paris and one of us (Nauenberg) recently propounded a detailed theory that shows how to construct a wave packet that is oriented in a particular direction in space. We found that, for large quantum numbers, a stationary solution of Schrödinger's equation exists that amounts to an "elliptical stationary state." This state is unusual. A conventional atomic state has a discrete energy value and a range of angular momenta [see "Highly Excited Atoms," by Daniel Kleppner, Michael G. Littman and Myron L. Zimmerman; SCIENTIFIC AMERICAN, May 1981]. The elliptical stationary state, however, consists of a well-defined linear superposition of these ordinary atomic states that centers within a spread of angular momenta. The eccentricity of the corresponding elliptical orbit determines the spread. The square of the magnitude of the wave function gives the probability for finding the electron at a particular position. Graphically, this probability appears as a bump on the orbit, representing the maximum value of the wave function [*see upper right illustration on next page*].

Classical arguments explain the presence of the bump. The quantum-mechanical state is analogous to an ensemble of electrons traveling on classical orbits. Because their velocity is at a minimum at apogee, the electrons will tend to bunch up there. The bunching yields the bump on a graphical representation of the elliptical state. It represents the region in which the electron is most likely to be found. Making the elliptical stationary state in the laboratory is substantially more complicated than forging a radial wave packet. A short pulse of laser light that excites an atom is not enough. The set of states needed to form the elliptical state turns out to require a superposition of many angular momentum states rather than many energy states. The laser beam cannot directly excite such a superposition. An additional field must be applied simultaneously with the laser pulses. Several solutions have been proposed. Two of us (Stroud and Yeazell) have excited such a state by employing a strong radio-frequency field in conjunction with a short optical pulse.

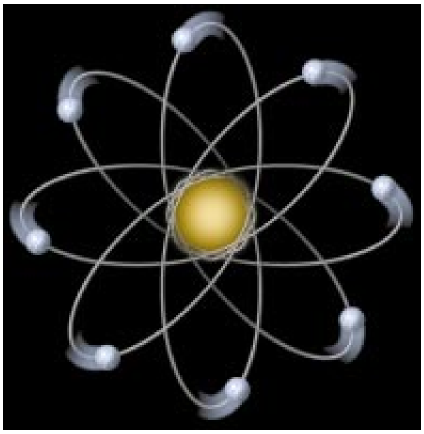
Although this elliptical state incorporates a definite angular orientation, it is stationary. It does not evolve in time. The final step in producing a classical state of the atom consists of making the wave packet move along the elliptical path [*see illustrations on pages 44 and 45*]. Although we have created such a wave packet as a solution of the Schrödinger equation on the computer, to date no one has succeeded in producing this state in the laboratory.

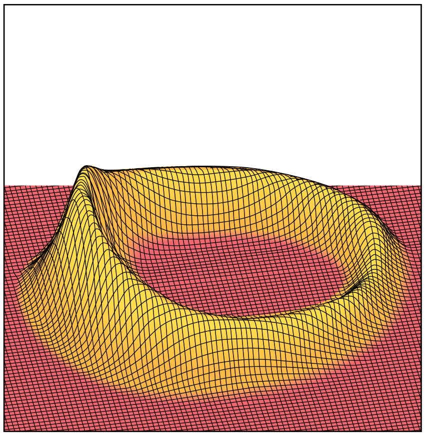
The theoretical wave packet we constructed is the most nearly classical state we know how to make. It shows striking classical properties but also maintains an underlying quantum-mechanical nature. As the wave packet moves around the elliptical path, it manifests one of its most obvious quantum properties. On each successive orbit, the wave packet spreads, a behavior akin to a classical group of electrons in which each particle moves at a different speed. Such a group would continue to spread indefinitely. But for the wave packet, a phenomenon quite distinct from classical behavior appears: quantum interference. This effect happens once the wave packet's head catches up to its tail and begins to interfere with it. Then, surprisingly, at a later, well-defined time, the wave packet reconstitutes itself, a behavior that does not have any classical analogue whatsoever. In between these full revivals, the state of the electron cannot be described as a single, spatially localized wave packet.

Indeed, windows in time exist in which the wave packet is localized in more complex structures. They constitute miniature replicas of the original wave packet that move classically as they maintain uniformly spaced positions on the orbit. These moments have been characterized as fractional, or partial, revivals. At a stage called the one-half revival, the wave packet has split into two smaller ones. Likewise, at the one-third revival, it has broken up into three packets, and so on. A classical particle by definition cannot spontaneously fracture and revive in this way, but a quantum particle can—and does.

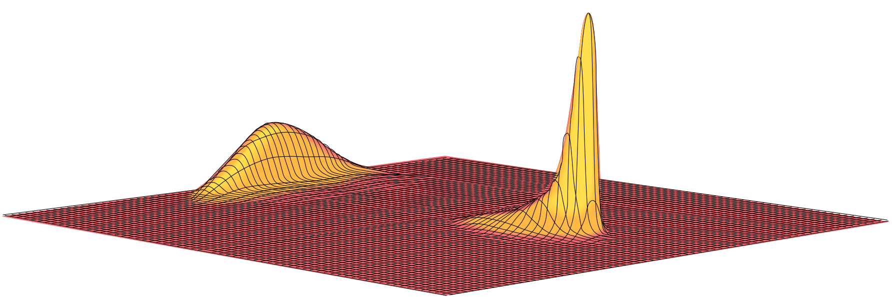
A classical analogy can explain many features of the quantum revivals. In particular, they can be likened to the bunching of runners on a racetrack. The runners represent the ensemble of electrons we use to imitate the quantum state. The racetrack contains a set of discrete classical orbits that satisfy Bohr's quantum conditions. At the beginning of the race, the runners line up at the start—that is, they are well localized. Each one runs in one of the quantized Bohr orbits. During the initial laps, the runners remain closely bunched. But after a few circuits, the runners have begun to spread around the track. It is not the quantum constraints or discreteness that causes this initial spreading. It is simply that the wave packet consists of a collection of waves of varying frequencies—a group of runners moving at different speeds.

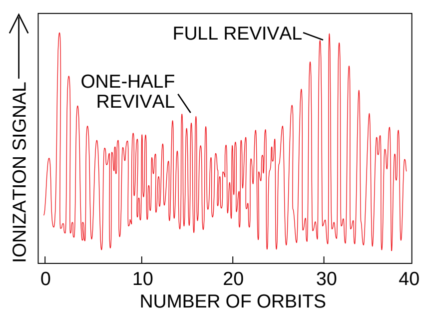
The quantum features begin to appear when the racers start to clump—that is, when the fastest runner catches up to the slowest runner. Further into the race, the quicker runners continue to pass the slower runners. Occasional-





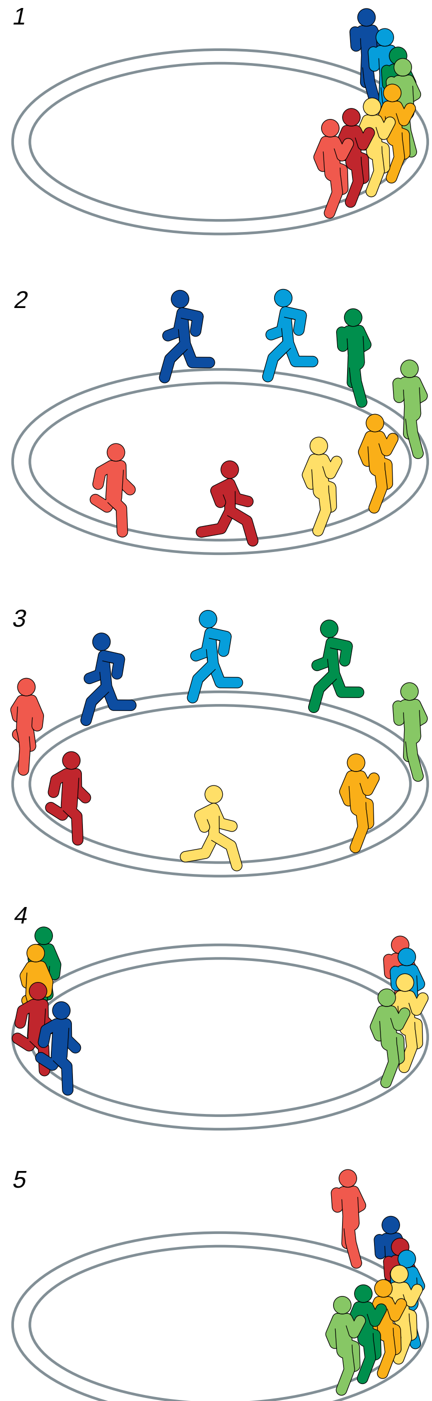
ENSEMBLE OF CLASSICAL ORBITS (*left*) is one way to describe a radial wave packet. The packet consists of a superposition of several energy levels; in effect, an electron moves simultaneously in many orbits that surround the nucleus. A more planetlike behavior would have the orbits lie in one plane. Such a state, called the elliptical stationary state, has been created (*right*). The bump on the left side represents the most likely location of the electron.





ONE-HALF REVIVAL of a wave packet (*left*)—that is, the formation of two smaller packets after the original has dispersed—takes place after about 15 orbits. It is indicated by ionization signals that appear twice as frequently (*right*). After about 30 orbits, the ionization signal returns to its original value, showing that the wave packet has fully revived.

RUNNERS ON A TRACK can portray the wave-packet revivals. At the start (*1*), the runners are bunched together, representing a well-localized wave packet. During the course of the race, the faster runners pull ahead (*2*); soon they begin to lap the slower competitors (*3*). Eventually two clumps of runners form (*4*), corresponding to a one-half revival. After many more circuits, they clump back into a single group (*5*). A problem with this model is that the full revival actually takes place on the side of the track opposite from the location of the clumped runners.



ly several runners may form a clump. Because of the particular distribution of speeds allowed by the quantum constraints, there is a moment when the runners form two clumps on opposite sides of the track. This clumping corresponds to the one-half fractional revival. Quantum constraints sort the runners into groups, so that one pack contains all the odd-numbered runners and the other all the even-numbered runners.

As the race continues, the runners spread out and eventually clump again, but into three groups. Finally, after many circuits, each runner has run a full lap farther than the next slower runner, so a full revival occurs. The number of such fractional revivals depends on the number of runners in the race. It requires at least two runners to form a clump. Similarly, in the atom the number of fractional revivals depends on the number of levels in the superposition. Neither the fractional nor full revivals would appear in this classical model without the imposition of the quantum constraints that place the runners into discrete orbits.

Investigations into this realm of physics have shown that despite Heisenberg's attempt to banish them, classical orbits remain a part of modern quantum mechanics. But their role is far more subtle than even Bohr realized. Wave packets that travel on classical trajectories are not produced by simply letting the quantum numbers of the system become large. Rather the formation of a special coherent superposition of states that have large quantum numbers is necessary for a wave packet to demonstrate two hallmark classical features: spatial localization and motion along an orbital path. These classical actions persist for only a limited period. For longer times, the underlying quantum dynamics manifests itself in previously unexpected wave phenomena that have no classical analogy.

Such results may best be understood in terms of theories that incorporate classical dynamics into quantum mechanics. Such semiclassical techniques are invaluable because conventional quantum-mechanical calculations are difficult and time-consuming, even on the largest supercomputers. Moreover, by themselves the resulting numerical solutions often cannot be understood or interpreted physically.

Although semiclassical methods have been used for a long time, especially in descriptions of a quantum system's energy, they have only recently been extended successfully to the time domain. They can now predict quantum behavior, even under nonlinear, or chaotic, circumstances. For example, Eric J. Heller of Harvard University and Steven Tomsovic of the University of Washington studied the motions of a wave packet trapped inside a "box." They showed that semiclassical methods describe the packet's chaotic motions as well as quantum calculations do. Such schemes also promise to illuminate other topics associated with quantum chaos that have received much attention lately. Among them are the microwave ionization of atoms and the behavior of atoms in strong electromagnetic fields.

Of course, short intense laser pulses can excite systems other than atoms. When a molecule is excited this way, its atoms can form wave packets. Presumably an appropriate tailoring of the laser pulse could control the internal dynamics of the molecule [see "The Birth of Molecules," by Ahmed H. Zewail; SCIENTIFIC AMERICAN, December 1990].

These techniques have also been used to form wave packets of electrons, or even positively charged holes, in semiconductor quantum wells. The coherent oscillations of the wave packets can then produce novel devices that cannot be made with more conventional means of excitation. Such devices would be bonuses that come packaged with the fundamental information we seek at the classical limit of quantum mechanics.

FURTHER READING

COHERENCE AND DECAY OF RYDBERG WAVE PACKETS. Jonathan Parker and C. R. Stroud, Jr., in *Physical Review Letters,* Vol. 56, No. 7, pages 716–719; February 17, 1986.

QUANTUM WAVE PACKETS ON KEPLER ELLIPTIC ORBITS. Michael Nauenberg in *Physical Review A,* Vol. 40, No. 2, pages 1133–1136; July 15, 1989.

LASER EXCITATION OF ELECTRONIC WAVE PACKETS IN RYDBERG ATOMS. G. Alber and P. Zoller in *Physics Reports,* Vol. 199, No. 5, pages 231–280; January 1991.

OBSERVATION OF FRACTIONAL REVIVALS IN THE EVOLUTION OF A RYDBERG ATOMIC WAVE PACKET. John A. Yeazell and C. R. Stroud, Jr., in *Physical Review A,* Vol. 43, No. 9, pages 5153–5156; May 1, 1991.

SEMICLASSICAL THEORY OF QUANTUM PROPAGATION: THE COULOMB POTENTIAL. I. M. Suárez Barnes et al. in *Physical Review Letters,* Vol. 71, No. 13, pages 1961–1964; September 27, 1993.

Emotion, Memory and the Brain

The neural routes underlying the formation of memories about primitive emotional experiences, such as fear, have been traced

by Joseph E. LeDoux

Despite millennia of preoccupation with every facet of human emotion, we are still far from explaining in a rigorous physiological sense this part of our mental experience. Neuroscientists have, in modern times, been especially concerned with the neural basis of cognitive processes such as perception and memory. They have for the most part ignored the brain's role in emotion.

Yet in recent years, interest in this mysterious mental terrain has surged. Catalyzed by breakthroughs in understanding the neural basis of cognition and by an increasingly sophisticated knowledge of the anatomical organization and physiology of the brain, investigators have begun to tackle the problem of emotion. One quite rewarding area of research has been the inquiry into the relation between memory and emotion. Much of this examination has involved studies of one particular emotion—fear—and the manner in which specific events or stimuli come, through individual learning experiences, to evoke this state. Scientists, myself included, have been able to determine the way in which the brain shapes how we form memories about this basic, but significant, emotional event. We call this process "emotional memory."

JOSEPH E. LEDOUX is interested in the neural foundation of memory and emotion. He studies the anatomy, physiology and behavioral organization of these aspects of mental functioning. LeDoux, who is a professor of neural science and psychology at New York University, is the recipient of two National Institute of Mental Health distinctions: a Merit Award and a Research Scientist Development Award. He has also received an Established Investigator Award from the American Heart Association.

By uncovering the neural pathways through which a situation causes a creature to learn about fear, we hope to elucidate the general mechanisms of this form of memory. Because many human mental disorders—including anxiety, phobia, post-traumatic stress syndrome and panic attack—involve malfunctions in the brain's ability to control fear, studies of the neural basis of this emotion may help us further understand and treat these disturbances.

Most of our knowledge about how the brain links memory and emotion has been gleaned through the study of so-called classical fear conditioning. In this process the subject, usually a rat, hears a noise or sees a flashing light that is paired with a brief, mild electric shock to its feet. After a few such experiences, the rat responds automatically to the sound or light even in the absence of the shock. Its reactions are typical to any threatening situation: the animal freezes, its blood pressure and heart rate increase, and it startles easily. In the language of such experiments, the noise or flash is a conditioned stimulus, the foot shock is an unconditioned stimulus and the rat's reaction is a conditioned response, which consists of readily measured behavioral and physiological changes.

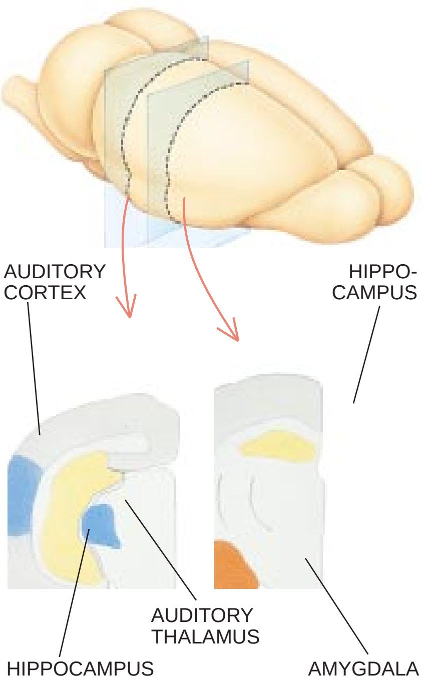
Conditioning of this kind happens quickly in rats—indeed, it takes place as rapidly as it does in humans. A single pairing of the shock to the sound or sight can bring on the conditioned effect. Once established, the fearful reaction is relatively permanent. If the noise or light is administered many times without an accompanying electric shock, the rat's response diminishes. This change is called extinction. But considerable evidence suggests that this behavioral alteration is the result of the brain's controlling the fear response rather than the elimination of the emotional memory. For example, an apparently extinguished fear response can recover spontaneously or can be reinstated by an irrelevant stressful experience. Similarly, stress can cause the reappearance of phobias in people who have been successfully treated. This resurrection demonstrates that the emotional memory underlying the phobia was rendered dormant rather than erased by treatment.

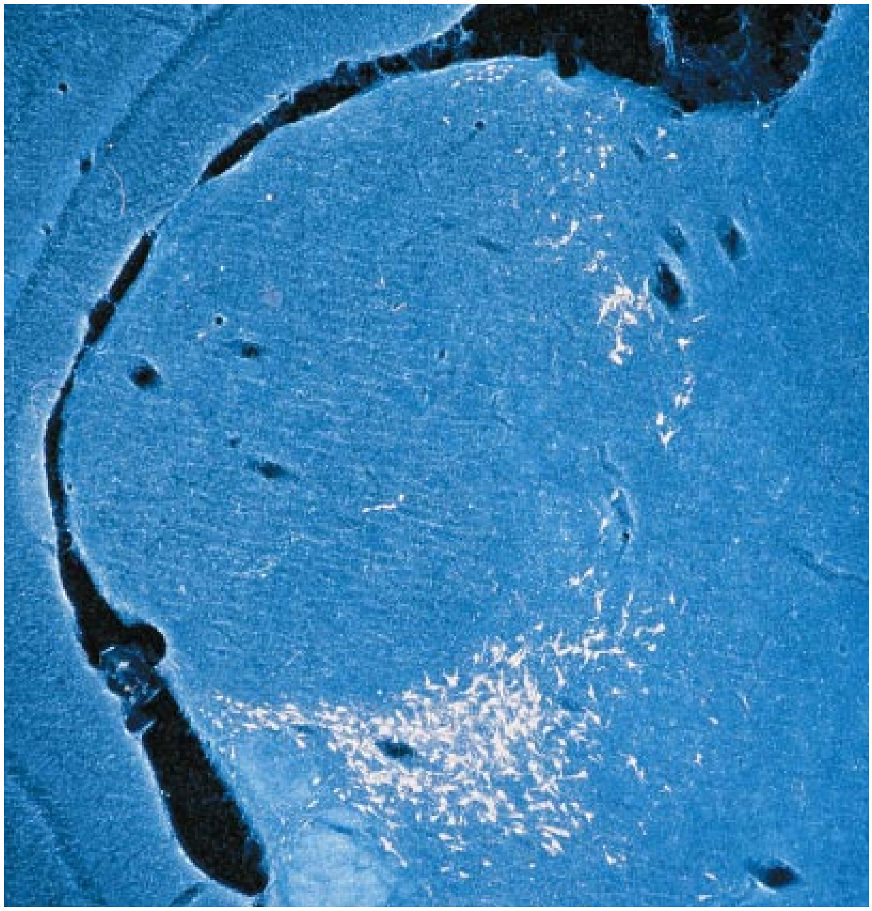
Fear conditioning has proved an ideal starting point for studies of emotional memory for several reasons. First, it occurs in nearly every animal group in which it has been examined: fruit flies, snails, birds, lizards, fish, rabbits, rats, monkeys and people. Although no one claims that the mechanisms are precisely the same in all these creatures, it seems clear from studies to date that the pathways are very similar in mammals and possibly in all vertebrates. We therefore are confident in believing that many of the findings in animals apply to humans. In addition, the kinds of stimuli most commonly used in this type of conditioning are not signals that rats—or humans, for that matter—encounter in their daily lives. The novelty and irrelevance of these lights and sounds help to ensure that the animals have not already developed strong emotional reactions to them. So researchers are clearly observing learning and memory at work. At the same time, such cues do not require complicated cognitive processing from the brain. Consequently, the stimuli permit us to study emotional mechanisms relatively directly. Finally, our extensive knowledge of the neural pathways involved in processing acoustic and visual information serves as an excellent starting point for examining the neurological foundations of fear elicited by such stimuli.

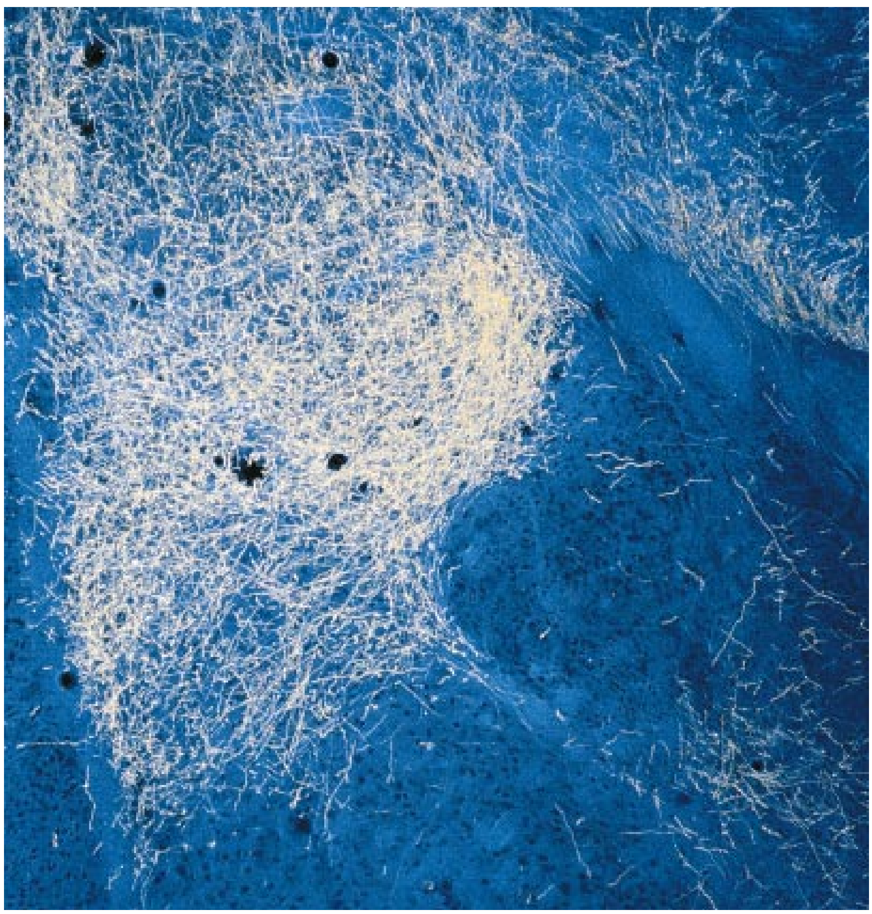
My work has focused on the cerebral

roots of learning fear, specifically fear that has been induced in the rat by associating sounds with foot shock. As do most other investigators in the field, I assume that fear conditioning occurs because the shock modifies the way in which neurons in certain important regions of the brain interpret the sound stimulus. These critical neurons are thought to be located in the neural pathway through which the sound elicits the conditioned response.

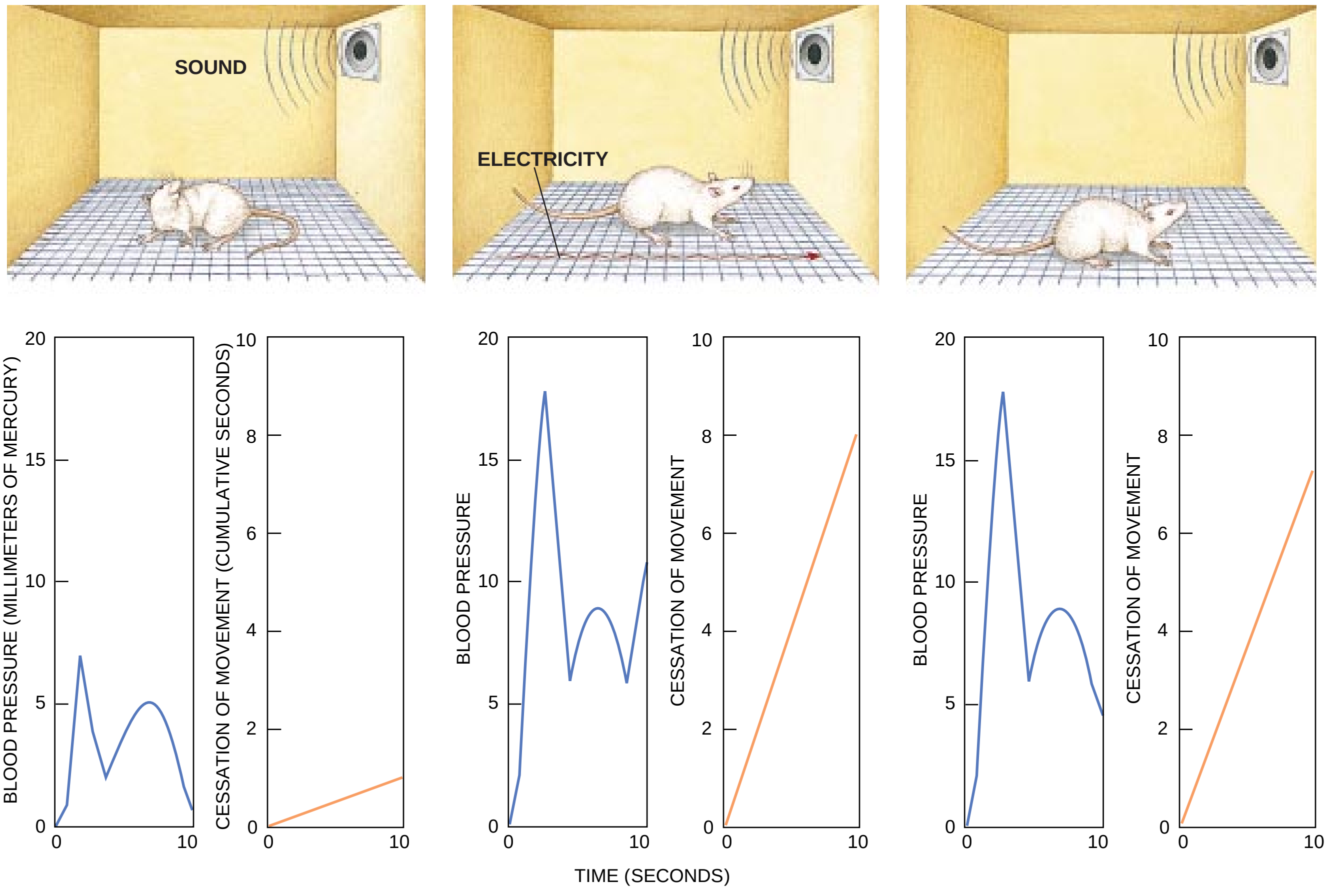
During the past 10 years, researchers in my laboratory, as well as in others, have identified major components of this system. Our study began when my colleagues at Cornell University Medical College, where I worked several years ago, and I asked a simple question: Is the auditory cortex required for auditory fear conditioning? In the auditory pathway, as in other sensory systems, the cortex is the highest level of







ANATOMY OF EMOTION includes several regions of the brain. Shown here in the rat (*above*), the amygdala, the thalamus and parts of the cortex interact to create memories about fearful experiences associated, in this case, with sound. Recent work has located precise areas where fear is learned and remembered: certain parts of the thalamus (*light pink at top right*) communicate with areas in the amygdala (*light yellow at bottom right*) that process the fear-causing sound stimuli. Because these neural mechanisms are thought to be similar in humans, the study of emotional memory in rodents may illuminate aspects of fear disorders in people.



CLASSICAL FEAR CONDITIONING can be brought about by pairing a sound and a mild electric shock to the foot of a rat. In one set of experiments, the rat hears a sound (*left*), which has little effect on the animal's blood pressure or patterns of movement. Next, the rat hears the same sound, coupled with a foot shock (*center*). After several such pairings, the rat's blood pressure rises at the same time that the animal holds still for an extended period when it hears the sound. The rat has been fear-conditioned (*right*): sound alone achieves the same physiological changes as did sound and shock together.

processing; it is the culmination of a sequence of neural steps that starts with the peripheral sensory receptors located, in this case, in the ear. If lesions in, or surgical removal of, parts of the auditory cortex interfered with fear conditioning, we could conclude that the region is indeed necessary for this activity. We could also deduce that the next step in the conditioning pathway would be an output from the auditory cortex. But our lesion experiments confirmed what a series of other studies had already suggested: the auditory cortex is not needed in order to learn many things about simple acoustic stimuli.

We then went on to make lesions in the auditory thalamus and the auditory midbrain, sites lying immediately below the auditory cortex. Both these areas process auditory signals: the midbrain provides the major input to the thalamus; the thalamus supplies the major input to the cortex. Lesions in both regions completely eliminated the rat's susceptibility to conditioning. This discovery suggested that a sound stimulus is transmitted through the auditory system to the level of the auditory thalamus but that it does not have to reach the cortex in order for fear conditioning to occur.

This possibility was somewhat puzzling. We knew that the primary nerve fibers that carry signals from the auditory thalamus extend to the auditory cortex. So David A. Ruggiero, Donald J. Reis and I looked again and found that, in fact, cells in some regions of the auditory thalamus also give rise to fibers that reach several subcortical locations. Could these neural projections be the connections through which the stimulus elicits the response we identify with fear? We tested this hypothesis by making lesions in each one of the subcortical regions with which these fibers connect. The damage had an effect in only one area: the amygdala.

That observation suddenly created a place for our findings in an already accepted picture of emotional processing. For a long time, the amygdala has been considered an important brain region in various forms of emotional behavior. In 1979 Bruce S. Kapp and his colleagues at the University of Vermont reported that lesions in the amygdala's central nucleus interfered with a rabbit's conditioned heart rate response once the animal had been given a shock paired with a sound. The central nucleus connects with areas in the brain stem involved in the control of heart rate, respiration and vasodilation. Kapp's work suggested that the central nucleus was a crucial part of the system through which autonomic conditioned responses are expressed.

In a similar vein, we found that lesions of this nucleus prevented a rat's blood pressure from rising and limited its ability to freeze in the presence of a fear-causing stimulus. We also demonstrated, in turn, that lesions of areas to which the central nucleus connects eliminated one or the other of the two responses. Michael Davis and his associates at Yale University determined that lesions of the central nucleus, as well as lesions of another brain stem area to which the central nucleus projects, diminished yet another conditioned response: the increased startle reaction that occurs when an animal is afraid.

The findings from various laboratories studying different species and measuring fear in different ways all implicated the central nucleus as a pivotal component of fear-conditioning circuitry. It provides connections to the various brain stem areas involved in the control of a spectrum of responses.

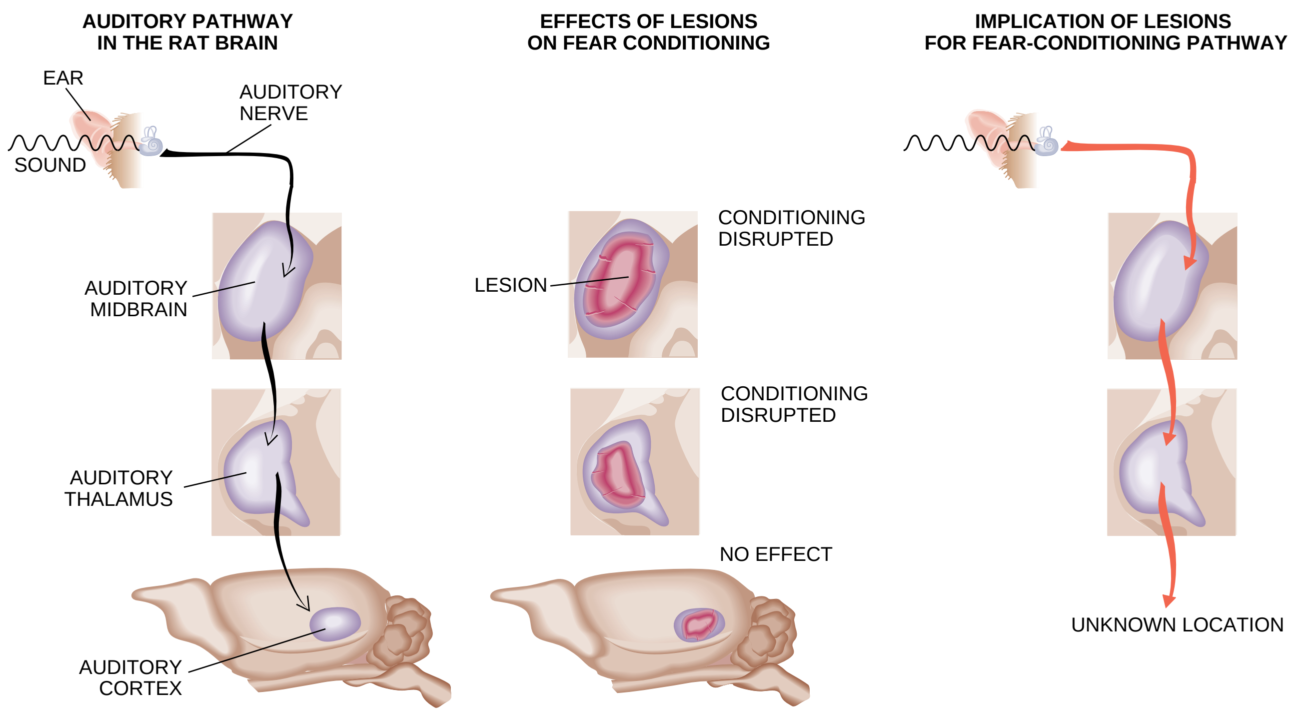
Despite our deeper understanding of this site in the amygdala, many details of the pathway remained hidden. Does sound, for example, reach the central nucleus directly from the auditory thalamus? We found that it does not. The central nucleus receives projections from thalamic areas next to, but not in, the auditory part of the thalamus. Indeed, an entirely different area of the amygdala, the lateral nucleus, receives inputs from the auditory thalamus. Lesions of the lateral nucleus prevented fear conditioning. Because this site gets information directly from the sensory system, we have come to think of it as the sensory interface of the amygdala in fear conditioning. In contrast, the central nucleus appears to be the interface with the systems that control responses.

These findings seemed to place us on the threshold of being able to map the entire stimulus response pathway. But we still did not know how information received by the lateral nucleus arrived at the central nucleus. Earlier studies had suggested that the lateral nucleus projects directly to the central nucleus, but the connections were fairly sparse. Working with monkeys, David Amaral and Asla Pitkanen of the Salk Institute for Biological Studies in San Diego demonstrated that the lateral nucleus extends directly to an adjacent site, called the basal or basolateral nucleus, which, in turn, projects to the central nucleus.

Collaborating with Lisa Stefanacci and other members of the Salk team, Claudia R. Farb and C. Genevieve Go in my laboratory at New York University found the same connections in the rat. We then showed that these connections form synaptic contacts—in other words, they communicate directly, neuron to neuron. Such contacts indicate that information reaching the lateral nucleus can influence the central nucleus via the basolateral nucleus. The lateral nucleus can also influence the central nucleus by way of the accessory basal or basomedial nucleus. Clearly, ample opportunities exist for the lateral nucleus to communicate with the central nucleus once a stimulus has been received.

The emotional significance of such a stimulus is determined not only by the sound itself but by the environment in which it occurs. Rats must therefore learn not only that a sound or visual cue is dangerous, but under what conditions it is so. Russell G. Phillips and I examined the response of rats to the chamber, or context, in which they had been conditioned. We found that lesions of the amygdala interfered with the animals' response to both the tone and the chamber. But lesions of the hippocampus—a region of the brain involved in declarative memory—interfered only with response to the chamber, not the tone. (Declarative memory involves explicit, consciously accessible information, as well as spatial memory.) At about the same time, Michael S. Fanselow and Jeansok J. Kim of the University of California at Los Angeles discovered that hippocampal lesions made after fear conditioning had taken place also prevented the expression of responses to the surroundings.

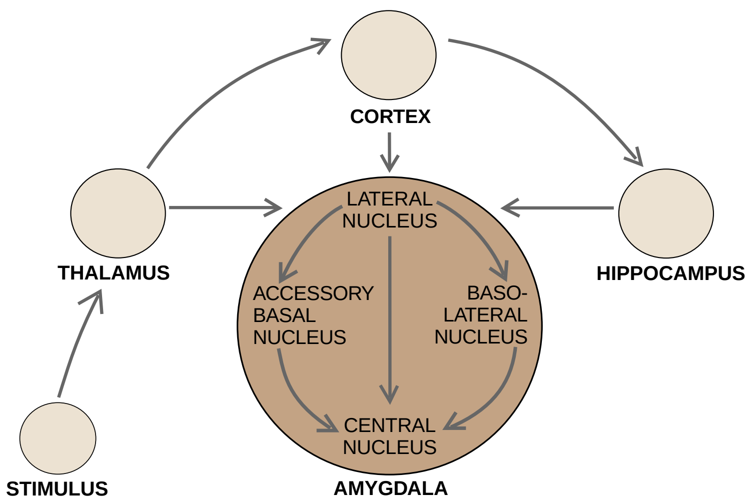
These findings were consistent with the generally accepted view that the hippocampus plays an important role in processing complex information, such as details about the spatial environment where activity is taking place. Phillips and I also demonstrated that the subiculum, a region of the hippocampus that projects to other areas of the brain, communicated with the lateral nucleus of the amygdala. This connection suggests that contextual information may acquire emotional signi-



BRAIN LESIONS have been crucial to pinpointing the sites involved in experiencing and learning about fear. When a sound is processed by the rat brain, it follows a pathway from ear to midbrain to thalamus to cortex (*left*). Lesions can be made in various sites in the auditory pathway to determine which areas are necessary for fear conditioning (*center*). Only damage to the cortex does not disrupt the fear response, which suggests that some other areas of the brain receive the output of the thalamus and are involved in establishing memories about experiences that stimulate fear (*right*).

Structure of the Amygdala

The amygdala's role in emotional behavior has long been considered important. Experiments in rodents have elucidated the structures of various regions of the amygdala and their role in learning about and remembering fear. The lateral nucleus receives inputs from sensory regions of the brain and transmits these signals to the basolateral, the accessory basal and the central nuclei. The central nucleus connects to the brain stem, bringing about physiological changes.



ficance in the same way that other events do—via transmission to the lateral nucleus.

Although our experiments had identified a subcortical sensory pathway that gave rise to fear conditioning, we did not dismiss the importance of the cortex. The interaction of subcortical and cortical mechanisms in emotion remains a hotly debated topic. Some researchers believe cognition is a vital precursor to emotional experience; others think that cognition—which is presumably a cortical function—is necessary to initiate emotion or that emotional processing is a type of cognitive processing. Still others question whether cognition is necessary for emotional processing.

It became apparent to us that the auditory cortex is involved in, though not crucial to, establishing the fear response, at least when simple auditory stimuli are applied. Norman M. Weinberger and his colleagues at the University of California at Irvine have performed elegant studies showing that neurons in the auditory cortex undergo specific physiological changes in their reaction to sounds as a result of conditioning. This finding indicates that the cortex is establishing its own record of the event.

Experiments by Lizabeth M. Romanski in my laboratory have determined that in the absence of the auditory cortex, rats can learn to respond fearfully to a single tone. If, however, projections from the thalamus to the amygdala are removed, projections from the thalamus to the cortex and then to the amygdala are sufficient. Romanski went on to establish that the lateral nucleus can receive input from both the thalamus and the cortex. Her anatomical work in the rat complements earlier research in primates.

Theodore W. Jarrell and other workers in Neil Schneiderman's laboratory at the University of Miami have shown that lesions in the auditory cortex disrupt fear conditioning to one of two stimuli that was paired with foot shock. Rabbits expressed fear responses only to the sound that had been coupled with the shock. After receiving auditory cortex lesions, however, the animals responded to both tones. When the auditory cortex was absent and animals had to rely solely on the thalamus and the amygdala for learning, the two stimuli were indistinguishable. This work suggests that the cortex is not needed to establish simple fear conditioning; instead it serves to interpret stimuli when they become more intricate. Schneiderman's findings are supported by research in primates showing that projections to the amygdala from sensory regions of the cortex are important in processing the emotional significance of complex stimuli.

Some of this work has been challenged by the intriguing studies of Davis and his team. They reported that damage to a region of the perirhinal cortex—a transitional region between the older and newer cortex—prevents the expression of a previously learned fear response. Davis argues, therefore, that the cortex is the preferred pathway to the amygdala and that thalamic projections are not normally used during learning, unless the cortex is damaged at the time of learning. Our general understanding of the effect of lesions administered after learning has taken place is that they interfere with long-term memory storage or retrieval. This interpretation seems applicable to Davis's work as well and is suggested by recent studies by Keith P. Corodimas in my laboratory. He showed that at least part of the deficit can be eliminated by providing reminder cues.

Once we had a clear understanding of the mechanism through which fear conditioning is learned, we attempted to find out how emotional memories are established and stored on a molecular level. Farb and I showed that the excitatory amino acid transmitter glutamate is present in the thalamic cells that reach the lateral nucleus. Together with Chiye J. Aoki, we showed that it is also present at synapses in the lateral nucleus. Because glutamate transmission is implicated in memory formation, we seemed to be on the right track.

Glutamate has been observed in a process called long-term potentiation, or LTP, that has emerged as a model for the creation of memories. This process, which is most frequently studied in the hippocampus, involves a change in the efficiency of synaptic transmission along a neural pathway—in other words, signals travel more readily along this pathway once LTP has taken place. The mechanism seems to involve glutamate transmission and a class of postsynaptic excitatory amino acid receptors known as NMDA receptors [see "The Biological Basis of Learning and Individuality," by Eric R. Kandel and Robert D. Hawkins; SCIENTIFIC AMERICAN, September 1992].

Various studies have found LTP in the fear-conditioning pathway. Marie-Christine Clugnet and I noted that LTP could be induced in the thalamo-amygdala pathway. Thomas H. Brown and Paul Chapman and their colleagues at Yale discovered LTP in a cortical projection to the amygdala. Other researchers, including Davis and Fanselow, have been able to block fear conditioning by blocking NMDA receptors in the amygdala. And Michael T. Rogan in my laboratory found that the processing of sounds by the thalamo-amygdala pathway is amplified after LTP has been induced. The fact that LTP can be demonstrated in a conditioning pathway offers new hope for understanding how LTP might relate to emotional memory.

In addition, recent studies by Fabio Bordi, also in my laboratory, have suggested hypotheses about what could be going on in the neurons of the lateral nucleus during learning. Bordi monitored the electrical state of individual neurons in this area when a rat was listening to the sound and receiving the shock. He and Romanski found that essentially every cell responding to the auditory stimuli also responded to the shock. The basic ingredient of conditioning is thus present in the lateral nucleus.

Bordi was able to divide the acoustically stimulated cells into two classes: habituating and consistently responsive. Habituating cells eventually stopped responding to the repeated sound, suggesting that they might serve to detect any sound that was unusual or different. They could permit the amygdala to ignore a stimulus once it became familiar. Sound and shock pairing at these cells might reduce habituation, thereby allowing the cells to respond to, rather than ignore, significant stimuli.

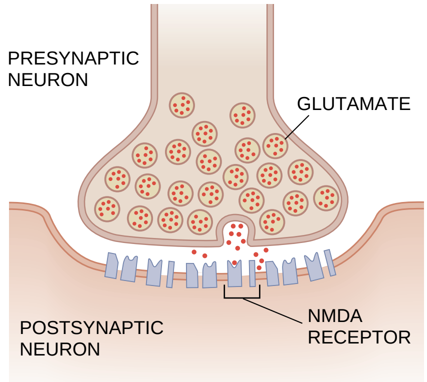
The consistently responsive cells had high-intensity thresholds: only loud sounds could activate them. That finding is interesting because of the role loudness plays in judging distance. Nearby sources of sound are presumably more dangerous than those that are far away. Sound coupled with shock might act on these cells to lower their threshold, increasing the cells' sensitivity to the same stimulus. Consistently responsive cells were also broadly tuned. The joining of a sound and a shock could make the cells responsive to a narrower range of frequencies, or it could shift the tuning toward the frequency of the stimulus. In fact, Weinberger has recently shown that cells in the auditory system do alter their tuning to approximate the conditioned stimulus. Bordi and I have detected this effect in lateral nucleus cells as well.

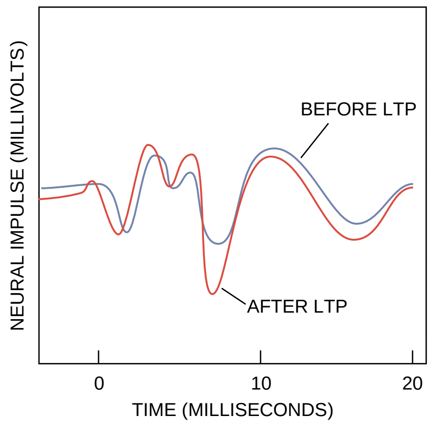
The apparent permanence of these memories raises an important clinical question: Can emotional learning be eliminated, and, if not, how can it be toned down? As noted earlier, it is actually quite difficult to get rid of emotional memories, and at best we can hope only to keep them under wraps. Studies by Maria A. Morgan in my laboratory have begun to illuminate how the brain regulates emotional expressions. Morgan has shown that when part of the prefrontal cortex is damaged, emotional memory is very hard to extinguish. This discovery indicates that the prefrontal areas—possibly by way of the amygdala—normally control expression of emotional memory and prevent emotional responses once they are no longer useful. A similar conclusion was proposed by Edmund T. Rolls and his colleagues at the University of Oxford during studies of primates. The researchers studied the electrical activity of neurons in the frontal cortex of the animals.

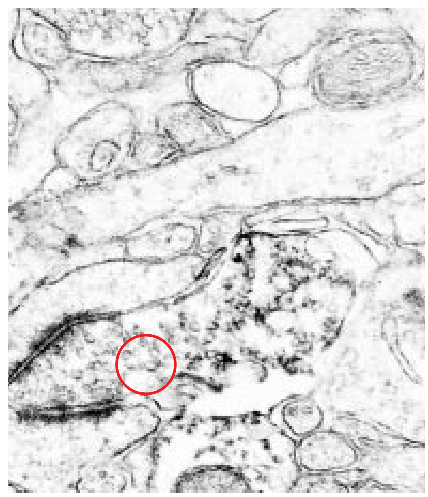
Functional variation in the pathway between this region of the cortex and the amygdala may make it more difficult for some people to change their emotional behavior. Davis and his colleagues have found that blocking NMDA receptors in the amygdala interferes with extinction. Those results hint that extinction is an active learning process. At the same time, such learning could be situated in connections between the prefrontal cortex and the amygdala. More experiments should disclose the answer.

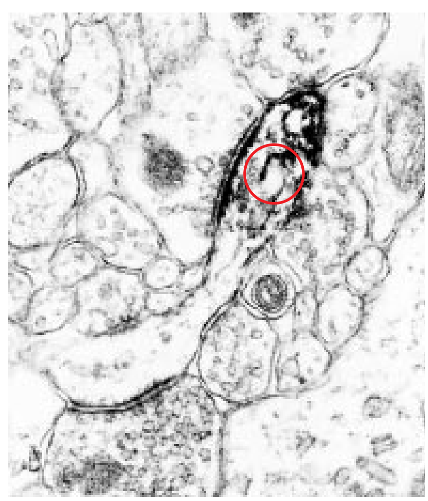
Placing a basic emotional memory process in the amygdalic pathway yields obvious benefits. The amygdala is a critical site of learning because of its central location between input and output stations. Each route that leads to the amygdala—sensory thalamus, sensory cortex and hippocampus—delivers unique information to the organ. Pathways originating in the sensory thalamus provide only a crude perception of the external world, but because they involve only one neural link, they are quite fast. In contrast, pathways from the cortex offer detailed and accurate representations, allowing us to recognize an object by sight or sound. But these pathways, which run from the thalamus to the sensory cortex to the amygdala, involve several neural links. And each link in the chain adds time.

Conserving time may be the reason there are two routes—one cortical and one subcortical—for emotional learning.









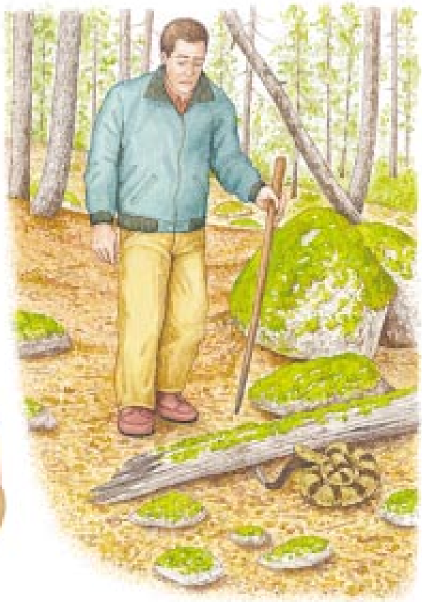
MEMORY FORMATION has been linked to the establishment of long-term potentiation, or LTP. In this model of memory the neurotransmitter glutamate and its receptors, called NMDA receptors (*top*), bring about strengthened neural transmission. Once LTP is established, the same neural signals produce larger responses (*top middle*). Emotional memories may also involve LTP in the amygdala. Glutamate (*red circle in top photograph*) and NMDA receptors (*red circle in bottom photograph*) have been found in the region of the amygdala where fear conditioning takes place.

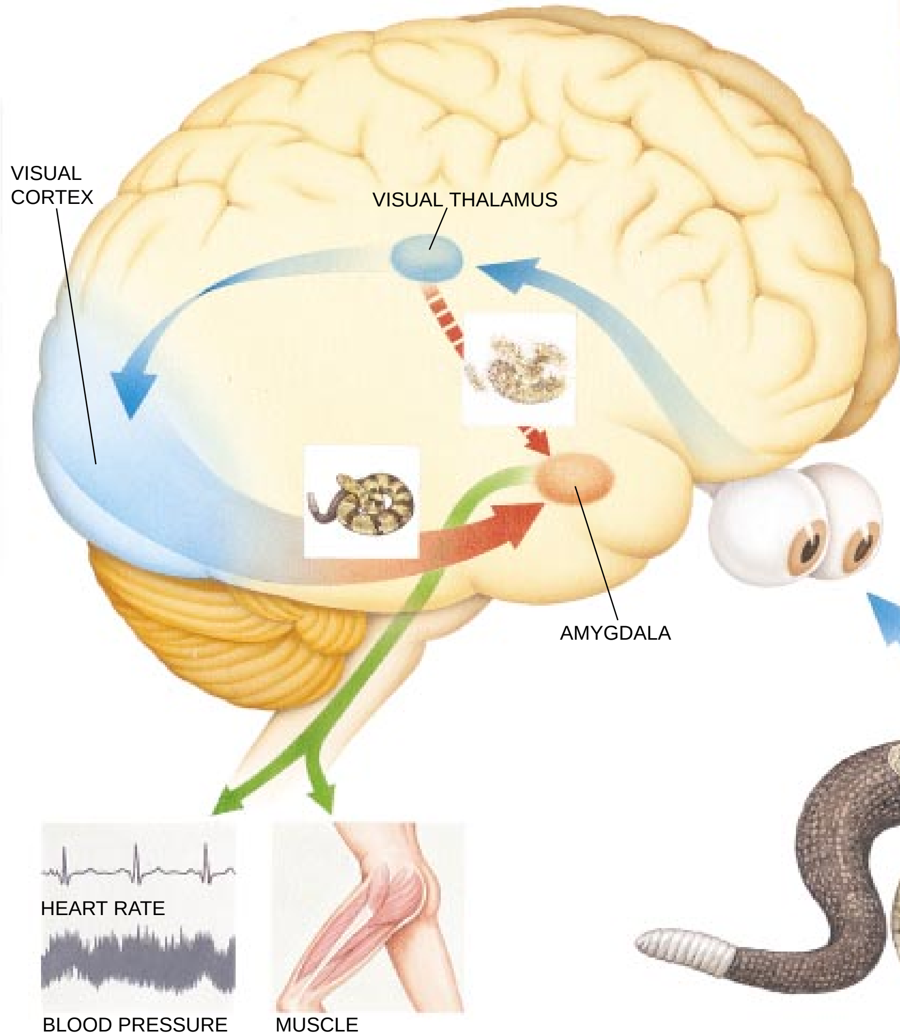
Animals, and humans, need a quick-and-dirty reaction mechanism. The thalamus activates the amygdala at about the same time as it activates the cortex. The arrangement may enable emotional responses to begin in the amygdala before we completely recognize what it is we are reacting to or what we are feeling.

The thalamic pathway may be particularly useful in situations requiring a rapid response. Failing to respond to danger is more costly than responding inappropriately to a benign stimulus. For instance, the sound of rustling leaves is enough to alert us when we are walking in the woods without our having first to identify what is causing the sound. Similarly, the sight of a slender curved shape lying flat on the path ahead of us is sufficient to elicit defensive fear responses. We do not need to go through a detailed analysis of whether or not what we are seeing is a snake. Nor do we need to think about the fact that snakes are reptiles and that their skins can be used to make belts and boots. All these details are irrelevant and, in fact, detrimental to an efficient, speedy and potentially lifesaving reaction. The brain simply needs to be able to store primitive cues and detect them. Later, coordination of this basic information with the cortex permits verification (yes, this is a snake) or brings the response (screaming, hyperventilating or sprinting) to a stop.

Although the amygdala stores primitive information, we should not consider it the only learning center. The establishment of memories is a function of the entire network, not just of one component. The amygdala is certainly crucial, but we must not lose sight of the fact that its functions exist only by virtue of the system to which it belongs.

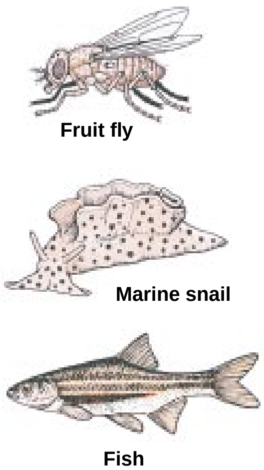
Memory is generally thought to be the process by which we bring back to mind some earlier conscious experience. The original learning and the remembering, in this case, are both conscious events. Workers have determined that declarative memory is mediated by the hippocampus and the cortex. But removal of the hippocampus has little

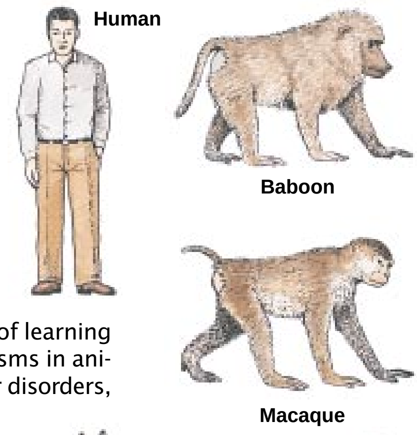




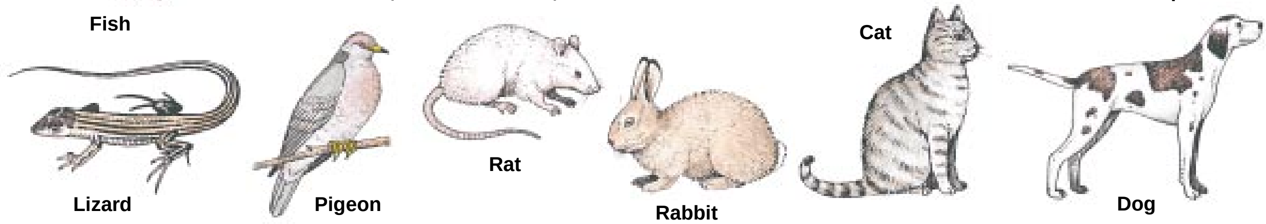
CORTICAL AND SUBCORTICAL PATHWAYS in the brain—generalized from our knowledge of the auditory system—may bring about a fearful response to a snake on a hiker's path. Visual stimuli are first processed by the thalamus, which passes rough, almost archetypal, information directly to the amygdala (*red*). This quick transmission allows the brain to start to respond to the possible danger (*green*). Meanwhile the visual cortex also receives information from the thalamus and, with more perceptual sophistication and more time, determines that there is a snake on the path (*blue*). This information is relayed to the amygdala, causing heart rate and blood pressure to increase and muscles to contract. If, however, the cortex had determined that the object was not a snake, the message to the amygdala would quell the fear response.

Some Species That Exhibit Fear Conditioning





Emotional memories brought about by fear-conditioning experiments have been observed in many animal groups. It appears that once a fearful memory has been established, it is relatively permanent: changes in behavior can be brought about by controlling the fearful response rather than by eliminating the emotional memory itself. This continuity between findings in diverse species suggests that brain pathways for this form of learning are similar. A fuller understanding of these mechanisms in animals may lead researchers to new treatments for fear disorders, such as panic attack or phobia, in humans.



effect on fear conditioning—except conditioning to context.

In contrast, emotional learning that comes about through fear conditioning is not declarative learning. Rather it is mediated by a different system, which in all likelihood operates independently of our conscious awareness. Emotional information may be stored within declarative memory, but it is kept there as a cold declarative fact. For example, if a person is injured in an automobile accident in which the horn gets stuck in the on position, he or she may later have a reaction when hearing the blare of car horns. The person may remember the details of the accident, such as where and when it occurred, who else was involved and how awful it was. These are declarative memories that are dependent on the hippocampus. The individual may also become tense, anxious and depressed, as the emotional memory is reactivated through the amygdalic system. The declarative system has stored the emotional content of the experience, but it has done so as a fact.

Emotional and declarative memories are stored and retrieved in parallel, and their activities are joined seamlessly in our conscious experience. That does not mean that we have direct conscious access to emotional memory; it means instead that we have access to the consequences—such as the way we behave, the way our bodies feel. These consequences combine with current declarative memory to form a new declarative memory. Emotion is not just unconscious memory: it exerts a powerful influence on declarative memory and other thought processes. As James L. McGaugh and his colleagues at the University of California at Irvine have convincingly shown, the amygdala plays an essential part in modulating the storage and strength of memories.

The distinction between declarative memory and emotional memory is an important one. W. J. Jacobs of the University of British Columbia and Lynn Nadel of the University of Arizona have argued that we are unable to remember traumatic events that take place early in life because the hippocampus has not yet matured to the point of forming consciously accessible memories. The emotional memory system, which may develop earlier, clearly forms and stores its unconscious memories of these events. And for this reason, the trauma may affect mental and behavioral functions in later life, albeit through processes that remain inaccessible to consciousness.

Because pairing a tone and a shock can bring about conditioned responses in animals throughout the phyla, it is clear that fear conditioning cannot be dependent on consciousness. Fruit flies and snails, for example, are not creatures known for their conscious mental processes. My way of interpreting this phenomenon is to consider fear a subjective state of awareness brought about when brain systems react to danger. Only if the organism possesses a sufficiently advanced neural mechanism does conscious fear accompany bodily response. This is not to say that only humans experience fear but, rather, that consciousness is a prerequisite to subjective emotional states.

Thus, emotions or feelings are conscious products of unconscious processes. It is crucial to remember that the subjective experiences we call feelings are not the primary business of the system that generates them. Emotional experiences are the result of triggering systems of behavioral adaptation that have been preserved by evolution. Subjective experience of any variety is challenging turf for scientists. We have, however, gone a long way toward understanding the neural system that underlies fear responses, and this same system may in fact give rise to subjective feelings of fear. If so, studies of the neural control of emotional responses may hold the key to understanding subjective emotion as well.

FURTHER READING

THE AMYGDALA: NEUROBIOLOGICAL ASPECTS OF EMOTION, MEMORY AND MENTAL DYSFUNCTION. Edited by John P. Aggleton. Wiley-Liss, 1992.

BRAIN MECHANISMS OF EMOTION AND EMOTIONAL LEARNING. J. E. LeDoux in *Current Opinion in Neurobiology*. Vol. 2, No. 2, pages 191–197; April 1992.

THE ROLE OF THE AMYGDALA IN FEAR AND ANXIETY. M. Davis in *Annual Review of Neuroscience*, Vol. 15, pages 353–375; 1992.

Adaptive Optics

Technology developed during the cold war is giving new capabilities to ground-based astronomical telescopes

by John W. Hardy

Atmospheric turbulence, which causes stars to twinkle and distant objects to shimmer, has frustrated astronomers ever since telescopes were invented. "The only Remedy is a most serene and quiet Air," wrote Sir Isaac Newton in 1704, "such as may perhaps be found on the tops of the highest Mountains above the grosser Clouds." Astronomers have followed this advice, which Newton offered in his *Opticks,* but even on the highest peaks atmospheric turbulence severely limits the power of big telescopes such as the 200-inch Hale telescope at Mount Palomar in California. The launch of the *Hubble Space Telescope* showed to what heights astronomers are willing to go to circumvent turbulence.

My colleagues and I at Litton Itek Optical Systems in Lexington, Mass., as well as workers at other institutions, have been pursuing another, earthbound, solution to the problem of atmospheric turbulence. Our approach, called adaptive optics, also has its roots in the development of space technology, but now, somewhat ironically, it is being applied to ground-based astronomical telescopes. Adaptive optics uses a deformable mirror or similar device to compensate, or correct, for the distortion of light caused by atmospheric turbulence. Adaptive optics technology is improving the ability of the next generation of earthbound telescopes to resolve fine detail and to detect extremely faint objects in the sky.

The challenge in building astronomical telescopes is to obtain the clearest possible image of a distant star, which should appear as a single point. Extended objects such as galaxies and planets can be regarded as collections of points. A distant star produces a spherical wavefront that travels vast distances through space until it reaches the earth's atmosphere, where air turbulence distorts it. Temperature variations associated with the turbulence generate changes in the air density, with the result that parts of the wavefront are slowed by different amounts, distorting the image. An adaptive optics telescope seeks to reverse this effect by restoring the spherical shape of the wavefront.

The first step is to determine how much each component of the wavefront is out of phase with the others. One way to that end is to divide the telescope's mirror into a number of zones and then measure the tilt of the wavefront in each zone. After processing by high-speed electronic circuits, this information is used to control actuators that determine the position of individual areas of the mirror's surface. The mirror is thereby deformed in such a way that any wave component arriving later than another actually travels a shorter distance to the focal point. This process of measurement and adjustment—a classic feedback setup—happens several hundred times a second. When the adaptive optics is working properly, all the components should arrive at the focal point in phase, to create a perfectly sharp image.

Radar engineers were the first to develop the notion of breaking a wavefront into parts and then bringing the parts into correct phase. The mathematical principles for compensating for distortion in a wavefront are virtually the same for optical images as for radar. In the early 1950s radar engineers began to divide antennas into segments so that the phase of the signal from each area could be independently adjusted. By phase shifting the wave components in this way, they were able to track moving objects with a fixed antenna or to focus the beam on objects at different distances.

The idea of applying adaptive principles to optical systems was first suggested in 1953 by Horace W. Babcock. He proposed that an electron beam be applied to control the thickness of a liquid film on a rigid mirror to compensate for errors in the phase of the incoming wavefront. The components of the wavefront that had phases preceding the others were delayed by passing them through a thicker film of liquid.

Babcock's ingenious concept would have required considerable development, and because the problem was of concern only to a fairly small community of astronomers, it was not pursued further. The simpler idea of stabilizing image motion with a tilting flat plate was used on a spectrograph of the Hale telescope in 1956. In an article in *Scientific American* in June 1956, Robert B. Leighton described the use of a tip-tilt mirror to obtain high-quality photographs of the planets.

Full correction of atmospheric turbulence, however, remained an unattained goal until the 1970s, when the U.S. military looked into the subject. Its interest stemmed from two sources. Pentagon scientists working on antiballistic missile defense needed a way to focus a laser beam on a distant

JOHN W. HARDY began working on adaptive optics in 1972 and, during the next two decades, developed the technology for applications in defense and astronomy. He was awarded a bachelor's degree in electrical engineering from London University in 1946 and has specialized in electro-optical technology. For his contributions to adaptive optics, Hardy won the Goddard Award of the Society of Photo-Optical Instrumentation Engineers in 1989 and the Michelson Medal of the Franklin Institute in 1992. He retired from Litton Itek Optical Systems in 1990 but is still active as a consultant and is now developing a low-cost adaptive optical device for small telescopes.

TELESCOPE equipped with adaptive optics is being tended by Robert Q. Fugate of the U.S. Air Force's Phillips Laboratory. Adaptive optical systems sharpen the images collected by ground-based telescopes by effectively erasing the blurring effects of the atmosphere.

target while protecting the ray from degradation in the atmosphere. The second objective was at the time even more urgent: the Soviet Union was launching great numbers of military satellites. The Defense Advanced Research Projects Agency (now ARPA) was searching for better methods of identifying those spacecraft. Photographs taken with ground-based satellite-tracking telescopes were too blurred by the atmosphere to yield useful images, even when digitally enhanced.

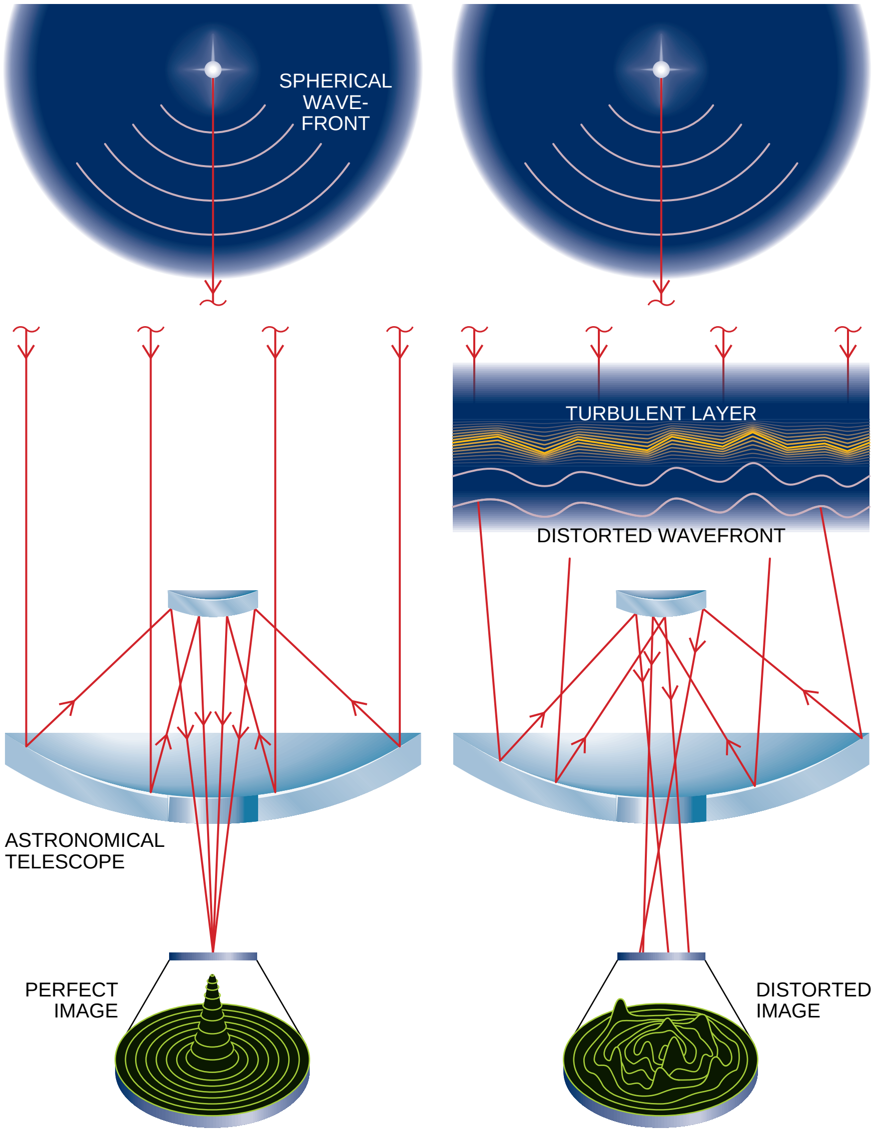
I was part of a team at Litton Itek Optical Systems in 1972 that won a contract with ARPA to develop a more effective approach. We decided to use adaptive optics to "undo" the distortion before the image was recorded—that is, to build a real-time atmospheric compensation system (RTAC).

Although the principle had been proved in radar applications, the components of an adaptive optical system had yet to be built. To create such a system, a key question had to be addressed: How finely must the incoming wavefront be divided to achieve a satisfactory reconstruction of the original image? The answer determines how many independently controlled actuators are required for the deformable mirror, which in turn determines the cost and complexity of the system. Fortunately, David L. Fried, then at North American Aviation, Inc., had provided a way to find the answer in 1966. Fried found that the optical effects of air turbulence, which at first appear complex and random, can be described in terms of simple wavefront shapes such as tilt, defocus and astigmatism (spherical and cylindrical curvature), which are familiar to all workers in optics. Furthermore, the strength of the turbulence can be represented as a single quantity, r_0. For conventional telescopes, r_0 is the diameter of the largest aperture that can be used before turbulence starts to degrade the image quality. As the turbulence gets stronger, r_0 becomes smaller. For earthbound observatories, it typically ranges between five and 15 centimeters at visible wavelengths, with an average value of 10 centimeters.

Most of the time, therefore, large telescopes cannot resolve objects such as double stars any better than can amateurs' small instruments. (Astronomers use large telescopes to collect enough light to enable them to record very faint objects. There are also periods when turbulence is quite low, enabling large telescopes to give good resolution.)

In adaptive optics, r_0 defines the size of each zone that must be adjusted to restore the image. To achieve good compensation at visible wavelengths, a four-meter telescope needs a deformable mirror controlled by about 500 actuators. The value of r_0 also depends on the wavelength of the incident light. In the infrared band, at two microns, an average value of r_0 is about 50 centimeters, so the number of actuators required by a four-meter telescope drops to about 60. We wanted to build a prototypical instrument equipped with a number of actuators sufficient to test the concept without the task's becoming too complex. So we settled, rather arbitrarily, on 21 actuators.

The only wavefront correctors available in 1972 were segmented mirrors that had been designed for remedying distortion in infrared laser beams. These devices were too slow and imprecise for our purposes. At first, a crystal of bismuth silicon oxide seemed a promising alternative. We found that we could adjust the phase shift of light passing through it by applying a voltage. But the crystal transmitted an in-



APPEARANCE OF STARS viewed from a great distance depends on the integrity of the spherical wavefronts of light they produce. If all the components of the wavefront can be focused, a star looks like a perfect point source of light (*left*). Atmospheric turbulence, however, randomly disrupts the wavefront's shape, which causes the components to arrive at a focal point out of phase (*right*).

sufficient amount of light, and its phase-correction capability was too small for atmospheric turbulence.

We considered using a flexible mirror made from a thin aluminized plate, which would reflect light efficiently and bend easily, but we struggled with the problem of stability. Although the surface of a deformable mirror moves less than 10 microns (one hundredth of a millimeter), it must be controlled with high accuracy—to a tolerance of as little as one fiftieth of a micron. My coworkers Julius Feinleib, Steven G. Lipson and Peter F. Cone, then at Itek, found that by mounting a thin glass mirror on a block of piezoelectric material fitted with electrodes, they could control deformations in hundreds of zones of the mirror to the required accuracy and speed. We called the device a monolithic piezoelectric mirror.

Next we addressed the problem of measuring distortion in the wavefront. At that time, the standard technique for precise measurement of optical wavefronts was a leisurely process in which photographs from a laser interferometer were manually scanned and digitized. With luck, information about the wavefronts was available the next day, a far cry from the one thousandth of a second response time needed for adaptive optics.

Fortunately, a new method of measuring wavefronts, called shearing interferometry, was under development. Interferometers are commonly used in optics to measure the phase of one wavefront by superimposing it on a second wavefront of known characteristics, thus producing an interference pattern. For adaptive optics, we need to know only the relative phase of each zone of the aperture with respect to its neighbors to determine the extent to which atmospheric turbulence has disturbed the wavefront's shape. Shearing interferometers accomplish this task by displacing ("shearing") two copies of the same wavefront by a known distance and then superimposing them. The intensity of the resulting interference pattern is proportional to the gradient, or slope, of the wavefront.

Conventional shearing interferometers, however, worked only with monochromatic light and produced only a fixed interference pattern. For adaptive optics, we needed to make rapid wavefront measurements using broadband white light from sunlit satellites. My colleague James Wyant was able to build a "white light" shearing interferometer, using a moving diffraction grating that produced an interference pattern in which the intensity varied sinusoidally. An array of photodetectors picked up the signal. The phase shift of this electrical signal when compared with a fixed reference was exactly proportional to the optical wavefront slope in the corresponding area of the aperture. This type of shearing interferometer is optically stable and reliable and needs little calibration. Later improvements increased the speed of this device so that it could measure 10,000 complete optical wavefronts per second, a speed sufficient to measure the worst atmospheric turbulence.

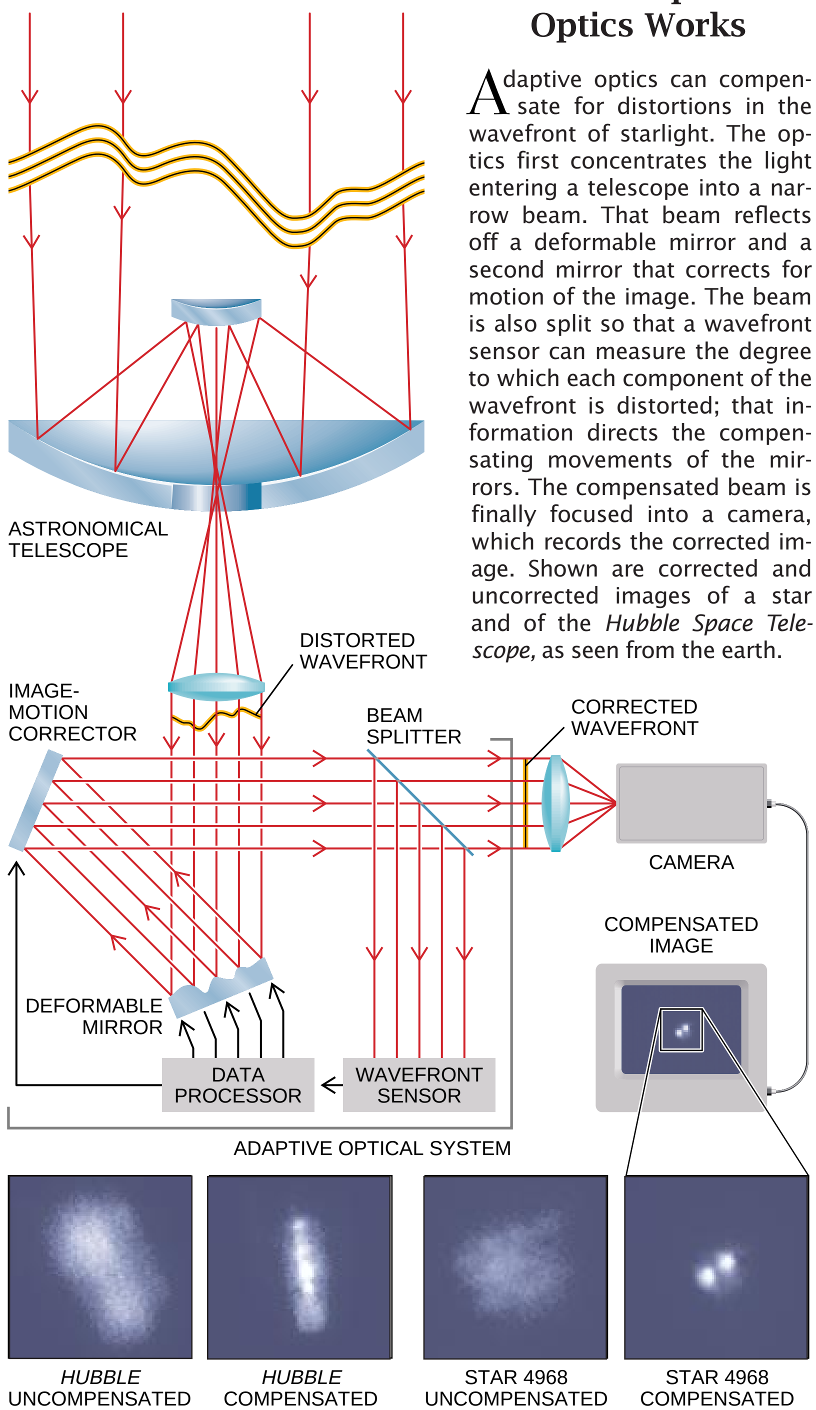
We needed one more element to complete the system: a fast method for synthesizing the individual wavefront measurements from each zone into a map of the continuous wavefront across the entire optical aperture. This wavefront reconstruction process is essential for determining the adjustment of the individual actuators. Serial computation, the most obvious method available to us, was problematic, given the small digital computers of that era, so we reverted to analog technology. Our group built a simple electrical network arranged in the same pattern as the actuators behind the deformable mirror. Electric currents representing the measured wavefront values were applied to the nodes in the network, which generated the exact voltages needed to adjust the corresponding actuators. This parallel network was extremely fast and could be expanded to manage a large number of actuators without losing speed—a vast advantage over techniques that utilized serial computation.

As the December 1973 test date for our real-time atmospheric compensation system approached, we grew increasingly concerned about whether the machine would operate stably. Each of the 21 actuators had its own feedback loop, but there was unavoidable cross-coupling between loops through the deformable mirror. In other words, a wavefront correction in one zone had a small effect on all the others. Our calculations showed that the new system should have been stable, but there was always the possibility of some unforeseen problem. We were concerned that if the RTAC burst into oscillation, as can happen with complex feedback systems, it might destroy the monolithic piezoelectric mirror that we had so painstakingly designed. When it was first tested, we were relieved to find that the system functioned perfectly and was as stable as a rock.

The RTAC demonstrated for the first time that adaptive optics could compensate extended images degraded by turbulence, and its basic architecture has been used in many subsequent systems. Still, it had too few actuators to be used on a large telescope. So in 1976, again with ARPA sponsorship, we started to build a much larger machine, called the compensated image system (CIS), which employed 168 actuators. My colleagues J. Kent Bowker, Richard A. Hutchin and Edward P. Wallner played important roles in the design of this pioneering system. We installed it in 1980 on the 1.6-meter telescope then operated by ARPA at Mount Haleakala in Maui, Hawaii. Once again we worried about stability: some members of the team predicted that large sections of the mirror would lock and render the entire system useless. When it was first tested in the spring of 1982 on bright stars, CIS proved to be perfectly stable. Since then, even larger adaptive optical systems have been built and operated without stability problems.

The CIS gave us the first real proof of how impressively adaptive optics can enhance the performance of ground-based telescopes. The results, particularly for double stars, were striking. The brighter star, with which the wavefront sensor measures turbulence, was easily visible before correction, but its companion was just a fuzzy, dim patch. Both stars exhibited a shimmering motion over several arc seconds. When the compensation loop was switched on, each star sprang into sharp focus and remained steady on the viewing monitor. The increase in brightness of the images was even more impressive than the improvement in sharpness.

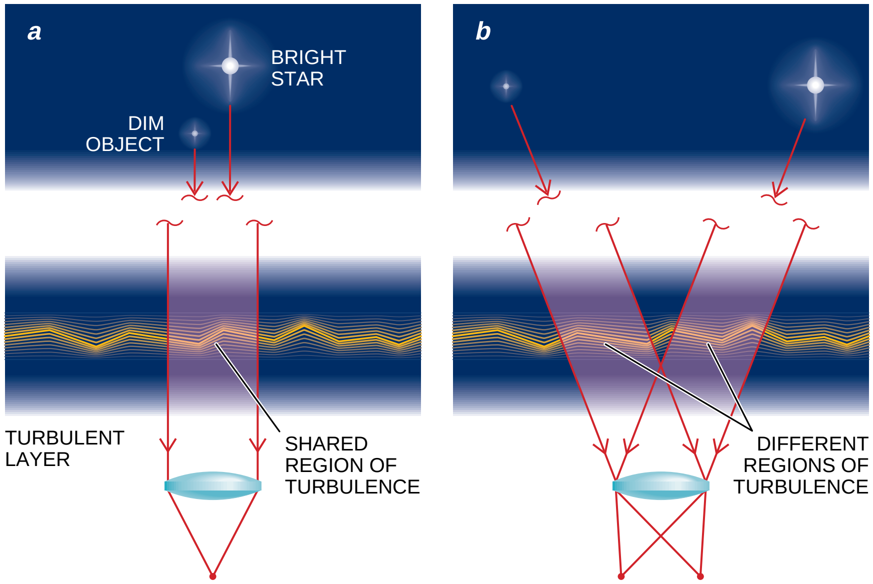
Since these early efforts, new types of deformable mirrors with more than 1,000 actuators have been made. Some of these are segmented mirrors, which consist of many flat plates, each mounted on three multiple-layer piezoelectric actuators. Segmented mirrors have the largest capacity to compensate for severe turbulence. Each segment is physically separate from the others, giving great freedom of movement. Unfortunately, the individual facets require frequent calibration. Because of discontinuities from one segment to the next, segmented mirrors tend to diffract some of the light, which affects an image's clarity. Consequently, astronomers generally prefer continuous faceplate mirrors. Such a mirror is made of a flexible aluminized glass faceplate mounted on actuators made of multiple layers of piezoelectric or electro-



How Adaptive Optics Works

Adaptive optics can compensate for distortions in the wavefront of starlight. The optics first concentrates the light entering a telescope into a narrow beam. That beam reflects off a deformable mirror and a second mirror that corrects for motion of the image. The beam is also split so that a wavefront sensor can measure the degree to which each component of the wavefront is distorted; that information directs the compensating movements of the mirrors. The compensated beam is finally focused into a camera, which records the corrected image. Shown are corrected and uncorrected images of a star and of the *Hubble Space Telescope,* as seen from the earth.

TO VIEW DIM OBJECTS, astronomers use brighter stars to gauge atmospheric turbulence (*a*). This technique, however, works only if a bright star is in the same part of the sky as the object viewed; if they are far apart, their light experiences different degrees of turbulence (*b*). Because few stars are bright enough to serve as references, this technique is useful only in a small part of the sky. One solution is to employ a laser beam directed through the atmosphere as an artificial reference star (*c*). With an array of laser beacons, astronomers can illuminate an entire field of view (*d*). A nearby star is still required to point the telescope.



strictive material, which expand or contract in response to a controlling voltage. The actuators are mounted on a rigid base plate. A continuous faceplate mirror tends to have better dimensional stability, requires less maintenance and provides a smoother correction over the telescope aperture. Another type of deformable mirror being developed is the bimorph mirror, consisting of flat piezoelectric elements glued to the back of a thin face sheet, which becomes bent when a voltage is applied. An adaptive optical system that employs a bimorph mirror is being engineered at the University of Hawaii. This system uses a wavefront sensor that directly measures wavefront curvature, simplifying the computation needed to control the actuators.

At infrared wavelengths, for which the effects of atmospheric turbulence are less severe, "modal correctors" potentially offer an elegant way of compensating for wavefront distortion. These devices are effective for correcting tilt, defocus, astigmatism and other more systematic distortions.

Other avenues currently under investigation include the application of neural networks to interpret signals from the optical sensor and to control the deformable mirror. Some researchers believe a neural network consisting of many interconnected decision-making cells can be trained to interpret these inputs better than the algorithmically based networks now in use.

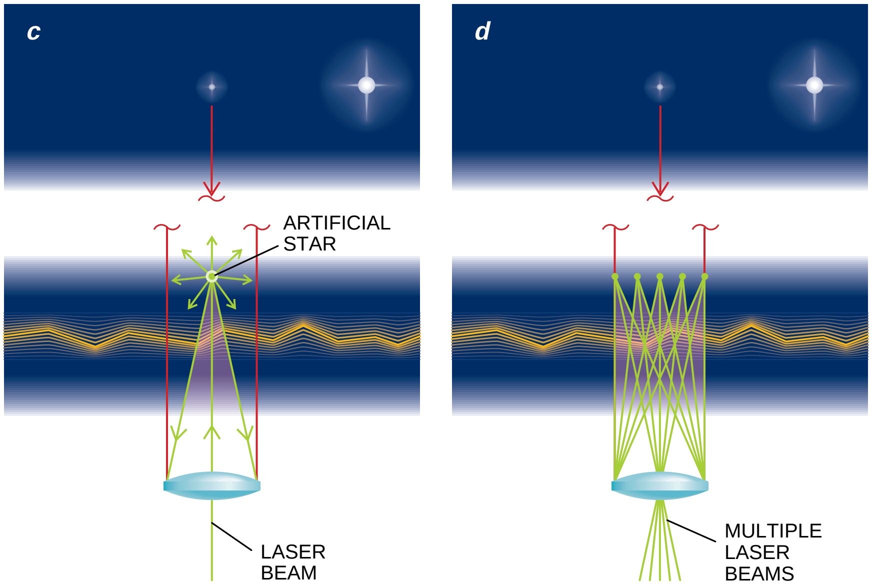
Although the initial application of adaptive optics has been very successful, the widespread adoption of this technology in observational astronomy is impeded by two fundamental problems. First, faint objects (a category that includes most of the objects of interest to astronomers) can be observed only when a bright star lies in close proximity to them. The need to make real-time measurements of turbulence at least as quickly as the atmosphere changes dictates how bright the guide star needs to be: enough photons must be collected within each small zone of the telescope aperture to make an accurate wavefront measurement. For an adaptive optical system working at visible wavelengths under average conditions, this means that at least 100 photons must be detected within each 10- by 10-centimeter zone in each one hundredth of a second. To meet that requirement, a guide star must be of the 10th visual magnitude or brighter. On average, only three stars of this magnitude are found in each degree-square patch of sky.

This limitation would be acceptable were it not for a second fundamental problem: adaptive compensation is effective only over an extremely small angle of sky, called the isoplanatic angle, which at visible wavelengths is usually less than five arc seconds across. Over a larger area the turbulence varies too widely from that measured by the wavefront sensor to get a consistently clear image. Thus, only the center would be correctly compensated, and the image would become increasingly blurred toward the edges. Because no more than a tiny area of the sky surrounding each guide star can be compensated, most of the sky is unavailable to adaptive optics using natural guide stars.

Researchers are working on two ways of circumventing these limitations. The first is to operate at longer (infrared) wavelengths, at which the optical effects of turbulence are much less severe. Because the value of r_0 at infrared wavelengths is five to 12 times greater than at visible wavelengths, each correction zone can be made correspondingly larger. Also, because disturbances in the wavefront take longer to change in large zones, more time is available for collecting light. As a result, dimmer stars can be used as guide stars. Furthermore, the isoplanatic angle is larger at longer wavelengths. Consequently, the area over which adaptive compensation is effective also increases. Taken together, these factors allow a visible guide star to be used to sharpen infrared observations over a much larger fraction of the sky than is possible at visible wavelengths.

The first infrared system, known as Come-On, was developed jointly by the European Southern Observatory and researchers in France in the late 1980s. It has been tested on the observatory's 3.6-meter telescope at La Silla, Chile, with excellent results.

The second approach is to use lasers to generate artificial beacons, or guide stars. Researchers at the Massachusetts Institute of Technology's Lincoln Laboratory and the U.S. Air Force's Phillips Laboratory fortuitously provided us with this powerful new way to measure atmospheric turbulence during their work for the Strategic Defense Initiative. In the 1980s they were studying how to fire a laser weapon so that it could deliver as much energy as possible to a target above the atmosphere. Because laser beams suffer the same kind of distortion at visible wavelengths that afflicts light from a distant star, the principles of adaptive optics can be applied. In 1982 M.I.T. researchers began employing a 69-actuator version of the CIS, known as the atmospheric compensation experiment (ACE), to correct distortion in a laser beam projected from the ground into space. In one experiment, the space shuttle *Discovery*



carried a retroreflector to bounce a laser beam back to the earth, where researchers used it to measure atmospheric distortion. In later tests, retroreflectors installed on rockets reached altitudes of 600 kilometers. By feeding this information into a deformable mirror, the workers were able to "predistort" a second beam so that it passed through the atmosphere and focused on a small target on the rocket. The ACE adaptive optics equipment has since been used successfully for astronomical work on the 60-inch telescope at Mount Wilson Observatory in California.

For astronomical telescopes, lasers create artificial guide stars in the upper atmosphere, either by producing backscatter from air molecules at altitudes of 10 to 40 kilometers, which is known as Rayleigh scatter, or by stimulating fluorescence in a naturally occurring layer of sodium vapor that lies at about 90 kilometers. Because a laser beacon is much closer to the telescope than is a natural star, the device generates a cone-shaped beam (rather than a virtually cylindrical beam), which passes through only a fraction of the turbulent atmospheric layer before it reaches the telescope aperture. This effect is more serious with the lower-altitude Rayleigh beacons and requires the use of more than one laser beacon.

In 1983 Robert Q. Fugate of Phillips Laboratory demonstrated the feasibility of using laser guide stars for wavefront measurement. Researchers at the M.I.T. Lincoln Laboratory created the first complete adaptive optical system depending on laser guide stars, known as SWAT (for short wavelength adaptive techniques). Between 1988 and 1990 at the U.S. Air Force's Maui Optical Site at Mount Haleakala, they used pulsed dye lasers operating at a wavelength of 0.512 micron to generate laser beacons at altitudes of four to eight kilometers. Compensation of turbulence was proved by comparing images of natural stars with and without the adaptive correction, and the experiment showed that two laser beacons yielded better results than one.

A different type of wavefront sensor, the Shack-Hartmann sensor, is generally used with laser beacons because it can handle either continuous or pulsed light sources. First used by Roland V. Shack of the University of Arizona in 1971, it employs an array of small lenses covering the optical beam, each producing an image of the guide star. Wavefront gradients are determined by measuring the image displacement in each zone.

Laser beacons should in principle permit the use of adaptive compensation on any celestial object, however faint, at any wavelength that can pass through the atmosphere. But the effectiveness of laser beacons is still limited by the need for a natural star by which to point the telescope. Laser beacons are useless for pointing because they are not fixed in the sky but vary in absolute position according to the effect of turbulence on the laser beam. Because of the need for a pointing star, adaptive optics can cover only roughly 30 percent of the sky at visible wavelengths. At infrared wavelengths, the sky coverage approaches 100 percent. Several organizations, including Phillips Laboratory, the University of Chicago and Lawrence Livermore National Laboratory, are pursuing the development of adaptive optics using laser beacons.

One of the unsolved problems of adaptive optics is how to create sharp images across large fields of view. For instance, no one has yet been able to obtain a fully compensated image of Jupiter's disk. The problem is that the disk is about 40 arc seconds across, which encompasses about 50 different isoplanatic patches, or areas in which atmospheric turbulence differs significantly. One often discussed approach is to employ multiple deformable mirrors in conjunction with an array of laser guide stars. Each mirror would in effect serve as a three-dimensional corrector by compensating for turbulence over a range of altitudes in the atmosphere. Multiple wavefront measurements covering the field of view would be made with the guide stars.

After 25 years of development, mostly for defense purposes, adaptive optics is now finding wider scientific uses, primarily in ground-based astronomy. Most large astronomical telescopes currently being planned or constructed include adaptive optics in their design.

A frequently raised question is why we continue to build large ground-based telescopes when we can escape atmospheric limitations completely by going into space. The answer is that space telescopes are far more expensive to build and maintain than those on the earth, even when the cost of adaptive optics is included.

Space- and ground-based astronomy should be considered complementary rather than competitive. Space telescopes have the advantage of being able to perform at wavelengths that are cut off by atmospheric absorption, detecting ultraviolet and x-ray radiation, for example. Ground-based telescopes, which at present have much larger apertures and baselines, are better suited for work at longer wavelengths, for which turbulence effects are more easily compensated. With each technology doing what it does best, our view of the universe will become clearer than ever.

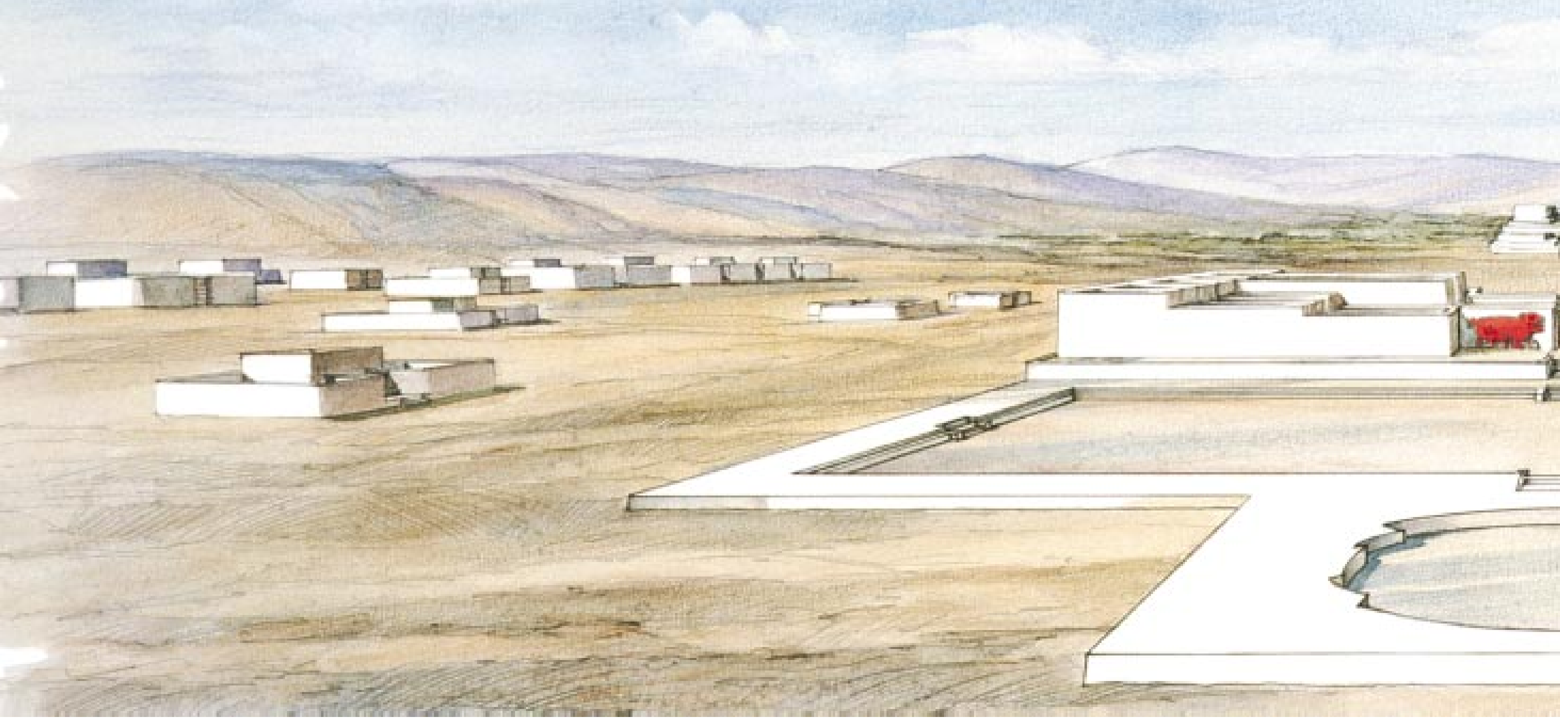
FURTHER READING

ACTIVE OPTICS: A NEW TECHNOLOGY FOR THE CONTROL OF LIGHT. J. W. Hardy in *Proceedings of the IEEE*, Vol. 66, No. 6, pages 651–697; June 1978.

ADAPTIVE OPTICS REVISITED. Horace W. Babcock in *Science*, Vol. 249, pages 253–257; July 20, 1990.

PRINCIPLES OF ADAPTIVE OPTICS. Robert K. Tyson. Academic Press, 1991.

ADAPTIVE OPTICS FOR ASTRONOMY: PRINCIPLES, PERFORMANCE AND APPLICATIONS. J. M. Beckers in *Annual Review of Astronomy and Astrophysics*, Vol. 31, pages 13–62; 1993.



Early Andean Cities

Some 3,800 years ago Pampa de las Llamas-Moxeke and Taukachi-Konkan were carefully laid-out urban centers that housed many hundreds of people

by Shelia Pozorski and Thomas Pozorski

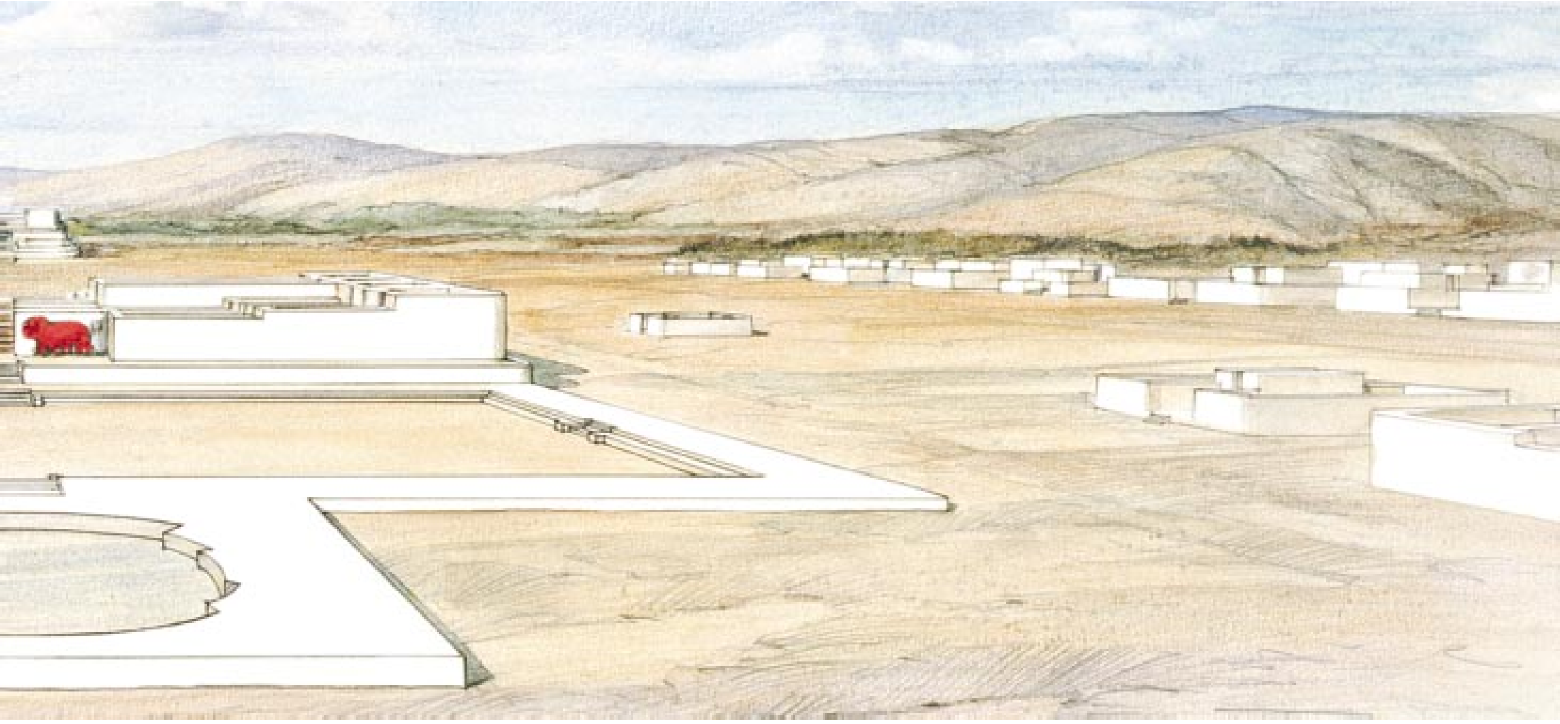
For more than half a century, archaeologists have believed that "true" civilization in Peru began with the Chavin Culture of what they call the Early Horizon, which extended from 1100 to 250 B.C. Yet on triangles of desert plain 350 kilometers north of Lima, nestled between coastward-pointing fingers of the Andes Mountains, lie the remains of cities at least 700 years older. Since 1985 we have been uncovering through a series of excavations two of these sites. They lie in the Casma Valley and testify to the existence of a complex civilization: there are temple-topped mounds as much as 30 meters high, administrative enclosures, large-scale irrigation systems and hundreds of houses for both rich and poor.

Not until we sent wood and charcoal samples from these sites for radiocarbon dating in 1980 did it become clear that the sites had been occupied during the Initial Period between 2200 and 1100 B.C., the time when ceramics, weaving and large-scale irrigation agriculture first appeared on the Peruvian coast. (This period corresponds roughly to the time of the Middle Kingdom in Egypt, a few hundred years after the building of the great pyramids.) Increasing knowledge about this civilization has yielded a new appreciation of its critical role in shaping subsequent cultures in the region.

Our understanding of the culture at these sites is largely the result of the peculiar climate conditions that prevail there. Significant amounts of rain last fell in Casma in 1983 and, before that, in 1925 and 1891. (Indeed, many of the structures we excavated had no roofs, as the climate makes them unnecessary.) As a result, the preservation of artifacts is nothing short of incredible: buried textiles, wood and even fragile leaves are often found in pristine condition. Furthermore, knowing that cultivators practiced irrigation—large-scale agriculture would have been impossible otherwise—gives archaeologists additional clues in piecing together the vanished culture's social structure.

While it flourished, Pampa de las Llamas-Moxeke, located along the southern branch of the Casma, was a bustling

SHELIA POZORSKI and THOMAS POZORSKI have carried out archaeological fieldwork in Peru since 1970. Both are associate professors of anthropology at the University of Texas Pan American in Edinburg, Tex. Shelia Pozorski is especially concerned with prehistoric subsistence and its relation to early societal development. She received her B.A. from Harvard University and her Ph.D. in anthropology from the University of Texas at Austin in 1976. Thomas Pozorski also earned his B.A. from Harvard and his Ph.D. from the University of Texas at Austin in 1976. His interests lie in the development of complex societies within the Andean area.



center that housed about 2,500 people within an area of two square kilometers. Two massive mounds dominate the site: Moxeke on the south and Huaca A on the north. The two structures face each other across a series of symmetrical plazas, forming a central axis that establishes the orientation for almost all other public architecture.

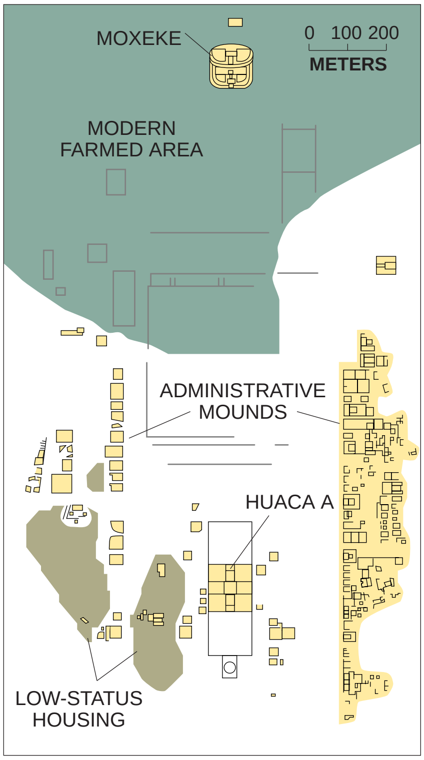
Flanking this axis are more than 110 intermediate-size administrative buildings arranged in parallel rows. Most face inward toward the site's center, and most are small mounds, like miniature, simple versions of Huaca A. There is also evidence of "urban renewal." Low-status houses appear to have been razed to gain space for new rows of intermediate-size mounds, several of which were never finished.

Moxeke, a U-shaped mound about 160 meters wide, 170 meters long and 30 meters high, probably served as a temple from which ceremonies were performed for large crowds in the plazas surrounding it. Huaca A (119 meters wide, 136 meters long and about 15 meters high) is reached via much smaller plazas at either end. Because of the downward ground slope from north to south, however, the tops of the two mounds appear to be about the same level. To the northeast of Huaca A is a sunken circular plaza as well. This architectural feature is common to many sites of the period.

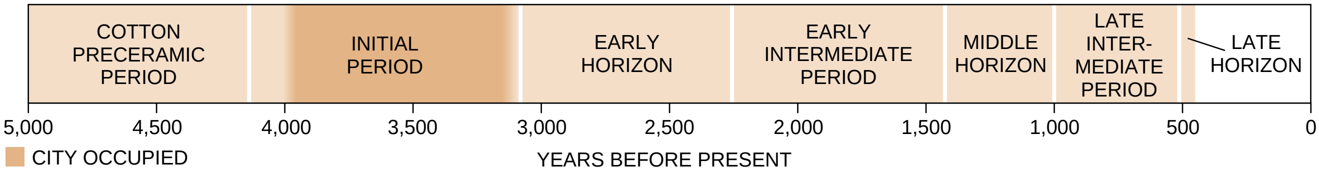
Planning is evident in all aspects of the site's architecture. All the public buildings exhibit internal symmetry—the two halves of the Moxeke mound are mirror images, for example. In keeping with its accessibility from plazas at either end, Huaca A has a four-way internal symmetry.

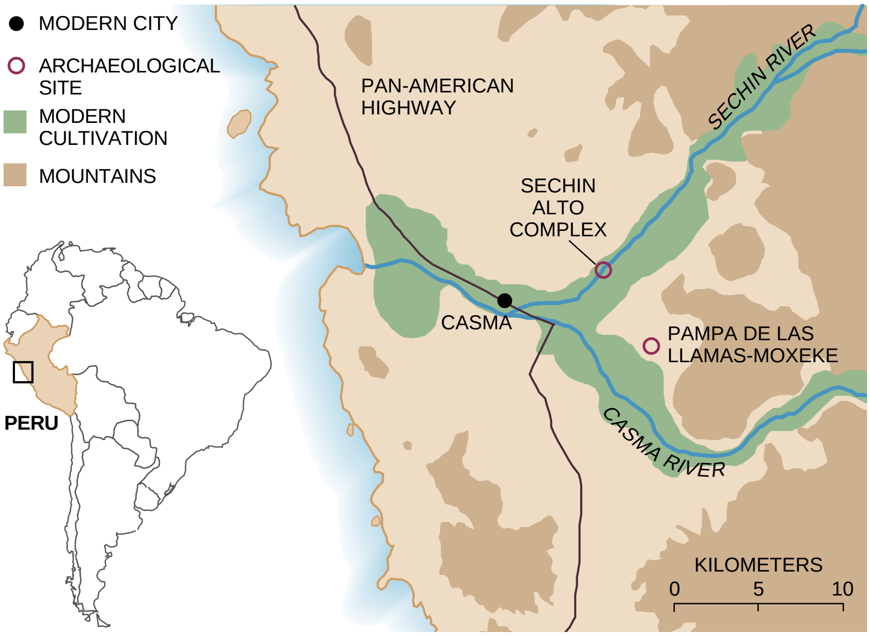
Huaca A is replete with what seem to be control devices: gates, barriers and walled compounds. These structures are probably related to its function as a huge warehouse. Pampa de las Llamas-Moxeke administered much of the surrounding area, and smaller settlements supplied labor and goods to its inhabitants. The mound's rooms appear to have held at least 4,400 cubic meters of agricultural products and luxury goods.

The rows of smaller mounds, meanwhile, appear to have served as lower-level bureaucratic centers. They probably contained offices and storage areas used by mid-level officials involved in



HUACA A mound at Pampa de las Llamas-Moxeke would have glistened white in the desert sun 3,800 years ago. The structure stood 15 meters high and stored at least 4,400 cubic meters of valuables and commodities. It was part of a city, built according to plan (*right*), that housed more than 2,000 people. The political entity to which Pampa de las Llamas-Moxeke belonged contained more than half a dozen cities scattered over 1,000 square kilometers. Civilization in the region evolved over a period of roughly 4,000 years (*below*).





PERUVIAN COASTAL DESERT is ill suited for agriculture but almost perfect for archaeologists. Arid conditions have preserved perishable items for more than 3,000 years. The sites investigated by the authors lie in one of more than 50 river valleys that cut through the desert.

the awesome task of acquiring, monitoring and redistributing the vast quantity of commodities stored and protected within Huaca A.

Residential architecture occupies a surprisingly small percentage of the total area of Pampa de las Llamas-Moxeke, but we have found both low- and high-status houses. Public officials probably occupied the finer homes, which had plastered stone walls as well as internal storage rooms and niches for their personal wealth. In contrast, most of the population—most likely farmers, laborers and artisans—lived in small, crowded, irregular clusters of rooms with mud-coated cane or wood walls. These dwellings left few remains except for low, stone-wall footings.

For more than 40 years after the Casma Valley sites were found, no one realized just how old they were. Investigations of the area began in 1937, when renowned Peruvian archaeologist Julio C. Tello excavated the mound of Moxeke. He uncovered several huge adobe friezes—high-relief sculptures depicting human figures and massive heads—on the front of the mound and extending slightly around to the sides. These friezes had been brilliantly painted in red, blue, green, black and white, and their discovery brought the valley to prominence in studies of early Andean civilization. Tello attributed the spectacular sculpture to the highland Chavin Culture, which characterized the Early Horizon.

In 1980 we examined the site, along with five others, as part of our effort to construct a precise chronography of the many "early" Casma Valley sites. We also needed to decide where to excavate on a larger scale. From the start, Pampa de las Llamas-Moxeke was full of surprises. Although the Initial Period is marked by the widespread introduction of pottery, we found relatively small amounts of very simple ceramics. Most were jars without necks, probably modeled after the gourd containers used before the invention of fired clay vessels. Simple woven textiles reflected the beginnings of weaving, but alongside them lay elaborate twined textiles (twisted together by hand rather than on a loom) characteristic of prepottery cultures. The remains of cultivated plants were abundant, but none of our more than 50 test pits yielded maize, a staple of later Andean populations.

All these data suggested that Pampa de las Llamas-Moxeke was a very early site; nevertheless, we were unprepared for the results of carbon 14 dating of charcoal and wood samples. Nine initial samples (and 15 collected during subsequent research) yielded calibrated dates between 2000 and 1500 B.C.—hundreds of years before the beginning of the Early Horizon.

In 1985 we returned to start excavations in earnest. We concentrated on the northern two thirds of the site because the southern third has been affected by both later prehistoric reoccupation and modern agriculture. We sampled large mounds, intermediate-size mounds and enclosures, and small and irregular house structures, seeking clues to the activities carried out within and around them and to their role in the site as a whole.

Tello had already worked extensively at Moxeke, so we chose Huaca A. The mound looked easy before we got under way: its four-way symmetry, readily visible from the surface, meant that excavation of only one quadrant would elucidate the structure of the entire edifice. Even before excavation, the walls were fairly straightforward to trace on the surface.

Our Peruvian workmen, 15 to 25 in number, began by moving huge fallen wall boulders weighing up to 150 kilograms. They used the wheelbarrow we furnished to shade the water jug, preferring to carry the stones atop a padded cushion on their backs in much the same manner as the original builders of Huaca A.

More than two meters down, we finally encountered a floor. The volume of the fallen rubble implies that the walls of the larger central rooms were originally at least five meters high. In the walls' upper reaches, we could see bare stones still held in place by mud mortar, their flat faces forming the wall face. The very top courses of the walls were probably made of cone-shaped adobe bricks, whose lighter weight would have eased construction.

The granite stones making up the walls were covered with a thick layer of coarse mud plaster, replete with the long finger strokes characteristic of prehistoric masons. The fine finishing plaster, rich in clay, had flaked off the upper parts of the walls, but lower down it was well preserved, as was the finely plastered floor. Traces of pigment reveal that the walls and floors were painted white. The entire mound would have glistened in the desert sun.

Ultimately we were able to define the walls for the north quadrant of Huaca A; from this, we reconstructed the configuration of the entire mound summit. The whole edifice consists of repetitions (in various sizes) of a basic construction unit: a square room with rounded exterior corners and two rounded and two square interior corners. The entrance of each room has a raised threshold; the upper walls contain large niches 1.25 to two meters above the floor. Twined and woven sedge mats carpeted the floors of these square rooms; their impressions remain in the clay-rich plaster of the floors.

The largest units lie along the central northeast-southwest axis of the mound; rooms become increasingly smaller as one moves toward the periphery. Spaces between the square-room units seem to have served mainly as access corridors. They generally contained neither matting nor niches.

To find out how people approached Huaca A and moved about on its summit, we cleared both the northeast and southwest atria and every entrance within the north quadrant. We uncovered evidence that access to the mound was so carefully restricted and monitored that the associated bureaucratic system boggles the mind. Yet the physical controls we found are more suggestive of a powerful underlying authority than of insurmountable physical barriers.

Intimidation (or inspiration) of a visitor began as he or she ascended either central staircase to enter one of the atria of Huaca A. The rear walls of both atria contained magnificent friezes on either side of the central entrance, clearly designed to impress the approaching visitor. The frieze within the northeast atrium is better preserved: there, sculpted in low relief in a medium of clay-rich silt, are the remains of two immense felines, probably jaguars.

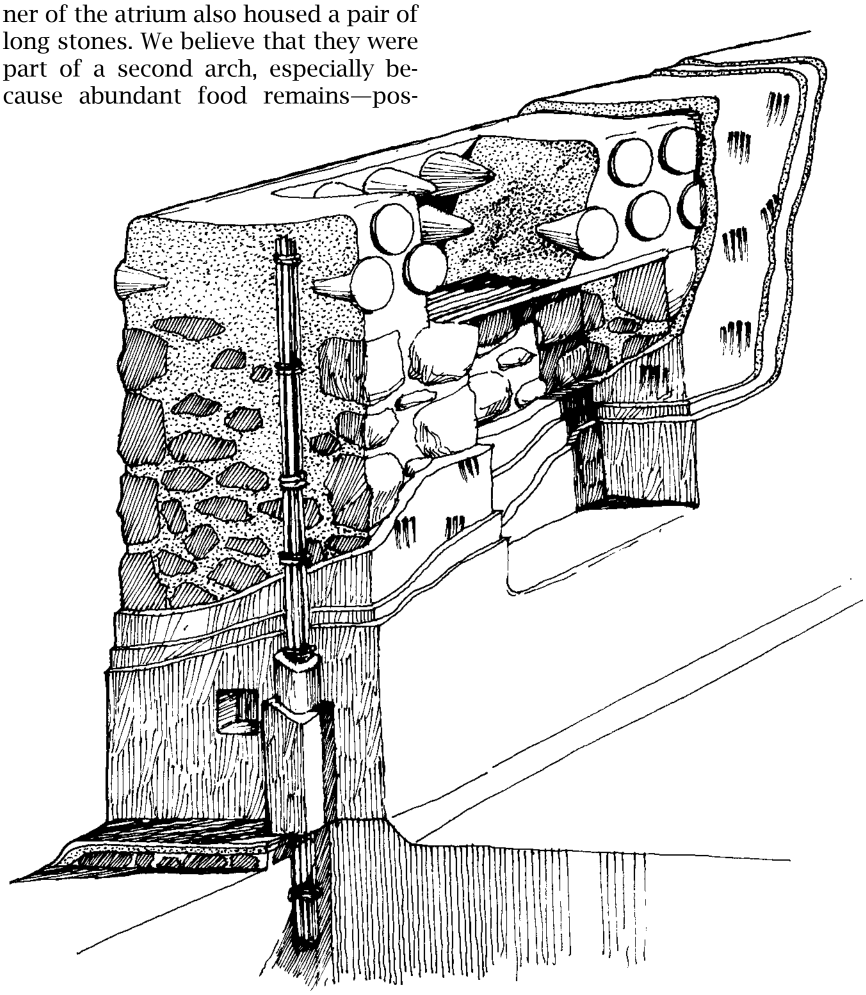
These colossal figures, once more than six meters high and 10 meters long, face each other across the northeast entrance into the mound. The curving walls of the atrium enhance the three-dimensional effect of the giant cats. Traces of red pigment suggest that they were once brightly colored. After 3,800 years, only the feet and serpent-shaped tails remain. Profile felines are frequently represented in Andean art, however, and therefore it is an easy task to reconstruct the body and head.

Along with the felines, the space immediately adjacent to the gateway contains circular reliefs whose central circles hold four evenly spaced rectangles. This design occurs on the ends of victorious warriors' clubs at the nearby site of Cerro Sechin, leading us to suspect that it is a symbol of authority.

In the southwest atrium, we found an elaborately carved stone almost 50 centimeters long. A double-bodied serpent design is incised into one face. An adjacent surface of the stone contains a surprisingly realistic carving of the impression of a right hand [*see illustration on page 72*]. This stone was accompanied by two much longer ones that probably supported it to form a small arch with the serpent facing outward and the hand downward—possibly a shrine or an altar. The west corner of the atrium also housed a pair of long stones. We believe that they were part of a second arch, especially because abundant food remains—possibly offerings—lay scattered nearby.

Each entrance within the mound was narrowed by opposing pilasters, supported internally by small branches lashed into a post-size bundle. Larger pieces of wood, a rare commodity in the desert, in all likelihood served as lintels above some entrances and over most of the wall niches within the rooms. A gate of large logs controlled access through the main northeast entrance of the mound. The log gate, located on the opposite side of the wall containing the frieze, consisted of four upright logs about 3.5 meters long mounted on either side of the entrance. Six additional logs, each almost exactly 4.1 meters long, fitted into the gap between the vertical logs and the walls. Their large and tapered ends alternated to create a level entrance barrier.

Other entrances on Huaca A—93 in all—were protected by one or more bar-closure mechanisms, devices never before described at early Peruvian sites. While first clearing these entrances, we noticed small square stone niches in their side walls. These niches always oc-



WALLS OF HUACA A were built mostly of granite stones covered with layers of coarse mud plaster and a fine clay-rich finishing coat. Niches held luxury goods and staple commodities. Room entrances were flanked by columns made of woody plants tied into bundles and coated with plaster. Conical adobe bricks (*left*) made up the top courses of the walls.

ANCIENT FRIEZES are still visible at the bottom of the walls that flanked the main entrance to the Huaca A mound. Their feline images once stood more than five meters high and were painted with red and white pigment.

curred in opposing pairs about half a meter above the entrance floor, and one member of each pair was consistently many times deeper than the other.

The purpose of the niches became clear as a result of simultaneous excavations in a nearby compound. An empty niche 35 centimeters deep was opposed by a niche that still had a horizontal wooden pole extending from its center. The pole was firmly embedded in rubble, so we cleared along the wall top and discovered that it was housed in a well-constructed stone chamber more than two meters long. Freed of debris, the pole was pulled across the opening to rest in the opposing niche, barring access to the compound as it had 3,800 years ago.

Why were all of the entrances barred? There is no evidence that anyone lived on Huaca A: neither cooking hearths nor refuse. The mound was neither a heavily guarded palace nor royal living quarters. Instead Huaca A was apparently used to store commodities and other valuable objects. We found only a few artifacts during the excavation—to be expected because valuables would have been removed when the site was abandoned. What remained were textile fragments in the wall niches, either from stored cloth or cloth containers. In addition, soil samples from the niches yielded pollen from cotton as well as from food plants, including beans, potatoes, sweet potatoes and peanuts, indicating that such commodities were stored there. (Contrary to commonly held beliefs that corn was an essential staple for New World civilizations, it was not among the agricultural products we found.) Thousands of rodent bones confirm the evidence of food storage.

There are tantalizing clues about the people who used Huaca A and monitored traffic within it. The finest textiles from the site were found on the mound, as were many of the turquoise beads and a unique wooden figurine. These items suggest that only the wealthier elite inhabitants of the city had ready access to Huaca A; if the lower classes had frequented the mound, we would have found evidence of their possessions there as well.

The square-room unit that dominates Huaca A is duplicated in the rows of intermediate-size public buildings that parallel the main site axis. Each of the compounds near the main mound consists of one such unit, and single square-room units form the central core of each of the smaller mounds in the longer rows. These buildings were apparently occupied by bureaucrats who monitored the flow of commodities into and out of the mound.

Behind several of the small mounds stand the remains of upper-class houses. Like the mounds and enclosures, these structures are aligned with the site axis; they are well made, with high walls of stone with mud plaster. There is also ample evidence of domestic activities. One room has a large, square, stone-lined hearth with a central circular depression for the fire. Small rooms containing large jars, rooms with wall niches and subfloor chambers were used for storage within the houses of the elite. The residences lack the bilateral symmetry of the public architecture, however, and household debris surrounds them.

Both Huaca A and the small mounds and enclosures were kept free of household refuse, and so this garbage yields the best information we have about the lives of the bureaucrats. Fragments of solid pottery figurines are common within and near the elite houses. We also found many fragments of stone mortars and pestles. These stones bear traces of red pigment that were used to paint friezes, walls and possibly the bodies of citizens.

Most interesting, in light of the site's developed bureaucracy, are what appear to be stamp seals and cylinder seals made of pottery. Both types are rare in the Andean area. Although later Andean cultures used stamp seals to impress designs on pottery, there is no evidence that they served this function at Pampa de las Llamas-Moxeke. Instead two examples contain red pigment residues, indicating that they were used to apply colored decoration on surfaces such as textiles or human skin. We think it is possible the seals served as symbols of power and authority, much as they did in the centers of early civilization in the Mediterranean and the Near East, where they are more frequently found.

Low-status housing at Pampa de las Llamas-Moxeke shares some characteristics with the elite domestic architecture; after all, both classes performed many of the same basic household chores. Refuse from food preparation and other domestic activities is abundant, and square, stone-lined hearths are common but much smaller. Many low-status residents cooked instead over cloverleaf-shaped hearths in which three upright stones supported round-bottomed pots. Moreover, the construction of low-status houses differed substantially from that of high-status residences: the rooms are smaller, irregular and clustered, and the stone parts of the walls (still extant) are but a few stones high. The upper walls were made of perishable cane or wood covered with mud plaster.

Although our excavations have revealed a society of unsuspected complexity for its time, continuing field research has made us aware of a much larger, integrated political entity. In the northern branch of the Casma Valley lies the Sechin Alto Complex—four sites that collectively cover 10.5 square kilometers, more than five times the area of Pampa de las Llamas-Moxeke. Before 1992, detailed scientific ex-

CARVED STONE (possibly an altarpiece) found in the atrium of Huaca A was incised on the front and bottom faces. In use, it was supported by two rough stones and probably surrounded by offerings of food.

cavations had been carried out only at the smallest site, Cerro Sechin. Since then, we have uncovered a wealth of evidence that links all the sites in the complex and ties them to Pampa de las Llamas-Moxeke.

Our excavations in 1992 and 1993 concentrated on the best-preserved Sechin Alto site, Taukachi-Konkan. Immediately we became aware that the people of Taukachi-Konkan used the same kinds of pottery and architectural elements we had seen five kilometers to the south. The large mounds contained the same square-room units, wall niches and restricted entrances as did the intermediate-size ones. The site's center is open, formed by several large rectangular plazas lined by intermediate-size mounds. Two large mounds open toward sunken circular plazas similar to the one adjacent to Huaca A. Radiocarbon dates of six samples range from about 2000 to 1300 B.C.

The sites are very similar in some ways, but their differences have proved even more significant, leading us to realize that the early society in the Casma Valley was more sophisticated than we had imagined. Taukachi-Konkan's largest mound, the Mound of the Columns, is unlike any of the others that have been excavated. It is located on the western edge of the site, and its summit is divided into two large central courts with numerous square-room units arranged symmetrically around them. On the front half of the summit we found the remains of about 100 round columns made of bundled cane, rope and stones covered with mud mortar and plaster. These columns supported a roof—a unique feature that presumably shaded members of the elite and screened them from view.

We believe the Mound of the Columns is probably the residence of an extremely important person, possibly the ruler of the Sechin Alto complex or the entire Casma Valley system. In addition to the roofed court, it contains a few roofed rooms lined with high niches that were most likely storerooms to house valuable goods both received and given. Other rooms have low niches more appropriate for seating and were possibly meeting or audience areas. Rooms on the rear half of the summit are irregular in layout and less accessible, leading us to believe they might have been living quarters. Two round sweathouse structures (enclosed, heated rooms intended for small religious ceremonies) are also located there.

The placement of this palace at Taukachi-Konkan is logical because the Sechin Alto complex probably served as the capital of a valleywide political entity. Although both sites are large and important centers, we now know that each formed a vital part of an even larger, more complex society that was well developed by 1800 B.C.

After enduring for the better part of a millennium, these cities were abandoned. The many buildings left unfinished at both sites suggest that local development stopped suddenly at its height, rather than declining gradually. Researchers first speculated that invaders from highland cultures made war on the coastal sites and disrupted the civilization. Evidence of the demise of the Initial Period polity coincides with the abrupt appearance of maize, domesticated animals, new kinds of artifacts and different building styles. We have recently uncovered evidence, however, that internecine conflict weakened the Casma Valley culture and helped the invaders' cause.

Stone carvings in Cerro Sechin depict, at life size, victorious warriors and their mutilated victims—a clear representation of battle. The warriors' facial decorations and the rectangle-in-circle symbols on their war clubs (like those on the Huaca A frieze) argue that the victors were local. Yet the clothing of the victims matches that of humanlike figures that decorate the front of the Moxeke temple mound.

We can only guess which factions warred against one another. Cerro Sechin may commemorate the victory of one segment of the elite over another; perhaps bureaucrats associated with the more secular activities of Huaca A deposed the religious hierarchy of Moxeke in a decisive conflict between church and state. Then again, the carvings may represent a rebellion by the general population against the proliferating bureaucracy. As these planned cities grew and their elite population expanded, demands on the laboring populace could have become excessive.

Although archaeologists may never know precisely how these cities rose and fell, their extreme age makes them crucial to the understanding of Peruvian prehistory. Pampa de las Llamas-Moxeke and Taukachi-Konkan do not represent an unknown or undiscovered civilization. Rather their new placement so far back in time demonstrates that the Initial Period was not merely a prelude to "true" civilization. Instead this time span witnessed the development of fundamental social, political, economic and religious principles that shaped Peruvian culture for the next 3,000 years.

FURTHER READING

RECENT EXCAVATIONS AT PAMPA DE LAS LLAMAS-MOXEKE, A COMPLEX INITIAL PERIOD SITE IN PERU. Shelia Pozorski and Thomas Pozorski in *Journal of Field Archaeology,* Vol. 13, No. 4, pages 381–401; Winter 1986.

EARLY SETTLEMENT AND SUBSISTENCE IN THE CASMA VALLEY, PERU. Shelia Pozorski and Thomas Pozorski. University of Iowa Press, 1987.

THE ORIGINS AND DEVELOPMENT OF THE ANDEAN STATE. Edited by Jonathan Haas, Shelia Pozorski and Thomas Pozorski. Cambridge University Press, 1987.

CHAVIN AND THE ORIGINS OF ANDEAN CIVILIZATION. Richard L. Burger. Thames and Hudson, 1992.

EARLY CIVILIZATION IN THE CASMA VALLEY, PERU. Shelia Pozorski and Thomas Pozorski in *Antiquity,* Vol. 66, No. 253, pages 845–870; December 1992.

The Sensory Basis of the Honeybee's Dance Language

Novel experiments, such as training bees to respond to sounds and recruiting them using a robot, have ended several debates surrounding the dance language

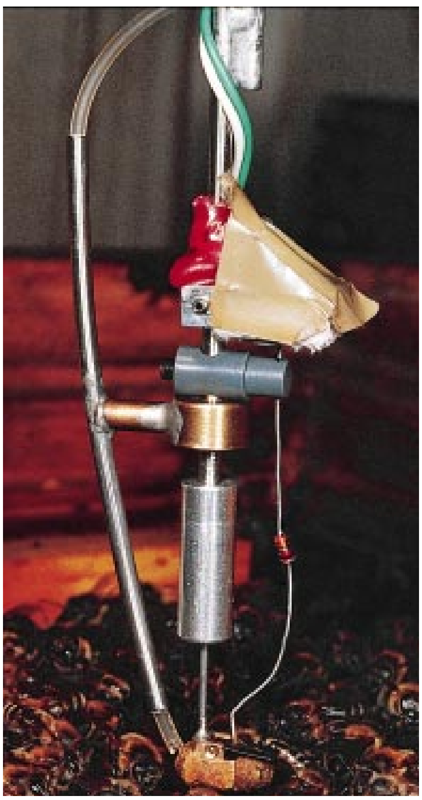
by Wolfgang H. Kirchner and William F. Towne

For many centuries, naturalists have observed that honeybees tell their nestmates about discoveries they make beyond the hive. Nevertheless, the system of communication that the insects use remained a mystery until the 1940s, when Karl von Frisch of the University of Munich in Germany first discovered the significance of bees' dances. In the hive the steps and waggles of a successful forager correlate closely with the exact distance and direction from the nest to the resource she has discovered. For the next two decades, most scientists believed bees relied primarily on these silent movements to communicate.

In the 1960s this view was challenged in two ways. The first challengers were Adrian M. Wenner, then a graduate student at the University of Michigan, now at the University of California at Santa Barbara, and Harald E. Esch of the University of Munich, now at Notre Dame University. Working independently, the two researchers discovered that the dances were not silent after all. As the bees dance, they emit faint low-frequency sounds, and Wenner and Esch both suggested that the sounds might play a critical role in the bees' communication. The use of sounds, they reasoned, might account for the bees' ability to communicate effectively in the complete darkness that prevails inside their nests. At the time, however, many scientists believed bees were deaf, and so the issue remained open.

Wenner later raised the second challenge to von Frisch's description of the dance language, rethinking his first hypothesis at the same time. Bees, he argued, use none of the information in the dances or the sounds. Instead he proposed that the insects rely on odors to find the new resource advertised by the dancer.

Now both of these debates have been resolved. Bees, it turns out, can hear, and their ears are well suited for detecting the sounds associated with the dances. Observation of how the insects respond to a robot that dances and sings like a live forager shows that both sound and dance are needed to communicate information about the location of food. Silent dances, the experiment demonstrates, communicate nothing, and sound without dance also fails. Odors too are involved but appear to lack the



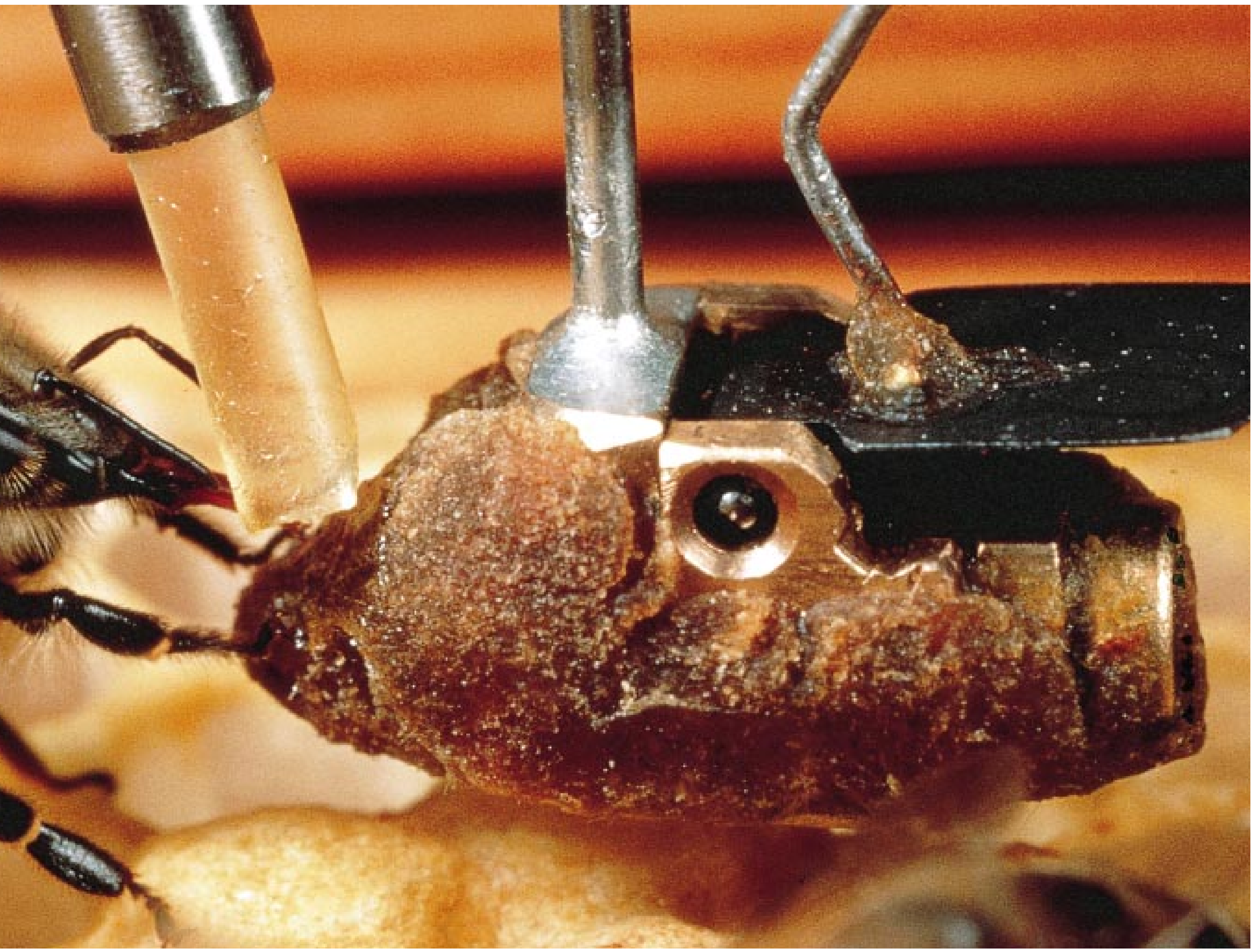
ROBOTIC HONEYBEE, directed by a series of step motors, can dance and emit sounds similar to those a live forager makes (*above*). Experiments using this robot have enabled researchers to explore how sounds are involved in the dance language. A syringe ending near the artificial dancer's head delivers sugar water to its followers, much in the same way that a live forager offers samples of food to her nestmates (*right*).

importance that Wenner ascribes to them. Beyond the resolution of these issues, we have also recently learned much more about the nature of the dance sounds, the bees' sense of hearing and the aspects of the dance that are most essential in the communication.

WOLFGANG H. KIRCHNER and WILLIAM F. TOWNE began their collaboration in 1987 and jointly discovered the auditory sense of the honeybee. Kirchner has studied the mechanisms and evolution of dance communication in honeybees for the past 10 years. He received a Ph.D. in 1987 from the University of Würzburg in Germany, where he currently holds a faculty position. Towne has studied the dance communication of honeybees since 1980. He received a Ph.D. in biology from Princeton University in 1985 and is an associate professor of biology at Kutztown University in Pennsylvania.

A certain pleasure comes from the solution of such a long-standing mystery. Aristotle himself documented the honeybee's ability to recruit her nestmates to a good food source but did not speculate on how the communication took place. He and other naturalists did, however, observe that a bee that finds a new food source returns to the nest and dances for her sisters, rather than feasting alone. Pliny reportedly constructed a hive that had a window made from transparent horn, through which he could monitor dancing bees. By studying bees kept in a glass-walled hive, von Frisch and his followers in the 20th century were able to recognize a pattern in the dance: The forager walks across the vertical sheets of comb hanging in the hive and traces out the shape of a figure eight; she pauses in each loop to shake her body from side to side. A few potential recruits chase after the dancer attentively for some time and then fly out on their own toward the target. Provided they like what they find, these recruits return to the nest and dance as well, sending even more bees from the hive to investigate the site. Eventually, the best food sources inspire the most dances and so attract the most bees.

In the 1920s von Frisch first proposed that the forager's dance somehow gave the other bees information about the food source. But it was not until 1943 that von Frisch discovered that the direction in which the dancer faced during her waggling run pointed toward the food site in relation to the sun: If she waggled while facing straight upward, toward the 12 on a clockface, then the food could be found in the direction of the sun; if she waggled 60 degrees to the left of 12, facing the 10 on a clockface, then the food lay 60 degrees to the left of the sun. In addition, he noticed that how fast the dancer completed her circuits corresponded to the distance between the hive and the feeding site: the closer the food, the more frenzied her pace. Von Frisch and



his colleagues made detailed accounts of the dance language. With a stopwatch close at hand, they could observe the dance, decipher its meaning and then locate the food supply of which it spoke. The accurate translations were a stunning accomplishment.

Nevertheless, several serious questions remained unanswered. First, investigators could not prove that the dance truly functioned as a language. Perhaps the correlation between the excited forager's movements and her finds in the field were merely coincidental and of no real significance to the bees themselves. Moreover, researchers had yet to figure out how the bees perceived and interpreted the dances. The scientists could clearly see the dancer through the glass wall and clock her movements, but the bees themselves, normally enclosed in dark nests, certainly did not have this benefit. What were the dances like from the bees' point of view?

Years after von Frisch interpreted the symbolism of the dances, Wenner and Esch independently found that dancing bees make sounds during their waggling run. Both men suggested that the sounds might help the dancer attract an audience in the dark nest. Many researchers doubted this premise because they thought bees could not hear airborne sounds. Still, the notion was not ignored altogether. Many insects, including bees, are quite sensitive to vibrations. Hence, some investigators speculated that the sounds the foragers produced could vibrate the combs under their feet as they danced. The comb vibrations might then advertise the dance to those bees who could not otherwise see the forager.

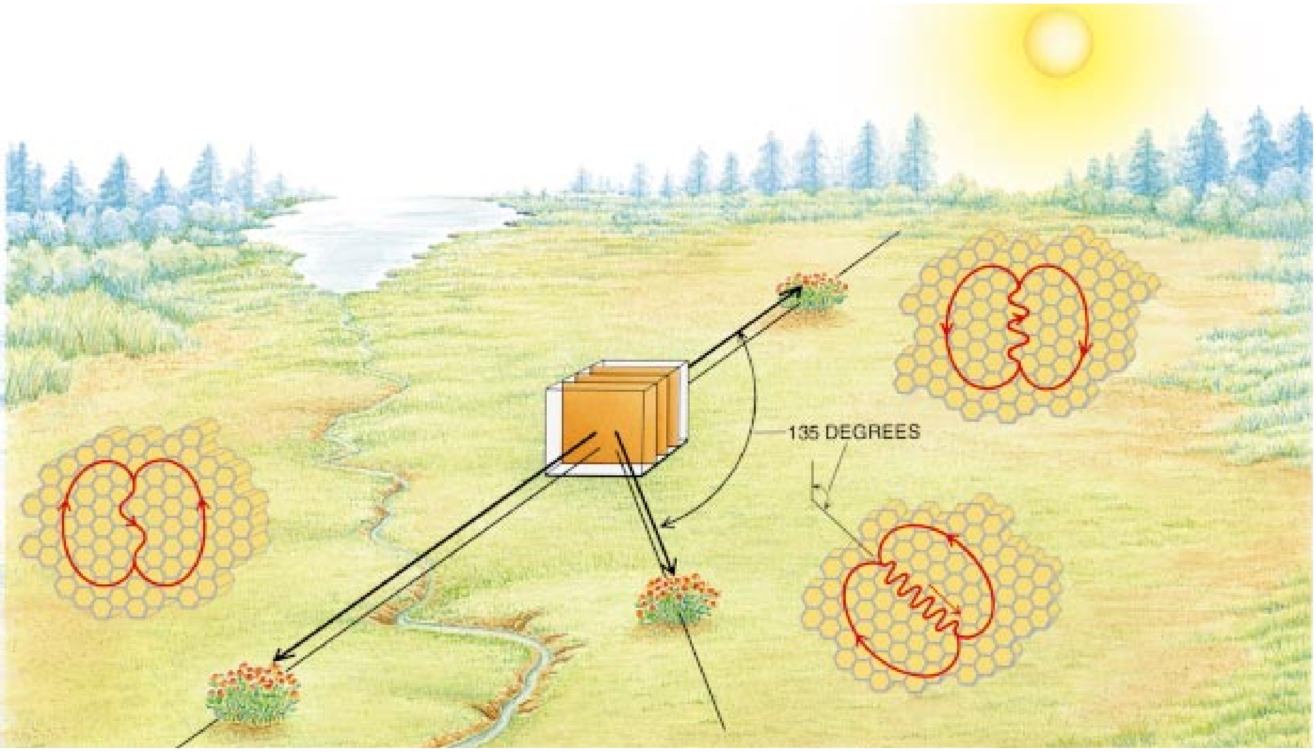
One of us (Kirchner), together with Axel Michelsen of Odense University in Denmark, answered part of this question several years ago. In their experiments, Michelsen and Kirchner aimed a laser beam at the comb near a dancing bee to determine whether or not the dance sounds generated vibrations in the comb. Surface vibrations, if any occurred, would cause minute changes in the light reflected from the comb. In this way, it was possible to measure the vibrations without touching the comb and possibly triggering additional tremors. These measurements revealed that dancing bees do not rattle the comb but that their audience does. The dance attenders sometimes emit a short squeak by pressing their thoraxes against the comb. This action vibrates the comb enough so that the dancing bee stops her movements. She then doles out small samples of the food she has collected so that her audience knows not only the direction and distance to the feeding site but how the food smells and tastes as well.

RETURNING FORAGER reports about her finds beyond the hive by dancing and beating her wings, thereby producing low-frequency airborne sounds. Her nestmates crowd around her and hear the sounds by using organs in the second joint of their antennae. They monitor the dance attentively for a short while and then fly from the nest to search out the feeding site alone.

Although this result was intriguing, we still did not know whether honeybees could hear the sounds a forager made during her waggling run. Because the sounds seemed likely to be involved in the dance communication, we have analyzed them thoroughly. Initially, we discovered that the forager made the dance sounds by beating her tiny wings and that the follower bees stood quite close to her when she did so. Therefore, our measurements of the sounds needed to be made in the same proximity.

In further collaboration with Michelsen, we gauged both the pressure changes and the air-particle movements near the dancer. Some insects' ears do not respond to sound pressures as human ears do but instead react to back-and-forth oscillations of nearby air molecules. In fact, it turned out that the pressure changes a dancing bee produces when she beats her wings are relatively small—which no doubt explains why the pressure-sensitive ears of scientists did not detect them for so long. But we found significant air movements within a few millimeters of the dancer's vibrating wings. We concluded that whereas the dance followers produce sounds by vibrating the comb, the forager transmits her own sound signals exclusively through the air. As a result, the noises from the forager's instructions and her followers' requests do not drown one another out.

Two research teams, using different experimental approaches, then tested the hypothesis that the forager transmits dance information acoustically. In the first line of inquiry, Kirchner and Kathrin Sommer, a graduate student at the University of Würzburg in Germany, changed the dance sounds simply by shortening the dancer's wings slightly. Clipped wings have a smaller vibrating surface and so produce sounds that have a higher pitch and a smaller amplitude. Sommer found that bees whose wings were experimentally shortened continued to forage and dance normally. The insects could not, however, re-



WAGGLING DANCE, performed on the vertical sheets of comb hanging in the nest, follows the path of a figure eight. In each loop the dancer waggles only when she is facing the direction of the food source in relation to the sun. If the food can be found by flying toward the sun, the dancer waggles facing straight up. If the food lies 135 degrees to the right of the sun, she waggles facing 135 degrees to the right; if the food can be found by flying away from the sun, then she waggles facing down. Her pace reveals the distance between the food and the hive: the faster the dance, the closer the food.

cruit nestmates. Similarly, bees from a mutant strain, called "diminutive wings," that have congenitally small wings cannot recruit at all—even though they fly and dance normally. Sommer studied a colony of honeybees, half of whom were normal bees and half diminutive-wing bees. She found that the normal dancing bees recruited both normal and short-winged bees equally well; in contrast, the mutant dancing bees recruited both strains poorly.

In the second series of experiments, Michelsen and Kirchner used a model bee that could perform the honeybee dances. Others had employed robotic bees in their research before, but none of the devices could make the correct sounds while it danced. Thus, a new attempt seemed worthwhile. Michelsen and his co-workers in Odense constructed the computer-controlled model bee, which danced in Würzburg for five summers.

Michelsen's team fashioned the model bee from brass and coated it with a thin layer of beeswax. The brass bee is slightly larger than a worker honeybee. Workers cut a razor blade to the appropriate size to make a set of wings. An operator can vibrate these metal wings on command by tweaking a stiff wire that connects them to an electromagnet. He or she can rotate the model in place by turning a thin rod attached to its back. A step motor at the far end of the rod steers the model's rotations automatically during its figure-eight dance. This same motor makes the model waggle from side to side. An x-y plotter, wired to a sliding metal sleeve placed around the rod, moves the model backward, forward, or left or right as needed. Finally, a thin plastic tube that ends near the model's head delivers food samples (sucrose solution) from a syringe. Yet another step motor regulates this mechanism. A desktop computer controls all the motors, which in turn direct the dance.

Each experimental session typically lasted for three hours. First Michelsen and Kirchner scented the model and its samples of sugar water with a faint floral fragrance, then placed baits in the field that gave off trace amounts of the same odor. At each of the baits, an observer recorded the approach of searching bees. The results showed repeatedly that the mechanical dancer could indeed recruit live bees. Most of the bees invariably went to the bait in the direction indicated by the model's dance steps.

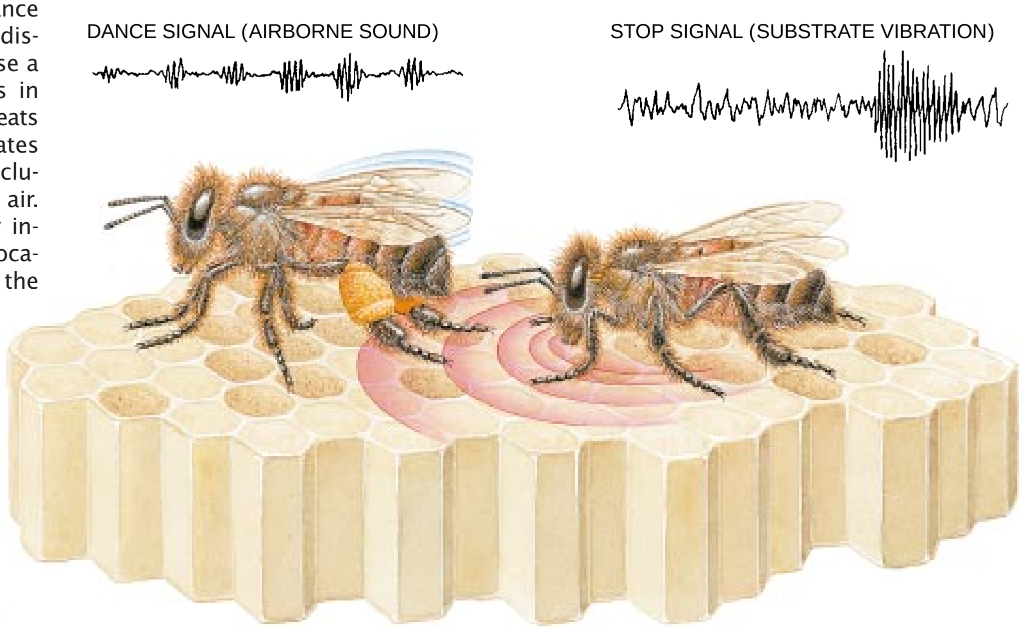
A number of additional experiments using the model bee followed. These trials were designed to determine how important various aspects of the dances are to the dance followers. For example, in some experiments, the model delivered food samples but did not dance. In this case, far fewer recruits ventured to the targeted bait. Also, the model failed completely to recruit live bees when its metal wings did not vibrate; these silent dances did not work, demonstrating that the sounds are truly an essential part of the honeybee's dance language.

The model bee's ability to recruit its nestmates showed that von Frisch was correct about the communicatory function of the dances. Earlier, however, despite what many researchers felt was compelling evidence supporting von Frisch's theory, some of our colleagues doubted that the bees used the distance and direction data encoded in the dances. Wenner and Patrick H. Wells of Occidental College and others have argued persistently that the coordinates given in the dance represent correlations only and are not signals. They believe recruits depend solely on odors to find feeding sites.

Von Frisch himself first attempted to test the significance of the dance movements. He found that recruits could no longer find the correct food source

The Sounds of the Dance Language

A forager emits dance sounds that are distinguishable from those a dance follower makes in the hive. The forager beats her wings and generates sounds that travel exclusively through the air. These sounds convey information about the location of food outside the hive. In contrast, a dance follower produces sounds by pressing her thorax against the comb; this action vibrates the comb and causes the forager to stop her dance and proceed to dole out food samples.



when he laid their hive horizontally on its side. The maneuver prevented the dancer from using gravity to orient the direction of her waggling run. As a result, the dance attenders could not interpret her actions correctly.

James L. Gould of Princeton University later punctured the odor hypothesis. He showed that a forager can dispatch her nestmates to a site she has never visited. Such a feat would be impossible if the recruits relied on odors alone to track down a feeding site. The event could occur, however, if the searching bees gave priority to the information they received from the dance. In these experiments, Gould placed a bright light in the hive, which the dance followers mistook for the sun. In doing so, they interpreted the dances erroneously. These misdirected bees most often searched in the field using the misaligned dance information and seemed to ignore other cues such as odor. Gould concluded that they evidently preferred the message given in the dance to the other signals. Finally, the experi-

DANCING ROBOT (*left*) successfully recruited its nestmates to food away from the hive. The experimenters placed eight baits around the hive and programmed the robot to dance concerning one site. Observers in the field recorded the approach of searching bees. Most of the robot's recruits went to the bait indicated by its dance (*two trials shown at right*).

ments using the model bee confirmed von Frisch's hypothesis; the dances do indeed represent a sophisticated form of communication.

The model bee has helped us answer other questions raised by von Frisch's observations, such as determining which components of the dance language represent what kinds of instructions. For example, when the model danced so that her waggling run appeared on the outside of the figure-eight path, the recruits followed the direction indicated in the waggling run, rather than that given by the orientation of the figure eight as a whole. Thus, the waggling run alone, during which the sounds are produced, tells the recruits the direction in which they can find the food.

Although the experiments using both the model bee and the diminutive-wing bees confirmed that the dance sounds are an important part of the dance language, a crucial element of the picture was absent: the identity of the structures through which bees hear airborne sounds. Several previous attempts made to resolve this issue had shown that bees seemingly could not hear at all. Yet because we now knew that the dance sounds traveled exclusively through the air, we felt inspired to reinvestigate the question. In these renewed trials, we tested the bees using sounds very similar to those that dancing bees make.

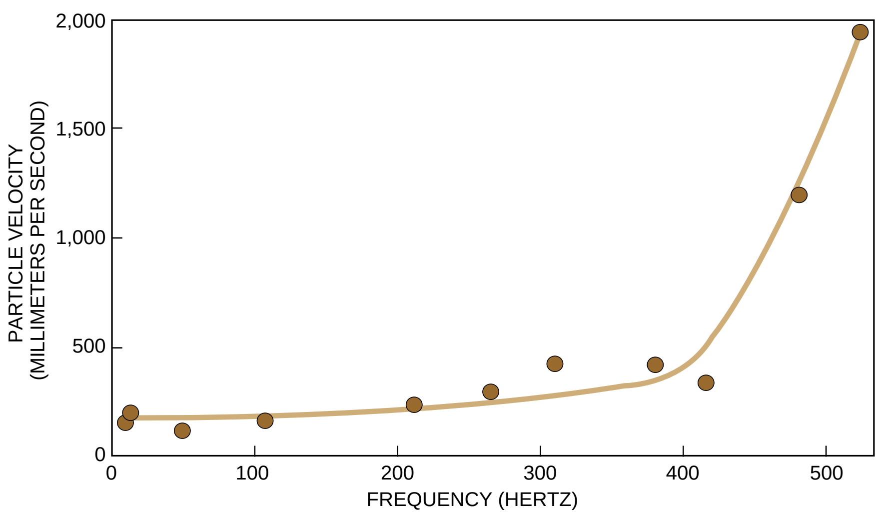
In our first series of experiments, we trained the bees to associate a sound, lasting for five seconds, with a very mild electric shock, arriving four seconds after the sound had started. We generated the sound at the open end of a narrow glass tube. The shock alone, if delivered while the bee was feeding, would drive the bee from the feeder for a few seconds; shortly thereafter she would return and continue feeding. We then posed the following question: If a bee repeatedly experiences a tone followed by an adverse stimulus, will she eventually learn to withdraw from the feeder within the first four seconds of the sound, before the shock? If so, then she can hear—and of course learn. We found that bees can indeed be trained to respond to airborne sounds, although they learned to do so very slowly.

More recently, we have employed a different training technique. In these experiments, a bee entered a very simple Y-shaped maze. We played a sound at one end of this two-sided feeder. The side from which the sound came changed unpredictably from one trial to next. If the bee turned toward the sound, she received a reward of sugar water; if she went away from the sound, she received nothing. We observed that the bees learned quickly to turn toward the sound. Claudia Dreller, a graduate student at Würzburg, used the procedure to explore the frequency and amplitude range in which the bees could hear. Dreller's work showed that honeybees sense only low frequencies, those below 500 hertz. They hear these tones with sufficient sensitivity to pick up the sounds of a dancing nestmate, which range from 250 to 300 hertz. They also show some ability to discriminate between frequencies in this range; they can discriminate between low- (20 hertz), medium- (100 hertz) and high-pitched (320 hertz) sounds. We do not yet know for what purpose the bees might use this latter ability.

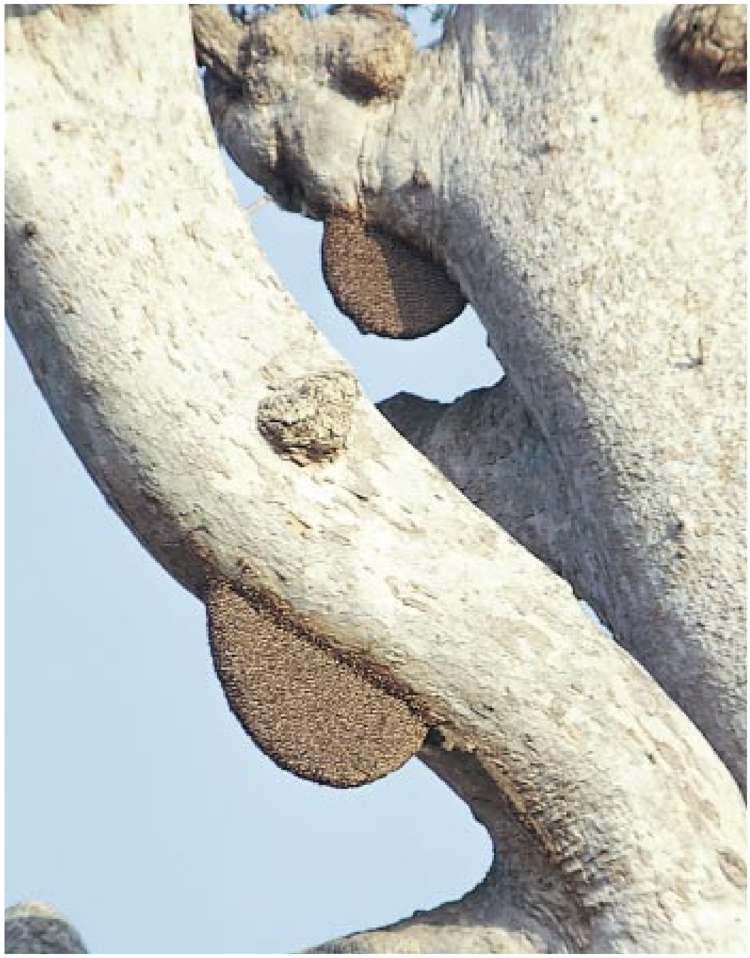
The same training technique enabled us to find out through what sensory structures the bees detect the sounds. We altered some of our trained bees by removing one antenna, fixing a certain antennal joint or removing sensory hairs from their head. We found that bees use a structure called the Johnston's organ, a chordotonal organ made up of nerve cells in the second joint of a bee's antennae, to pick up airborne sounds. Some flies and mosquitoes rely on the same structure to perceive sounds.

If we now put the pieces together, we can see how the dance language works. The dancer emits sound signals that help the dance followers determine where the dancer is and how she is moving, which in turn offers them critical information regarding the direction and distance to the feeding site. The dance attenders receive these signals through the Johnston's organs located in their antennae, which are always held near the dancer. Because these organs are bilateral—one on the left and one on the right—the dance followers can use them to judge their position with respect to the dancer and therefore understand the direction to the food. At the same time, the followers emit sounds that vibrate the comb. The forager stops her dance when she receives these signals and delivers samples of the food she has collected. These appetizers give the dance followers additional hints about the taste, smell and quality of the food source. The bees attend the dancing for a while and then fly out to find the food source on their own. If they are lucky, they will find the food. If they fail, they will return to the nest and try again.

This dance language is clearly a very complicated, highly developed system of recruitment. To understand how such a system evolved, we and other scientists have examined the recruiting techniques of related species. The genus *Apis,* to which all honeybees belong, has no close living relatives. Bumblebees and stingless bees are their nearest kin, and, unfortunately, it seems that bumblebees do not recruit at all. Many species of stingless bees, on the other hand, do recruit. But, as far as we know, none has developed a symbolic language similar to the dance language of *Apis.* All four species of honeybees studied so far (three of which live in Asia) speak some variant of the dance language. As Martin Lindauer discovered in the 1950s, when he was at the University of Munich, all species use similar distance and direction codes, even though there are some differences.



FREQUENCY RANGE of the sounds a bee can detect extends well below the range heard by human ears. The graph shows how fast the air particles near a dancer's wings must travel to generate audible signals. Within this range, the bees show an ability to differentiate between sounds having varying frequencies.



NESTING HABITS may have influenced the evolution of the dance language, as it is spoken by different species of honeybee. *Apis florea* (*bottom left*), considered the most primitive in behavior, does not produce dance sounds. These bees nest in the open and dance only during the day. *A. dorsata* (*above*) also nests in the open but sometimes dances at night. *A. dorsata* makes dance sounds, as do *A. cerana* (*top left*) and the familiar western bee, which usually dance in dark nests.

Although we have looked for dance sounds in four species of honeybees, we have found that only three produce them. These three species hold something else in common: they all must occasionally dance in the dark. Two of them, the familiar western bee, *A. mellifera,* and the Asian bee, *A. cerana,* nest in lightless, enclosed areas such as hollow trees or other similar cavities. The third sound producer, the giant bee, *A. dorsata,* nests in the open on single sheets of comb, hanging under rock outcrops or thick branches of trees. Fred C. Dyer of Michigan State University first showed that *A. dorsata* sometimes dances at night, and Kirchner discovered only recently that this bee produces sounds. *A. dorsata's* signals were very difficult to detect because these sounds are particularly low in pitch.

The single species that dances silently, the dwarf bee, *A. florea,* dances in the open like *A. dorsata,* but only during the day. Dancers of this species make gestures that may, in daylight, serve as visual signals to attract dance followers in the same way that sounds assist those bees who dance in the dark. Because some indications suggest that *A. florea's* habits are the most primitive, we assume that the complicated acoustical communication system of the other three species most likely evolved from a visual display when these bees developed habitations that cut them off from light.

Now that we can finally listen to the bees' language and even speak it a little, we face a host of new questions. For example, we have yet to learn for what purpose the bees possess the ability to distinguish between different pitched sounds. In addition, perhaps they use similar airborne sounds in ways that we do not even suspect at this time. In the hopes of revealing their communication system in full, we will continue to eavesdrop on their conversations.

FURTHER READING

THE DANCE LANGUAGE AND ORIENTATION OF BEES. K. von Frisch. Harvard University Press, 1967.

COMMUNICATION AMONG SOCIAL BEES. Martin Lindauer. Harvard University Press, 1971.

THE HONEY BEE. James L. Gould and Carol G. Gould. Scientific American Library, 1988.

ACOUSTICAL COMMUNICATION IN HONEYBEES. W. H. Kirchner in *Apidologie,* Vol. 24, No. 3, pages 297–307; July 1993.

The Ethnobotanical Approach to Drug Discovery

Medicinal plants discovered by traditional societies are proving to be an important source of potentially therapeutic drugs

by Paul Alan Cox and Michael J. Balick

In 1785 William Withering, a British physician, reported that ingestion of dried leaves from the foxglove plant eased dropsy, an accumulation of fluid now known to be caused by the heart's failure to pump adequately. Withering credited an unexpected source for his information. "I was told," he wrote, that this use of foxglove (a member of the genus *Digitalis*) "had long been kept a secret by an old woman in Shropshire, who had sometimes made cures after the more regular practitioners had failed."

Digitalis has been helping cardiac patients ever since. Today two of its components—the glycosides digoxin and digitoxin—are prescribed to hundreds of thousands of people throughout the world every year. Indeed, these glycosides currently serve as the treatments of choice for rapid atrial fibrillation, a dangerous cardiac irregularity. Given the importance of *Digitalis,* we have no doubt that many readers of *Scientific American* are alive in 1994 because Withering investigated the secret remedy of "an old woman in Shropshire."

PAUL ALAN COX and MICHAEL J. BALICK met in the late 1970s, when they were doctoral students at Harvard University. Cox is dean of general education and honors and professor of botany at Brigham Young University. He is also president of the Society for Economic Botany. Balick is Philecology Curator of Economic Botany and director of the Institute for Economic Botany of the New York Botanical Garden in Bronx, N.Y. He is also past president of the Society for Economic Botany. Cox and Balick, who are collaborating on a book about ethnobotany for W. H. Freeman and Company, serve as advisers to a variety of governmental, academic and industrial research groups and foundations.

A decade ago this story would probably have been regarded as nothing more than a historical anecdote, of little relevance to contemporary drug discovery. By the mid-1980s most pharmaceutical manufacturers had abandoned exploring folk uses of plants in their search for new drugs. Now, however, the pendulum is beginning to swing back toward an appreciation that plants used in traditional medicine can serve as a source of novel therapeutic agents.

Such appreciation has emerged in part because of recent discoveries made by a small but growing group of ethnobotanists—researchers who study the relationships between plants and people. Fieldwork exploring the medicinal uses of plants by indigenous peoples in remote parts of the world, coupled with the introduction of sophisticated assays able to determine whether plants exert a biological effect, has facilitated the discovery of bioactive molecules made by medicinal plants. Some of these molecules show promise as possible therapies for a range of diseases, including AIDS and cancer.

In the U.S. the drug development process is a long and arduous one, designed to ensure that therapies released to market are effective and safe. It can thus easily take many years for a substance to become commercially available as a drug. There seems little doubt, however, that within a decade several new agents derived from ethnobotanical research will be introduced. We cannot claim that most drugs of the future will be found in this way. Yet the strategy, as will be seen, has many merits.

Until the 1950s, almost all pharmaceutical research relied heavily on vascular plants as sources of medicines. Flowering plants and ferns (as opposed to microscopic organisms and fungi) have given rise to about 120 commercially sold drugs and account for some 25 percent of all prescriptions issued every year in North America. Many of these agents are now synthesized in the laboratory, but others are still isolated from plants. Most were discovered by studying indigenous uses of plants.

For instance, the drug reserpine, which is still occasionally prescribed in the U.S. for hypertension, was isolated from the root of the climbing shrub *Rauvolfia serpentina* (Indian snakeroot) after scientists began analyzing Ayurvedic remedies—the traditional treatments used by the peoples of India. Other examples include aspirin, opiates, pilocarpine (prescribed for glaucoma and dry-mouth syndrome) and two cancer treatments—vincristine and vinblastine. Vincristine and vinblastine, both of which are still extracted from *Catharanthus roseus* (rosy periwinkle), have been prescribed for pediatric leukemia and Hodgkin's disease, respectively, since the 1960s.

Plants have been a rich source of medicines because they produce a host of bioactive molecules, most of which probably evolved as chemical defenses against predation or infection. Nevertheless, several forces conspired by the close of the 1970s to cause plants to lose much of their appeal as drug sources for the pharmaceutical industry. Microorganisms and fungi that inhabit soil, which are easy to collect and culture, had provided a dazzling array of antibiotics. Advances in synthetic chemistry and molecular biology promised to supply new means for designing drugs in the laboratory. And few major discoveries of plant-derived drugs had followed the identification of vincristine and vinblastine. Given these conditions, many pharmaceutical firms simply stopped searching for therapeutic compounds in higher plants.

In spite of this discouraging state

of affairs, in the late 1970s and early 1980s several groups of scientists independently set off to different corners of the world for the express purpose of finding innovative drugs through ethnobotanical research. Some of the researchers, such as Gunnar Samuelsson and Lars Bohlin of the University of Uppsala, were former students of the distinguished Swedish ethnobotanist Finn Sandberg. A number of American researchers trace much of their inspiration to another leading ethnobotanist, Richard Evans Schultes of Harvard University.

The two of us are among that latter group. When we were graduate students at Harvard in the late 1970s, Schultes encouraged us to continue our studies of traditional uses of plants in remote regions of the earth. Balick, in his doctoral work, extended Schultes's pioneering ethnobotanical research in the Amazon. Meanwhile Cox, who had been a missionary in Samoa for the Mormon church during his undergraduate years, returned to the South Pacific to study the ecology of the rain forest and the uses of plants by the Samoan people. After receiving our degrees, we continued to investigate herbal medicine in two culturally and geographically distinct areas: Central America (Balick) and Polynesia (Cox). Both of us have studied intensively with indigenous healers.

The ethnobotanical approach is actually one of several methods that can be applied in choosing plants for pharmacological studies. It is estimated that 265,000 flowering species grace the earth. Of these, less than half of 1 percent have been studied exhaustively for their chemical composition and medicinal value. In a world with limited financial resources, it is impossible to screen each of the remaining species for biological activity. Some kind of collection strategy is needed.

Investigators, for example, can gather vegetation randomly in an area supporting rich biological diversity. Unfortunately, random searches yield relatively few new drug possibilities. One notable exception is taxol. In 1992 the Food and Drug Administration approved taxol, derived from *Taxus brevifolia* (the Pacific yew tree), as a treatment for ovarian cancer and in 1994 approved it for treating metastatic breast cancer unresponsive to other therapies. Taxol was found in the course of a random-screening program conducted by the National Cancer Institute (NCI), which has maintained a plant-screening program, with varying degrees of energy, since 1960. (Today the NCI employs several plant-gathering strategies. It also examines specimens not only for their anticancer effects but also for their ability to impede the functioning of the human immunodeficiency virus, or HIV, which causes AIDS.)

Other plant-collecting methods, including the ethnobotanical approach,

ANDREW RAMCHARAN, a healer living in Belize, Central America, is collecting the leaves and flowers of the plant *Cornutia pyramidata* to include in a mixture used for skin rashes. Ethnobotanists interested in drug discovery often rely on healers to identify plants that are likely to contain potent bioactive chemicals. There is some urgency to this work: many healers are elderly and lack apprentices. As they die, much of their knowledge of local vegetation dies, too.

Drugs Discovered from Ethnobotanical Leads

The well-established drugs listed below are among dozens that were developed after scientists began to analyze the chemical constituents of plants used by traditional peoples for medicinal or other biological effects. For instance, Western researchers isolated reserpine in 1952 from the climbing shrub *Rauvolfia serpentina* (*photograph*), which has been employed in India for many centuries to treat snakebite and mental illness.

DRUG	MEDICAL USE	PLANT SOURCE
Aspirin	Reduces pain and inflammation	*Filipendula ulmaria*
Codeine	Eases pain; suppresses coughing	*Papaver somniferum*
Ipecac	Induces vomiting	*Psychotria ipecacuanha*
Pilocarpine	Reduces pressure in the eye	*Pilocarpus jaborandi*
Pseudoephedrine	Reduces nasal congestion	*Ephedra sinica*
Quinine	Combats malaria	*Cinchona pubescens*
Reserpine	Lowers blood pressure	*Rauvolfia serpentina*
Scopolamine	Eases motion sickness	*Datura stramonium*
Theophylline	Opens bronchial passages	*Camellia sinensis*
Vinblastine	Combats Hodgkin's disease	*Catharanthus roseus*

Rauvolfia serpentina
INDIAN SNAKEROOT

Filipendula ulmaria
MEADOWSWEET

Papaver somniferum
OPIUM POPPY

Datura stramonium
JIMSON WEED

Cinchona pubescens
FEVER TREE

Catharanthus roseus
ROSY PERIWINKLE

are more targeted. In phylogenetic surveys, researchers choose close relatives of plants known to produce useful compounds. In ecological surveys, they select plants that live in particular habitats or that display characteristics indicating they produce molecules capable of exerting an effect on animals. Collectors might, for example, focus on specimens that seem to be immune from predation by insects. The absence of predation suggests a plant may produce toxic chemicals. Many chemicals that are toxic to insects also show bioactivity in humans, which means they might be capable of achieving some therapeutic effect.

Finally, the ethnobotanical approach assumes that the indigenous uses of plants can offer strong clues to the biological activities of those plants. For instance, if our colleagues at the NCI or at pharmaceutical firms asked us to focus on collecting plants that might serve as therapies for HIV, we would pay special attention to plants a society uses against diseases we know are caused by viruses. Plants exploited for their poisonous effects are also of interest. Blowgun poisons such as curare have in the past been found to contain compounds able to serve as anesthetics.

The history of drug discovery implies that the ethnobotanical approach is the most productive of the plant-surveying methods, and recent findings confirm that impression. Some of the data were collected by Cox with his student Rebecca Sperry and with Bohlin and Mervi Tuominen of the University of Uppsala. This group found that 86 percent of the plants used by Samoan healers display significant biological activity in a variety of assays. Balick and Rosita Arvigo of the Ix Chel Tropical Research Foundation in Belize discovered that crude extracts of the plants that one healer (Don Elijio Panti, in Belize) considered to be his most powerful gave rise to four times as many positive results in a preliminary laboratory test for activity against HIV than did specimens collected randomly.

We should note, however, that few of the compounds exhibiting activity in laboratory tests will become new drugs. Some will turn out to be identical to, or less potent than, existing agents; others will prove too toxic for commercial use. Nevertheless, demonstrating activity in a bioassay is a necessary first step in the drug development process.

How do ethnobotanists choose the societies they study? Researchers differ in their criteria. Because not all cultures are equally likely to make use of plants having significant pharmaceutical activity, the two of us focus our efforts on those possessing three characteristics. First, the societies should be located in a floristically diverse area, such as a tropical rain forest. Such diversity dramatically increases the number of plants available; it thus enhances the likelihood that plants with pharmacologically active molecules will be pressed into service.

Second, the societies should have remained in the region for many generations. Groups who have resided in one place for a long time presumably have had ample opportunity to explore and experiment with local vegetation. According to this requirement, aboriginal peoples who have populated Australia for many thousands of years would be a better choice for study than European settlers.

Third, the cultures must have a tradition in which healers transmit their plant knowledge from generation to generation, usually through apprentices. Consistent application of a given species for an ailment over millennia generates information rather analogous to that produced by large-scale clinical trials. Such repetitive, long-term use of botanical species can be expected to have identified both the most effective medicinal plants and those that are too toxic for use.

As these criteria suggest, we are especially interested in recording the specialized information possessed by healers; that is, knowledge not necessarily shared by most other individuals in the community. Many members of indigenous societies can identify some common plants that have medicinal uses, much as many Europeans realize that *Aloe vera* is helpful for burns. But only a healer might know the best plant to choose for treating serious or uncommon diseases, such as yellow fever.

Other ethnobotanists—notably, Brent Berlin of the University of California at Berkeley and Walter H. Lewis and Memory Elvin-Lewis of Washington University—prefer to study plants used by a broad array of villagers. This approach, called the consensus technique, is based on the supposition that awareness of highly effective medicinal plants will spread rapidly throughout a culture. We concentrate on seeking esoteric information because many healers no longer have apprentices. As healers die, their arcane wisdom is likely to be extinguished forever. Botanical insights that are widely held should be accessible longer.

Interestingly, healers often practice both as generalists and specialists. In Belize, caregivers provide some level of primary health service to residents of their villages and surrounding regions, yet they often also have unique areas of expertise. Hortense Robinson of Ladyville specializes in midwifery and other health care concerns of women and children. Andrew Ramcharan of Ranchito focuses on treating snakebite in an area that harbors many venomous snakes. The knowledge such healers possess can be quite remarkable. In the islands of Samoa, most healers are female and specialize in herbal medicine; only bone setters are commonly male. The herbalists, or *taulasea,* typically make use of more than 100 species of flowering plants and ferns.

We find healers mainly by asking people to identify the individuals to whom they turn when they are sick. Having located healers, we take pains to explain our mission to them and to tribal chiefs. This process is analogous to "informed consent" in clinical settings. The interactions go most smoothly if Western researchers are trained in both botany and anthropology and have a good understanding of the societies they visit. If the healers are amenable to teaching us, we spend many weeks, months or even years with them and take careful notes on the properties and uses of the plants they show us. During such fieldwork, we attempt to live in accordance with local customs, and we usually learn the local language.

When, with the guidance of the healers, we have decided on the plants we want to collect, we gather a kilogram (about two pounds) of each species. (Prior to collection, we always obtain permission from the healers, village chiefs, landowners and governments.) Whenever possible, we collaborate with local scientists or students. We save four or five samples of a plant to serve as "voucher," or reference, specimens. These specimens are pressed flat, dried and labeled carefully. The labels indicate species name and indigenous name and application, as well as any directions needed for locating the plant again. Eventually the specimens are deposited in herbaria (collections of preserved plants) in several parts of the world for consultation by other botanists.

We then prepare the rest of the material for transport to our laboratories. Typically we dry the plants or preserve them in aqueous alcohol. Later, in the laboratory, we or our colleagues usually extract different kinds of molecules from the plants by immersing them in various solvents. Next, we separate out the solvent and deliver the resulting solvent-free extracts (or sometimes the plant samples themselves) to collaborators at other universities, governmental agencies or drugmakers for screening.

In the 1960s standard screening assays for many potential drugs consisted of injecting test material into a rodent and waiting to see if the animal fell ill, got well or displayed some other change in behavior or health. Screening was thus a time-consuming, costly and imprecise endeavor. Thirty years later bioassays are faster (usually being automated) and are significantly more specific. At the NCI, tiny amounts of material can be screened rapidly against an array of up to 60 distinct human tumor cell lines. Many other assays assess the ability of an extract to influence the activity of a single enzyme involved in the biochemical interactions that underlie a disease. For example, scientists looking for an AIDS treatment might evaluate whether the extracts inhibit the activity of the enzyme reverse transcriptase in cells. Reverse transcriptase is coded for by the genetic material of HIV and is needed for replication of the viral particles.

When an extract displays significant activity in a bioassay, we return to the field with our local collaborators and retrieve a bulk sample of the original plant, typically 50 to 100 kilograms' worth. Chemists need these extra supplies in order to isolate the molecule responsible for the observed activity and to determine (by spectroscopic techniques) its structure. Newly emerging techniques that require very little plant material may render such return collection trips unnecessary.

Once the molecular structure is identified, researchers compare it with that of known chemicals. If the isolated molecule is novel, or has already been found but has not been studied as a possible drug, it may be analyzed further as is. In other cases, a synthetic version will be examined instead—that is, if workers can determine how to con-

ETHNOBOTANIST MICHAEL J. BALICK, one of the authors, is drying plants that were collected by his colleagues in Belize. Drying is one of the techniques by which investigators preserve their specimens in the field.

struct the substance and can do so at a reasonable cost. Still other times, the molecule will serve as raw material that is altered to produce some desired activity. Even if an isolated molecule or a synthetic version clearly cannot serve as a drug for some reason, its discovery might suggest previously unconsidered avenues for attacking a disease.

After investigators settle on a test molecule having significant activity—also known as a lead compound—evaluation proceeds as it would for any other potential drug. Researchers study the compound for such characteristics as its strength of binding to a particular target molecule and its toxicity to cells. If it passes these tests, a company or governmental agency may decide to invest the money needed (an estimated $200 million) to bring it to market. At that point, it becomes a drug candidate. To make its way to pharmacy shelves, the drug candidate must prove its worth in additional laboratory analyses and in clinical trials.

So far the rigorous ethnobotanical field searches that have been conducted since the early 1980s have generated many lead compounds. Some of these have been identified as drug candidates, and a few may eventually reach pharmacy shelves. We cannot be sure of the exact number of leads or drug candidates, because most pharmaceutical makers will not discuss products in the early stages of development. We do know, however, that many of the lead compounds derived from ethnobotanical research exhibit potent antiviral, antifungal or anticancer effects—properties that are sorely needed in the pharmaceutical arsenal.

Two drug candidates that have reached clinical trials are being developed by Shaman Pharmaceuticals in South San Francisco, Calif. The active component in both drugs is derived from a plant that grows in Central and South America. One formulation shows promise against respiratory viruses. The second version may be administered topically for treating infections caused by the herpes simplex virus.

Another exciting agent in the drug pipeline is a powerful antiviral compound called prostratin. The story of prostratin's identification serves as a good paradigm of the drug discovery process. When Cox was in Samoa in 1984, several healers told him that they treat yellow fever by giving patients brews made by steeping in water pulverized wood of the rain-forest tree *Homalanthus nutans.* Cox collected samples and brought them to the U.S., where he prepared freeze-dried solvent extracts. Because of the prevailing attitudes at the time, no drug companies were inclined to analyze his specimens.

Fortunately, Michael R. Boyd and Gordon M. Cragg of the NCI agreed to test them. The extracts exhibited strong activity against HIV in the test tube. Chemists John H. Cardellina II and John A. Beutler, also at the NCI, then discovered that the active component was a chemical called prostratin. The chemical was known but was not being studied as a drug, because it was a phorbol, a class of chemicals that promotes the development of tumors.

Worry over tumor promotion initially dampened interest in prostratin at the NCI. But Cox argued that if this particular phorbol stimulated the growth of cancer, its carcinogenic properties would have become evident in Samoa; its continued use by healers suggested it was safe. When Peter M. Blumberg, a cell biologist at the NCI, examined the drug's carcinogenic effect in mice, he found, to his co-workers' surprise, that it did not promote tumor formation, even though it activated an enzyme (protein kinase C) that participates in malignancy. The NCI is now planning to accept bids from drug companies for the right to investigate prostratin further as a possible drug.

An additional compound identified through Cox's ethnobotanical research highlights the striking precision of the knowledge possessed by many healers. Samoan *taulasea* apply the bark of a local tree, *Erythrina variegata,* to the skin to treat inflammation. In describing this tree, they insisted that only one of the two types of *E. variegata* in the region was helpful. Sure enough, when a team lead by Vinod R. Hegde and Mahesh G. Patel of Schering-Plough Corporation in Kenilworth, N.J., evaluated the ability of bark samples to inhibit an enzyme involved in inflammation (phospholipase A_2), they found the healers were absolutely correct. Bark from only one variety of *E. variegata* displayed anti-inflammatory activity. From that bark, the team isolated the active component—a chemical known as a flavanone. The chemical is now under development as a topical anti-inflammatory at Schering and at Phyton Catalytic in Ithaca, N.Y.

In Thailand a team led by Hans T. Beck of the New York Botanical Garden and Weerachai Nanakorn, then of the Royal Forest Department in Bangkok, learned from healers that the roots of *Curcuma comosa,* a relative of the ginger plant, eased stomach pains and other gastrointestinal disorders. Tannis Jurgens and colleagues at Merck & Co., in Rahway, N.J., then isolated from the roots a novel compound that kills parasitic worms in the stomach. The compound is now under investigation as a potential treatment for parasitic diseases. In Peru, Lewis and Elvin-Lewis obtained samples of a tree sap that Jivaro Indians and others apply for hastening the healing of wounds. The active ingredient in the sap, called taspine, is being tested for that same purpose in animals. If those tests are successful, human trials are likely to follow.

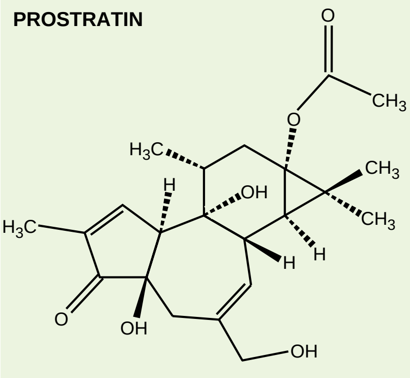
The probability that drug companies, universities and Western scientists will profit financially from information provided by healers in traditional societies raises a serious question: What is being done to protect the interests of healers and their communities? We are well aware that the healers with whom we work provide significant intellectual guidance. Indeed, we refer to them as "colleagues," "guides" and "teachers," instead of as "informants." Given the significant participation of indigenous peoples in our research programs, we believe they are entitled to the same intellectual property rights enjoyed by other investigators. Hence, we do all we can to see that those rights are protected.

For instance, in the case of prostratin, the NCI and Brigham Young University, where Cox is a professor, have guaranteed that a significant part of any royalties earned from the drug will be returned to the Samoan people. Virtually all ethnobotanists active in drug research are involved in making similar arrangements for the cultures they study. In fact, now that ethnobotanical inquiry is expanding, formal guidelines are being devised.

To many cultures, protecting the forests around them is more important than receiving money. Hence, many researchers are devoting effort to protecting the rain forests in the regions where they work. For example, Thomas Eisner of Cornell University helped to convince Merck & Co. to invest $1 million in a project designed to inventory and ultimately preserve a part of the Costa Rican rain forest. In return, Merck can study chemicals extracted from living organisms in Costa Rica's national parks and elsewhere.

Four village-owned and village-managed reserves encompassing 64,000 acres have now been formed in Samoa with funds raised through the Seacology Foundation of Provo, Utah (an organization Cox helped to create). These protected territories include the Falealupo Rain Forest Reserve, the site where the plant that produces prostratin was

PROSTRATIN



SAMOAN HEALER MARIANA LILO, who died in 1993, is shown cutting the stem of *Homalanthus nutans.* She went on to macerate the wood and steep it in hot water, producing a brew to be drunk by patients with yellow fever (a viral disease). The active ingredient, prostratin (*above*), was isolated by the National Cancer Institute, which plans to license it to a drug company for investigation as a possible treatment for HIV infection. Prostratin is one of a range of potential drugs that have been discovered through recent ethnobotanical research.

first collected. In the U.S. territory of American Samoa, Cox participated in establishing the 50th national park of this nation. By mandate of Congress, the park will allow and encourage Samoan healers to continue to use medicinal plants in a sustainable way—that is, in a way that guarantees their ongoing availability.

Balick and Arvigo and their colleagues Gregory Shropshire of the Ix Chel Tropical Research Foundation and Leopoldo Romero of the Belize Association of Traditional Healers recently helped to establish the world's first ethnobiomedical forest reserve. It is an entity designed specifically to ensure that medicinal plants will be available for local use. The protected area, called the Terra Nova Rain Forest Reserve, comprising almost 6,000 acres of forest in the Yalbak region, is to be managed by the Association of Traditional Healers. The reserve was created not only to provide needed medicinal plants but also to teach young people about their uses. It is hoped that such teaching will encourage youth to preserve the knowledge of their elders. Researchers will also be collaborating with the healers to devise ways to make sure that medicinal plants are harvested sustainably.

In spite of its apparent successes, ethnobotany is unlikely to ever become a major force behind commercial drug discovery programs. Its application is limited by the paucity of properly trained ethnobotanists who have the time to conduct rigorous, long-term fieldwork in remote areas of the world. Also, many funders of drug research still perceive the ethnobotanical approach as archaic, unscientific and unworthy of attention.

Yet the demonstrated ability of ethnobotany to generate exciting leads for drugs suggests to us that, for the near future at least, the approach will occupy an expanding role in drug development. Those who understand the value of such work are truly in a race against time. In Samoa, two healers who first provided Cox with information leading to the discovery of prostratin—Epenesa Mauigoa and Mariana Lilo—died in 1993. Generations of accumulated medical wisdom died with them. Ethnobotanists can capture much of the remaining knowledge, but only if the research effort is expanded soon. Sadly, plant knowledge seems to be disappearing even faster than the forests themselves.

FURTHER READING

ISLANDS, PLANTS, AND POLYNESIANS: AN INTRODUCTION TO POLYNESIAN ETHNOBOTANY. Edited by Paul A. Cox and Sandra A. Banack. Dioscorides Press, 1991.

A NONPROMOTING PHORBOL FROM THE SAMOAN MEDICINAL PLANT *HOMALANTHUS NUTANS* INHIBITS CELL KILLING BY HIV-1. K. R. Gustafson et al. in *Journal of Medicinal Chemistry,* Vol. 35, No. 11, pages 1978–1986; May 29, 1992.

RAINFOREST REMEDIES: 100 HEALING HERBS OF BELIZE. R. Arvigo and M. Balick. Lotus Press, Twin Lakes, Wis., 1993.

THE DEVELOPMENT OF AN ETHNOBIOMEDICAL FOREST RESERVE IN BELIZE: ITS ROLE IN THE PRESERVATION OF BIOLOGICAL AND CULTURAL DIVERSITY. M. J. Balick, R. Arvigo and L. Romero in *Conservation Biology,* Vol. 8, No. 1, pages 316–317; March 1994.

Grading the Gene Tests

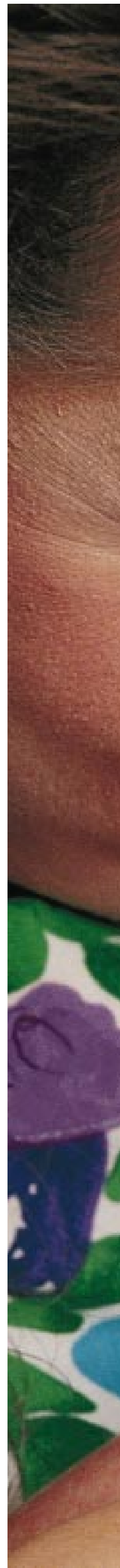
by John Rennie, *staff writer*

Almost nine months before she was born, Brittany Nicole Abshire passed the most important test she will ever take. Her parents, Renee and David, are both healthy carriers of the trait for Tay-Sachs disease, a cruelly disabling and ultimately lethal inheritable disorder. After they lost one daughter to Tay-Sachs in 1989, they swore they would never have another child unless they could be sure that it would be free of the disease. Genetic tests could diagnose the condition before birth, but the Abshires' religious beliefs ruled out abortion as a way of screening for healthy fetuses.

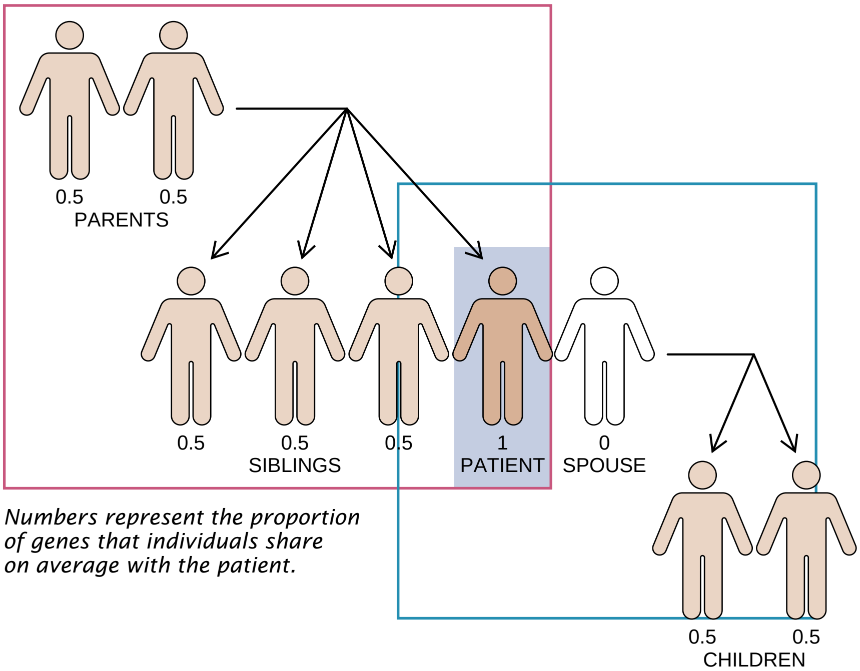
There seemed to be no hope until the Abshires learned about a new technology called preimplantation genetic testing. The experimental procedure had already been used to screen more than a dozen children for a different genetic disorder, cystic fibrosis. Gary D. Hodgen and specialists at the Jones Institute for Reproductive Medicine at the Eastern Virginia Medical School collected ova and sperm from the Abshires and successfully fertilized seven ova in vitro. After three days, when those seven had developed to about the eight-cell stage, Hodgen's team plucked a cell from each pre-embryo and tried to analyze its DNA.

For four of the pre-embryos, the analysis worked: one of them showed the deadly combination of genes, but three were not even carriers. Those three pre-embryos were implanted in Renee, and one survived to become Brittany, who was born this past January. Courtesy of genetic testing, Brittany is the first child ever certified to be free of Tay-

EIGHT-CELL EMBRYOS like the one shown at the left can now be screened for the presence of some genetic illnesses. Baby Brittany Nicole Abshire (right) *was tested in this way because both of her parents are carriers of the fatal Tay-Sachs disease gene.*



From just a snippet of DNA, geneticists can sometimes forecast a patient's health. But ethical problems surrounding this testing are as ominous as the diseases themselves



GENES ARE SHARED by members of a family. If one person carries a gene for a disease, then each of his or her parents, siblings and children has a 50 percent chance of carrying the same gene. That fact affects the privacy of genetic data.

Sachs disease before entering her mother's womb.

As the era of genetic testing dawns, miracles such as Brittany could become commonplace. Genetic testing is the fastest-growing area in medical diagnostics: according to the Office of Technology Assessment (OTA), the number of genetic tests will increase 10-fold over the next decade. "Potential new genetic tests roll off the conveyor belt of the Human Genome Project almost once a week," remarks Norman Fost of the University of Wisconsin–Madison Medical School. Hundreds of thousands of fetuses are already being tested every year by techniques such as amniocentesis and chorionic villus sampling.

Tests are not just for the unborn: many can also be used to diagnose illnesses more accurately in children and adults. In the past year alone, researchers have found genes associated with Alzheimer's disease, Huntington's disease and colon cancer; they expect to find a breast cancer gene almost any day now. Tests based on those and other discoveries could warn people that they are at special risk for those diseases. And used in conjunction with prospective therapies that replace defective genes with working ones, genetic tests could lead to real cures.

But many human geneticists and other observers are concerned that the rapid growth of genetic testing is already posing ethical, legal, social and scientific quandaries for which there will be no easy answers. They fear that new tests are proliferating without adequate supervision. Because the genetics of disease is proving to be more complex than anticipated, the predictive power and utility of some tests are in question. Yet states are enthusiastically instituting programs for screening newborns that may be unnecessary. Meanwhile investigators are beginning to see evidence that "genetic discrimination" is costing some people jobs and insurance.

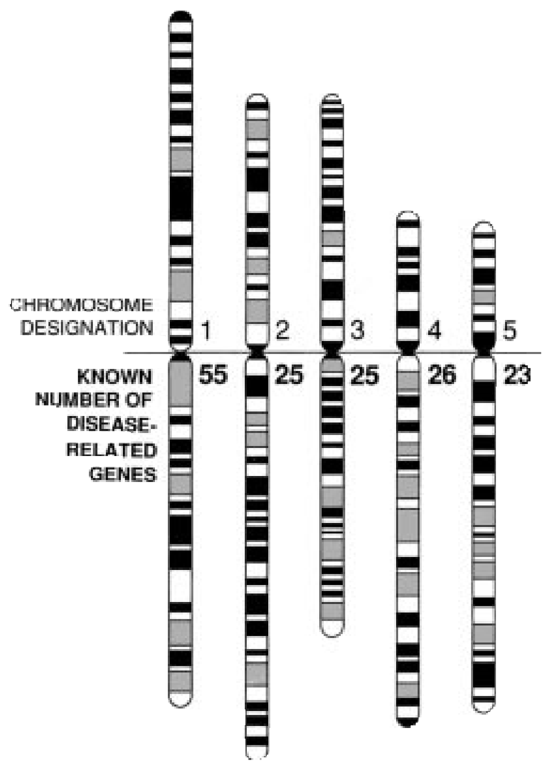
"We need to proceed cautiously, because there is a potential for doing harm with this technology," warns geneticist Michael M. Kaback of the University of California at San Diego, a pioneer in population screening. When we know more about a human's genetic makeup than ever before, will we know what to do with all that information?

A Mythological Model

Genetic testing is not a single technology. Rather it refers to a broad range of methods for gauging the presence, absence or activity of genes in cells. At the relatively low-tech end, researchers can count the chromosomes in a patient's cells or measure the amount of telltale proteins in his or her blood. At the most sophisticated level, researchers assay a cell's DNA with molecular probes that can find a specific genetic sequence among the three billion base pairs that make up human DNA. Some tests cost as little as $50, whereas others are more than $1,000. With these tests, medical geneticists can try to predict the course of a patient's health.

Unfortunately, the more that researchers have learned about human genetics, the more they have come to appreciate that even seemingly straightforward diseases are complicated. Notions of genetic illness have often been built around the single-gene model, in which a defect in a gene causes a particular health deficit. Some diseases do work this way: the deformed blood cells of sickle cell anemia are caused by a gene that makes an abnormal form of hemoglobin; the fatal miseries of Tay-Sachs result from the lack of an enzyme that breaks down fatty substances in neurons. But this model is turning out to be an oversimplification.

No more than about 3 percent of all human diseases are caused by defects in a single gene, and none of those are major killers, as are heart disease and cancer. The more complex conditions involve a host of genes that merely nudge a person's predisposition to develop an illness. According to most estimates, everyone carries at least five to 10 genes that could make him or her sick under the wrong circumstances or could adversely affect children. "We're all mutants," Kaback summariz-



es. “Everybody is genetically defective.”

The truth is that the rules for what constitutes a genetic disease are not clear-cut. If researchers someday find a gene that confers a 60 percent predisposition for gross obesity, is that a genetic defect? What about a gene that gives a 25 percent predisposition for cardiovascular disease at age 55? Or—moving into an even more ambiguous area—a gene that predisposes to antisocial behavior?

Furthermore, even some diseases that once appeared to fit the single-gene model on the basis of their hereditary patterns are more variable than had been assumed. Cystic fibrosis, one of the most common hereditary disorders among people of European descent, is a useful example. Its symptoms include the accumulation of suffocatingly thick mucus in the airways and often severe digestive problems. Today drug treatments allow half of all cystic fibrosis sufferers to live to about age 30, but just a couple of decades ago patients rarely survived into their twenties, and some still die in infancy.

Researchers had often prayed for a genetic test that could find carriers of the disease in the general population. Then, in 1989, investigators at the University of Michigan and the University of Toronto found the gene responsible for cystic fibrosis on chromosome 7; it encodes a protein in cell membranes that affects the intracellular balance of chloride ions. DNA-based tests for mutations commonly associated with the disease soon appeared.

But those tests have revealed a new class of cystic fibrosis patients—people with relatively minor symptoms, such as asthma or bronchitis, who never thought of themselves as genetically ill. Some male patients are perfectly healthy except that they are infertile—for reasons not yet understood, they lack a vas deferens, the tube in the reproductive system that conducts sperm cells from the testicles. Basically, cystic fibrosis is not a single disease after all.

Moreover, molecular biology has revealed that cystic fibrosis is not caused by a single type of mutation. Although one mutation is associated with 70 percent of all cases, and two others with another 15 to 20 percent, more than 360 mutations have been linked to cystic fibrosis so far. No one has yet been able to correlate firmly the severity of the disease with different mutations. And DNA tests designed to catch one mutation will miss others.

All these discoveries make it much harder to interpret the results of genetic tests for cystic fibrosis. A positive test result does not indicate how severely afflicted a patient will be, and a negative test result could be misleadingly reassuring. Thus, the DNA tests usually need to be confirmed by biochemical assays and monitoring for symptoms.

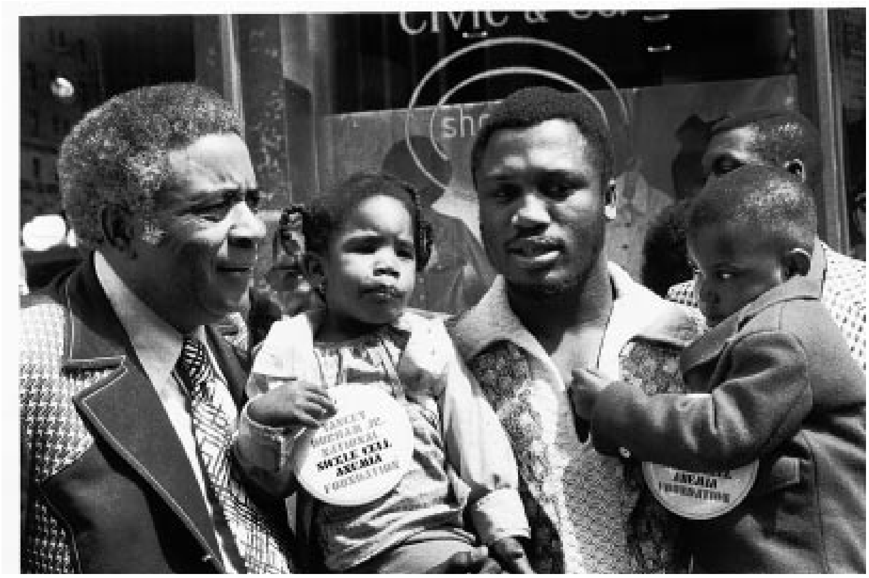
The problem confounding the single-gene disease model and the straightforward interpretation of the tests is what many observers call the myth of genetic determinism. “Genetic determinism is one of these simpleminded errors that we were prone to commit when we thought genes linked to diseases in a kind of inevitable, ineluctable fashion,” explains Thomas H. Murray, director of the Center for Biomedical Ethics at the Case Western Reserve University School of Medicine and former head of a task force for the Human Genome Project. “It invites you to think that ‘genes equal fate.’ ” Environmental circumstances, in the form of modified diets, therapeutic drugs, behavioral changes and other influences, can avert many disastrous outcomes foretold in the DNA. Conversely, because cystic fibrosis, heart disease, cancer, autoimmune disorders, multiple sclerosis and other conditions arise from an unfortunate confluence of genetic and environmental factors, genetic tests for those illnesses can never by themselves predict an individual’s future with perfect clarity.

Damned by DNA

Against this backdrop of genetic interpretation (and misinterpretation), the drama of population screening is being played. In recent decades, several screening programs aimed at detecting genetic diseases in large groups of people have been attempted, some with good results, some with bad.

One nightmarish example of well-intentioned testing gone wrong is the screening campaign for sickle cell anemia during the early 1970s. In response to a groundswell of demands that something be done about the disease, which is a scourge of the African-American community, the federal government funded a screening program to detect carriers of the sickle cell gene. The program was easy to implement because carriers could be identified from just one drop of blood through an inexpensive test. At first, the screening enjoyed popular support. Some ministers conducted tests on their congregations; some members of the Black

CHROMOSOME MAPS of the genetic traits for a variety of illnesses are still being compiled. Disease-related genes have already been located on each of the chromosomes, and new ones are constantly being discovered. The relation between genes and diseases is rarely simple, however. Almost all diseases are likely to have some genetic component.

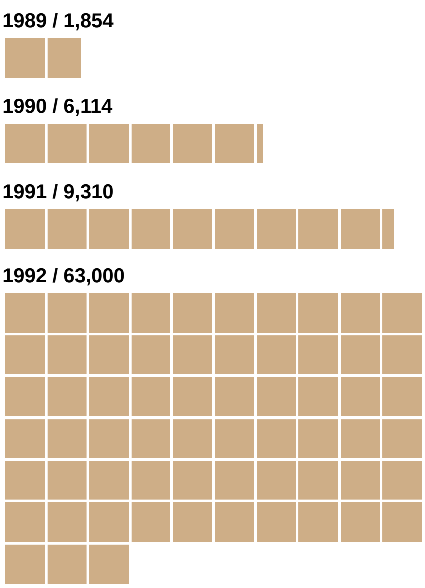


MASS SCREENING for sickle cell anemia was at first popular among African-Americans during the early 1970s. Boxer Joe Frazier (*center*) is shown promoting one screening drive. Because the public was poorly educated about the meaning of the tests, the results were sometimes misused to discriminate against healthy carriers of the trait.

Panthers were offering the tests door-to-door in black communities.

But soon things turned ugly. Because people were rarely educated about the meaning of the tests, many perfectly healthy carriers of the trait were led to believe they were sick. This ignorance extended to some state governments as well. The Massachusetts legislature, for instance, passed a law requiring that all children at risk for "the diseases" sickle cell anemia and sickle cell trait be screened before enrollment in school. "By legislative fiat, sickle cell trait became a disease," Kaback moans.

GROWTH IN THE USE OF CYSTIC FIBROSIS CARRIER TESTS
(YEAR / NUMBER OF TESTS)



SOURCE: Office of Technology Assessment, 1992

TESTING FOR CARRIERS of cystic fibrosis has mushroomed in recent years.

Some insurance companies began to deny coverage to black carriers on the grounds that they had a preexisting medical condition or that their children were bad risks. The U.S. Air Force Academy rejected black applicants who were carriers. Some commercial airlines refused to hire carriers as flight attendants because of the erroneous belief that such individuals were particularly likely to faint at high altitudes. Prominent scientists suggested on network television that the best solution to the anemia problem would be for blacks carrying the gene to forgo breeding—a suggestion that naturally fed fears that the screening was really genocidal in intent. In short, the testing did more harm than good by becoming a tool of long-standing prejudices.

The Tay-Sachs program, which Kaback organized, is a shining example of what can go right with such an effort. Tay-Sachs disease is especially prevalent among Jews of eastern European descent. (The Abshires are not Jewish, but they come from a small community in Louisiana where the disease is also common.) Since the early 1970s more than a million Jews throughout the world have volunteered for testing to learn whether they are carriers of the recessive Tay-Sachs trait. When couples who both carry the mutation decide to have children, they typically elect to have prenatal testing. If a fetus has the disease, they usually abort it rather than give birth to a child who would succumb within five years to a horribly slow, painful death. More important, however, the tests also set at ease the minds of fearful couples who might otherwise never risk having children.

Why did the Tay-Sachs program succeed where the one for sickle cell failed? Kaback and others credit the care that went into its implementation. Before the pilot programs in Baltimore and Washington, D.C., began, 14 months were spent establishing contacts within the Jewish community and educating potential patients about the tests and their implications. People received extensive genetic counseling both before and after the tests. Unlike the screening for sickle cell anemia, the Tay-Sachs program was always voluntary, which meant that people had the opportunity to prepare for the consequences of the testing.

Another important difference, Kaback argues, lies in the diseases themselves. Because Tay-Sachs is always so uniformly hideous in its progression, extremely few people believe an affected child should be brought into the world. Because testing can prevent that tragedy, it carries a clear benefit. The benefits of sickle cell testing are less distinct. The severity of the disease is variable and intermittent, and with prompt medical attention, immunizations and antibiotic treatments (which alleviate the frequent bacterial infections that accompany the anemia), patients can live for many decades. Relatively few people believe aborting an anemic fetus on the basis of a genetic test is humane, so the test has less obvious utility.

Learning the Lessons

The Tay-Sachs model for screening has been successfully adapted for some other diseases. For instance, prudent testing has greatly cut the incidence of thalassemia, a genetic blood disorder common among many people of Mediterranean descent: in Sardinia, the rate of thalassemia has declined from one in 250 births to one in 1,200 births during the past 20 years.

Applying the lessons learned from the Tay-Sachs and sickle cell experiences is not always easy, however, especially for diseases in which the population at risk is extremely large. Once again cystic fibrosis serves as a useful example. In families with histories of the disease, screening is usually accurate and beneficial. But 80 percent of the children with cystic fibrosis are born to families without such histories, so nearly all couples in the U.S. would need to be screened to prevent most cases. "How do you get four million pregnant women a year into an educational setting?" Fost asks. Given that the number of professional genetic counselors in the U.S. is barely more than 1,000, cystic fibrosis screening alone would swamp the country's counseling resources. Most patients would have to get counseling information from their primary physicians—many of whom have little or no training in genetics, according to surveys of the medical profession.

Yet many researchers believe, as Fost puts it, cystic fibrosis screening is "metastasizing, despite the lack of any evidence that it works." For the past eight years, investigators in Wisconsin have been conducting a double-blind study to determine whether biochemically screening newborns for cystic fibrosis is beneficial. So far they have found no evidence that identifying the children at birth is better than waiting for any symptoms to emerge. Nevertheless, in 1989 the states of Colorado and Wyoming both made cystic fibrosis screening mandatory for all newborns.

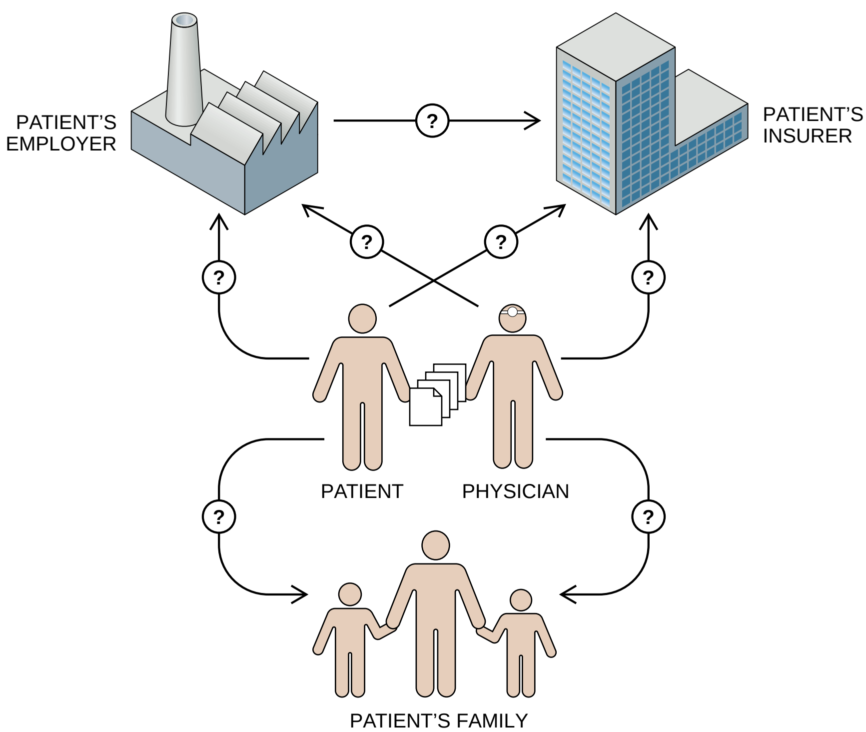
Because all genetic tests have some margin of error, excessive newborn testing has the capacity to do harm, if only by worrying parents unnecessarily. One study looked at families in which a child was initially diagnosed as having cystic fibrosis but was later shown to be healthy; one fifth of the parents nonetheless continued to fear that their children had the disease. Geneticists and ethicists often express concern about how stigmatizing children with a disease label may warp personal development.

In a report issued last November, the Institute of Medicine (IOM) of the National Academy of Sciences offered numerous recommendations about how genetic testing should be conducted to minimize its potentially adverse effects. All testing, it suggests, should be voluntary, and the results should be kept confidential to prevent misuse. All testing should be linked to genetic counseling, so that the patient understands the results and their implications. For the most part, tests should be restricted to those conditions for which some beneficial intervention is possible, either as therapy or as reproductive planning. Tests should not usually be performed purely for information's sake; rather the patient should be able to use the result to make an informed decision about an issue of immediate relevance, such as having a child or electing a medical treatment. The IOM also strongly believes an individual should decide freely whether or not to be tested, without social pressure or financial inducement.

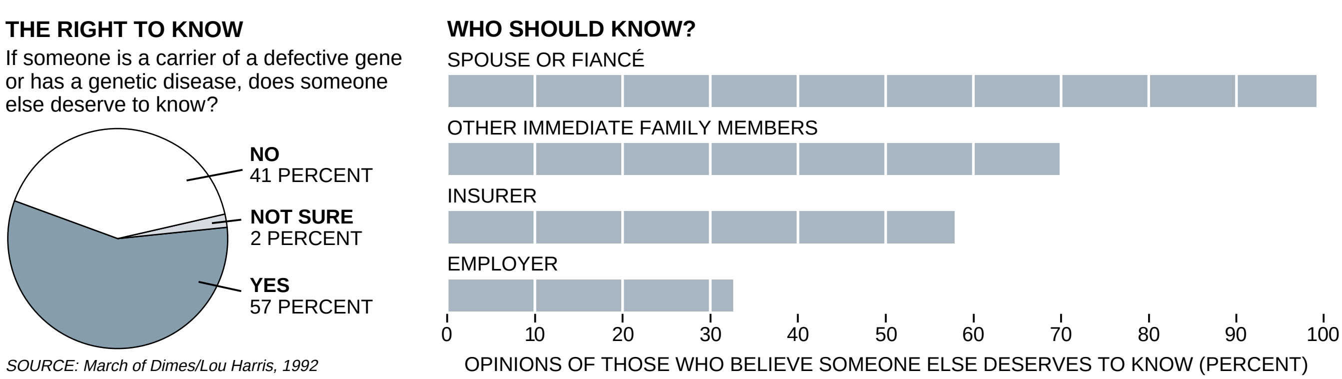
Unregulated Testing

As good as those guidelines may be, abiding by them will be difficult. One problem with trying to set any limits is that genetic testing is so easy to do; another is that the field is largely unregulated. "Currently there is very little to stop someone from implementing genetic tests on a population basis without the sort of institutional review and informed consent required for other new technologies," Fost says.

Neil A. Holtzman, a health policy expert at the Johns Hopkins University Hospital and an editor of the IOM report, notes that most companies in the business of genetic testing offer it as a laboratory service and not as a kit for physicians to use. They are therefore not obliged to submit their methods for appraisal and approval by the Food and Drug Administration. The companies do fall under the jurisdiction of the Clinical Laboratory Improvement Amend-



WHO SHOULD KNOW the results of genetic tests is an unsettled ethical and legal issue. Because the results of genetic tests may carry health and financial implications for others, a patient or physician may be obliged to divulge that information. In a 1992 poll, a majority of Americans said the privacy of test results should not be absolute.



BREE WALKER LAMPLEY, a television and radio celebrity in Los Angeles, has ectrodactyly, a genetic condition that causes a deformity of the hands and feet. She was publicly criticized for having a baby in 1991 because her child had a 50 percent chance of inheriting that trait. (Her son did indeed inherit that gene.) Her case illustrates the social pressures that can be imposed on individuals who carry genetic defects.

ments of 1988. This law ensures that laboratories conducting interstate commerce in Pap smears and other biomedical tests meet certain standards of reliability. Unfortunately, Holtzman says, the Health Care Financing Administration, the agency empowered with enforcing those rules, "has really dragged its feet on setting up guidelines for these new genetic tests."

According to the IOM report, only 10 states have established any licensing requirements for genetic testing laboratories, and only New York State has comprehensive regulations that pertain to DNA tests. This lack of oversight worries Holtzman and the rest of the IOM committee about the possible margin of error in the test results. "I think we've already seen a couple of examples of companies seemingly getting tests out there without any regulatory brakes being put on," Holtzman charges. "If we don't get the proper regulatory authority in place, it's going to become a problem of too little, too late."

Critics also point out that many genetics researchers have financial ties to companies in the business of testing. "What I see my colleagues doing is isolating a gene, finding a single mutation and then jumping into population screening. That's not the way it should happen, in my opinion," Kaback says. "We've got entrepreneurial interests influencing judgments that people are making about when tests are ready to be deployed in the population."

Kaback and the rest of the IOM committee also worry about pressures being put on physicians to use genetic tests. "There are private companies sending letters to doctors all over the country, telling them they should be offering cystic fibrosis carrier testing to all their pregnant couples," he says. In these litigious times, many doctors may feel it is safer to order a test than to face charges of negligence.

Because of such concerns, several professional groups have issued statements that emphasize the experimental status of most genetic tests. The American Society of Human Genetics has twice announced that offering screening for cystic fibrosis carriers to the general public should not be considered the standard in medical practice. This past March, the National Advisory Council for Human Genome Research at the National Institutes of Health warned that screening for cancer should not be performed widely until more research on its reliability and consequences could be determined.

Genetic Privacy

Even if the accuracy and utility of genetic testing are assured, maintaining the privacy of that information will remain a problem. Your genes are not exclusively your own: you share half of them with each of your parents, siblings and children. If you discover that you carry a worrisome gene, you may have an ethical, if not legal, obligation to tell them. "There are ripples from genetic testing that don't have analogues in most other kinds of medical testing," remarks Arthur Caplan, a bioethicist at the University of Minnesota.

A 1992 March of Dimes poll reported that 57 percent of the public thinks someone other than a patient deserves to know that he or she carries a defective gene. Of those who believed so, 98 percent thought a spouse or betrothed should know. More surprisingly, however, 58 percent thought insurance companies should also be informed, and 33 percent thought an employer should be told. In a different study, physicians themselves disclosed a willingness to violate patient-doctor confidentiality in some cases: 54 percent said that, even over a patient's objections, they would tell relatives at risk about the results of a test for Huntington's disease (a lethal neurodegenerative disorder that usually manifests in middle age). Twenty-four percent said they would tell the patient's employer, and 12 percent would tell an insurance company.

Industry has also revealed an appetite for individuals' genetic information, although the limited predictive ability and high cost of the current tests seem to have restricted their appeal. A survey by the OTA released in 1991 tried to determine whether employers were conducting genetic tests as conditions of employment; such tests could theoretically reveal workers who would be particularly bad (or expensive) health risks. It found that in 1989 only 12 of 330 Fortune 500 companies were monitoring or screening for any reason. But more than half of the polled companies found the idea of monitoring acceptable, and 40 percent admitted that a person's health insurance costs might affect his or her chance of employment.

Bioethicist Thomas Murray has also found interest in genetic data among insurance companies. "If you ask the question narrowly, 'Are insurers requiring DNA tests of customers?', the answer is no," he says. "The tests are too expensive, not cost-effective, there aren't enough of them yet—there are a lot of reasons. But are insurers interested in genetic information, and will they be in the future? The answer to that is a resounding yes. The best genetic information right now comes not from genetic testing but from one's personal health history—what did your parents die of?"

The insurance industry itself confirms that view. A joint report issued in 1991 by the American Council of Life Insurance and the Health Insurance Association of America stated that although insurers would not be requiring tests in the foreseeable future, they should be entitled to know the results of any genetic tests or other evidence of possible disease in a policyholder's record. Insurers maintain that they need this information not to discriminate genetically but to set fair rates for all policyholders. It would be wrong, they argue, to make healthy people pay higher premiums because they had been lumped in with those at higher risk. They also do not want to exempt genetic testing results from scrutiny because they say this precedent might someday preclude them from using other medical and statistical indicators of risk.

These arguments do not persuade Murray. "Insurers are applying a model of 'actuarial fairness' and justice they developed in the realm of commercial insurance," he replies. Those principles, in his opinion, are not relevant to health insurance, which serves social and humanitarian functions that deserve consideration. Because no one can control his or her genetic makeup, it seems wrong to penalize individuals for it.

Many observers also doubt the ability of insurance companies and other agencies to interpret genetic information wisely. Paul R. Billings, a geneticist now at the Palo Alto Veterans Administration Medical Center, has been collecting examples of people who were apparently discriminated against because of genetic information that became known to insurers, employers, health maintenance organizations (HMOs) and adoption agencies. He and his colleagues first published some of their results in the *American Journal of Human Genetics* in 1992. Several of the cases they described concerned people who were healthy carriers of genetic diseases or had extremely mild symptoms yet were still denied jobs or insurance coverage. One woman who applied to become an adoptive parent was rejected because her family history of Huntington's disease made her "too great a risk."

In another case, parents who had one child with cystic fibrosis conceived again, and prenatal testing confirmed that this second child, too, would have the disease. When the family's HMO learned that the couple intended to proceed with the pregnancy, it moved to withdraw or limit the entire family's health coverage. Only after threats of a lawsuit did the HMO change its mind.

Billings says a second report, listing about 100 cases of genetic discrimination, is almost done. With funding from the Human Genome Project, he and his co-workers have also been conducting a further survey of 30,000 people with genetic conditions. "Our preliminary view of that study is that it confirms genetic discrimination is occurring," he notes. A 1992 OTA report also found that about 15 percent of genetic counselors said some of their clients had experienced discrimination in insurance.

Outside the Law

Anyone looking to the law for protection from genetic discrimination would find a thin shield at best. At the federal level, most experts agree, the strongest statute for barring discrimination in the workplace is the 1990 Americans with Disabilities Act. But the relevance of that act to genetic information is still uncertain. The Equal Employment Opportunity Commission, which enforces the act, has already offered the opinion that healthy carriers of genetic diseases would probably not qualify as having a disability.

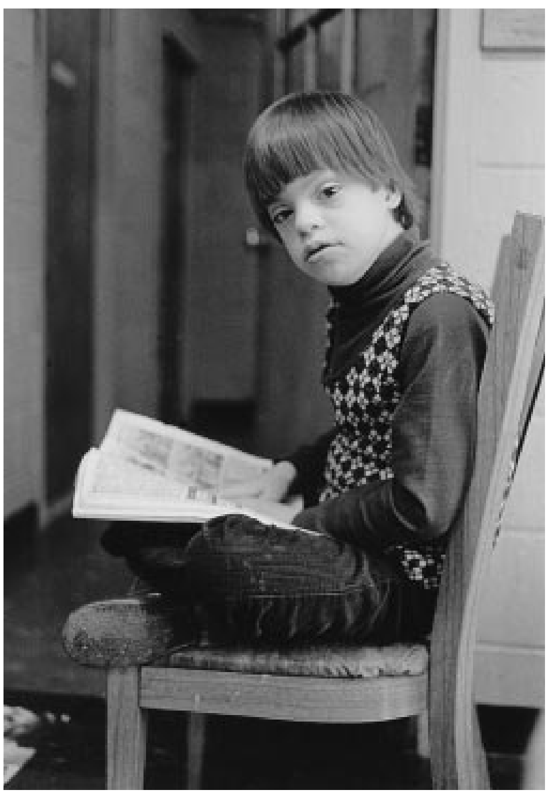
Some states have laws that protect the carriers of genes for sickle cell anemia and a few other specific traits found disproportionately among African-Americans and other groups whose members have been historical targets of discrimination. Somewhat broader legislation has been passed in Wisconsin and a few other states and is being debated in a number of others. In 1991 the California legislature passed a law that prevented any form of genetic discrimination, but it was subsequently vetoed by Governor Pete Wilson.

Because most concern about discrimination revolves around insurance coverage, insurance reform may be the key to a solution. "Our task force recommended that all individual risk information be excluded from decisions about who gets insured, what they get insured for and how much they get charged," Murray says. "We see no other practical, sustainable plan for health care coverage than community rating." Community rating, which was the basis for the first health insurance programs during the 1930s, is a system in which a customer's premiums are determined by the health profile of his or her community. Genetic information about individuals would be irrelevant.

The IOM committee and others also favor community rating for that reason. Insurers disagree, arguing that individual risk rating serves the public welfare more equitably at less expense. Nevertheless, for many reasons, the popularity of individual risk rating has been declining. In recent years, New York, Maine and a few other states have shifted to programs based at least partially on community rating.

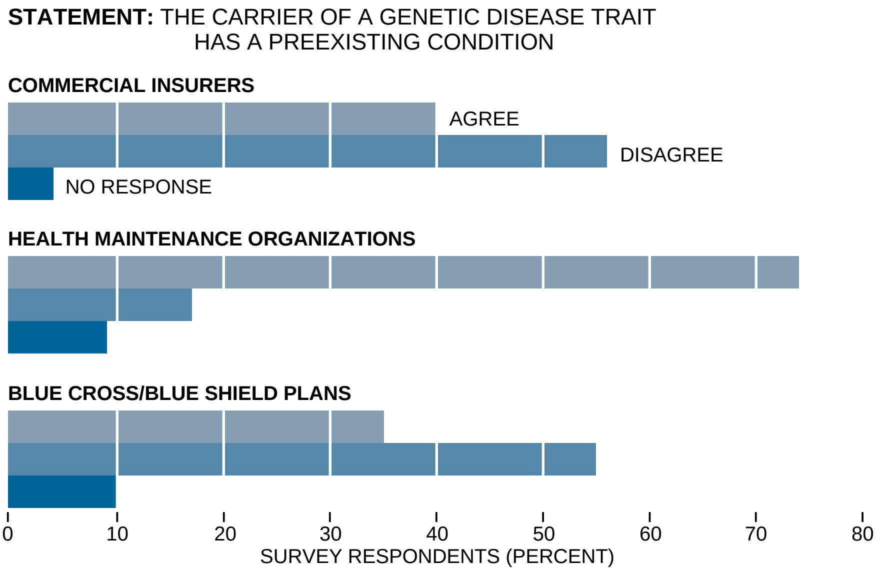
Both insurers and their critics agree that reforms in health care financing are likely to render moot some—but not all—of the disputes about genetic discrimination in insurance. The details of the new health care system will be important. For example, some insurance plans that would meet the Clinton administration's standard of universal coverage might still be able to disallow coverage for certain genetic conditions. People with those traits might then have to buy additional insurance or pay some costs out of pocket. Under the current system, insurance companies and HMOs often resist paying for many genetic tests or for the costs of genetic counseling. It remains to be seen how reforms in health care financing will affect those policies.

Universal, comprehensive coverage could also become the in-



DOWN SYNDROME patients, who carry three copies of chromosome 21, exhibit a range of mental and physical impairments. But given the proper education and medical attention, many of them are accomplishing much more than was once believed possible. Predicting the limits of people with genetic handicaps is uncertain.

strument of a demon that has long haunted genetic technology: eugenics. The word immediately summons memories of the Nazi genocidal horrors and other atrocities, such as the forced sterilization of 30,000 people as "mental defectives" in the U.S. before World War II. Yet eugenics can arise through a seemingly more benign movement to ration social resources. Last December, China announced a policy of discouraging people at risk for hereditary diseases from having children, on the grounds that the genetically ill impose too much of a burden on society.



SOURCE: Office of Technology Assessment, 1992.

INSURERS' ATTITUDES toward people who carry the genes for disease traits may shape policies concerning the eligibility of those persons and their offspring for health care coverage. Critics worry that genetic discrimination could occur.

A Eugenic Democracy

The U.S. is not immune from such thinking. "If we get universal health care, people are going to want to know why they should be paying for the 'genetically irresponsible,'" Caplan believes. "In this country, eugenics is not going to come from a Hitlerian dictator saying, 'You must do this.' It's probably going to come from a society saying, 'You can have a kid like that if you want, but I'm not paying.'"

A counterargument might be that in a democracy, the public should be entitled to set limits on what costs individuals can impose on everyone. Yet such arguments, like those surrounding insurance, invariably depend on a definition of fairness. Moreover, whatever motivations lie behind the pressures being put on people making reproductive choices, their net effect is the same: society influences who will or will not be born. As Billings says, "Whether we want to dress it up in economic incentives, if you have to sell your house and go broke to have a child of a certain type, that's eugenics."

Because of its history, "eugenics" is perhaps too loaded a word to describe some applications of genetic testing. Holtzman and many others prefer to reserve it to describe circumstances in which a government or other agency intervenes in reproductive decisions. That is why, Holtzman says, the IOM committee so emphasized the importance of preserving personal autonomy: couples should weigh for themselves whether the risks of having a sick child outweigh other considerations.

"I think the goal of eliminating diseases and disabilities is a good one," Caplan affirms. "I don't think there's anything wrong with encouraging women at risk to be screened for spina bifida or to ask Jewish couples of eastern European descent to be screened for Tay-Sachs. It's wrong to confuse the goal of eliminating disease with the moral problem of coercion."

The troublesome fact remains, however, that the line between genetic diseases and undesirable traits is blurred. Caplan cites the example of albinism, which is not a disease but is associated with a higher risk for skin cancer, faulty vision and social stigma. "I think the standard that medicine wants to go with is, is there clear dysfunction or disorder? If not, I would argue that medicine ought not to be testing for it and counseling for it. If we're in the gray zone, I suspect we should try to stay out of those areas, because there's so much else of clear value that we could be doing," he suggests.

Yet Caplan admits that—as much as he would like to—he cannot frame a convincing standard that would allow parents to select some of their children's traits but not their sex, height or other cosmetic features, presuming technological feasibility. Distaste for abortion will stop many parents from exercising that veto for now. But such choices could be circumvented by emerging technologies for screening embryos before they are implanted—or even before conception. "So I think the stance that we will deal only with clear-cut disorders will last about five minutes," Caplan answers ruefully. "Once you can actually do that testing, the interest will swamp my objections. The ability to choose the traits of your child will roar through with a whoosh."

The consequences of those choices may be very hard to predict. Murray recalls that at a meeting of the American Society for Human Genetics a few years ago, he heard about a deaf couple who had asked one geneticist whether the hearing ability of a fetus could be determined prenatally: they wanted to go through with the pregnancy only if they knew that the child *would* be deaf.

As Brittany Abshire and millions of other healthy children and adults prove, genetic testing can immeasurably improve the quality of life for individuals, even entire families. To ignore the good it can do would be an act of immoral blindness and cowardice. But using these technologies wisely will demand foresighted social and legal policies. The record is discouraging: in the past, when genetic discoveries have made tests possible, policymakers have often either encouraged them prematurely or acted too late.

"The paradigms of eugenics are programs of unsurpassed evil. They're not going to get any less evil just because our genetics got better," Murray muses. "We need to be very conscious of what we're dabbling in."

FURTHER READING

CYSTIC FIBROSIS AND DNA TESTS: IMPLICATIONS OF CARRIER SCREENING. Congress of the United States, Office of Technology Assessment. Report OTA-BA-532, August 1992.

MOLECULAR BIOLOGY: IMPACT ON HUMAN DISEASE. Theme issue of *FASEB Journal,* Vol. 6, No. 10; July 1993.

ASSESSING GENETIC RISKS: IMPLICATIONS FOR HEALTH AND SOCIAL POLICY. Edited by Lori B. Andrews, Jane E. Fullarton, Neil A. Holtzman and Arno G. Motulsky. National Academy Press, 1994.

Cyberspace Cadets

Gates and McCaw take new venture into orbit

The proposal by Teledesic in Kirkland, Wash., to launch an 840-satellite communications network bears a peculiarly American imprimatur: an appeal to the utopian strain in the national character and a confidence that no vision is beyond the reach of innovation and hard work. But like many such utopian visions, it also manifests a grandiosity that pushes sanity—in this case, the engineering and fiscal varieties—to its limit.

The plan calls for spending $9 billion to circulate some 900 satellites (840 active and up to 84 spares) in 435-mile-high orbits to deliver telephone, video and computer data services to the entire world. That armada outnumbers by a factor of at least two all the commercial communications satellites currently in orbit.

Teledesic must position the 900 satellites precisely so that they hurtle in 21 loops around the North and South Poles; each orbital plane would contain up to 44 satellites. A satellite would communicate with eight other satellites: two to the front and two to the rear in the same orbital path, and two others to each side in adjacent orbits.

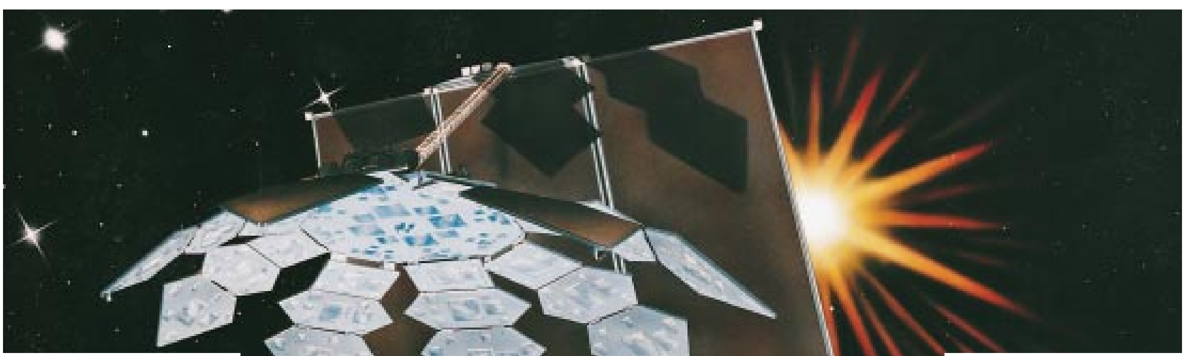
Each bird, weighing 1,700 pounds at launch and measuring about 40 feet in diameter when fully deployed, carries an advanced packet switch for relaying voice, video and data. These asynchronous transfer mode switches, as they are known by techies, have only recently entered commercial use in terrestrial telephone networks.

To meet an ambitious 2001 deadline, Teledesic would have to launch about a rocket per week or more over a two-year period, each one carrying eight or more satellites. The estimated number of components required for the total complement of Teledesic satellites would be sufficient to create an industry: the startup's application to the Federal Communications Commission cites the need to make 500 million gallium arsenide chips, more than have ever been manufactured commercially.

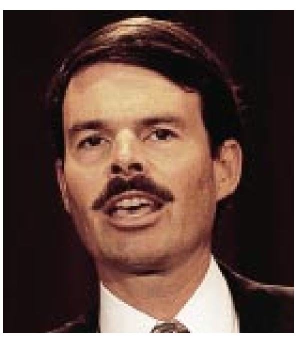
"The state of the art is such that we don't know how to launch them, we don't know how to keep them in orbit and we don't know how to build them at that price," says Burton I. Edelson, a satellite communications specialist at George Washington University. "We have a lot of problems maintaining system viability with a small number of satellites, even weather satellites."

Also perplexing is the market for these services. Teledesic portrays itself to the FCC as an Internet for Montagnard tribesmen. "Today, the costs to bring modern communications to poor and remote areas is so high that many of the world's people cannot participate in the global community," reads Teledesic's FCC application. "Yet the benefits of the communications revolution should be extended to all of the world's citizens, including those...who do not have ready access to doctors, hospitals, schools, or libraries, and who are at risk of being shunted aside." Hamlets and wadis may be well served. Not so the hordes of laptop computer and handheld device owners: the technology operates with a fixed ground antenna to provide large amounts of digital communications capacity for video and data users. Teledesic's market, Edelson says, could be more economically serviced by a few big satellites in the much higher geostationary orbit where they would cover a broad swath of the planet.

Winning endorsements from dozens of national telecommunications authorities will prove as difficult as raising the projected $9 billion. Furthermore, the intended cost of each satellite, at less than $10 million, is a small fraction of the price tag for other advanced communications satellites. "I don't know



TELEDESIC



SYGMA

SYGMA

SATELLITE is one of 840 in a worldwide network envisaged by the newly formed corporation Teledesic, in which Craig McCaw* (left) *and Bill Gates* (right) *have invested.

that anybody finds the $9-billion figure credible; it will cost at least $25 to $35 billion," says Alan L. Parker, president of Orbital Communications, a Virginia-based company that plans to begin the launching of 26 small satellites for data transmission this year.

A recipe for a technologically impossible and financially dubious program on the scale of the Strategic Defense Initiative? Then why are two of the initial financial backers, William H. Gates III and Craig O. McCaw, preeminent among entrepreneurs in computer software and telecommunications?

Neither tycoon is betting a life's savings on Teledesic's success. Within days of the Teledesic announcement, Microsoft, of which Gates is chairman and a major shareholder, agreed to build a wireless network for data transmission with Mobile Telecommunications Technologies, the largest U.S. paging concern.

In fact, Gates and McCaw have each invested but $5 million in Teledesic. A subsidiary of McCaw's cellular telephone company has put down an additional $3.7 million. If Teledesic somehow became a success, these relatively small investments could yield large payoffs. John Pike, an analyst with the Federation of American Scientists, thinks Gates and McCaw may be interested in playing in low-earth orbit for some of the same reasons that the U.S. Department of Justice antitrust lawyers are interested in Microsoft. The game is called monopoly.

Teledesic does not intend to supplant the local telephone company; it will sell communications capacity on its satellites only to existing service providers, many of them in developing nations. Bolstered by the names Gates and McCaw, Teledesic may try to convince state-held telephone companies to sign on as investment partners for this orbiting version of the Internet.

If governments cast their lot with Teledesic, the competition—companies laying fiber-optic cable, for example—might find entry into these emerging markets blocked by the coopted bureaucracies. Even if a national telecommunications market is open to competitors, the early arrival of "an 800-pound gorilla," Pike's characterization of Teledesic, would complicate later attempts by others to enter this business.

Gates and McCaw could emerge as the Harriman and Vanderbilt of the information age, Pike says. "Most of the other cyberspace projects are just a better way to watch *Gilligan's Island,*" he adds. "If this thing happens like they hope it will happen, it would be a big deal, literally one of the defining realities of the third millennium, in the way that Ma Bell defined reality for most of the 20th century in the U.S."

Gates's and McCaw's thoughts about their critics are hard to know because they do not share them publicly. (SCIENTIFIC AMERICAN's requests for interviews were turned down.) In any event, the original idea for Teledesic came from neither man, but from a venture capitalist, Edward F. Tuck. For the past 10 years, Tuck has been involved in starting new companies, among them an enterprise named Calling Communications, which was intended to become a worldwide telephone service using a network of communications satellites.

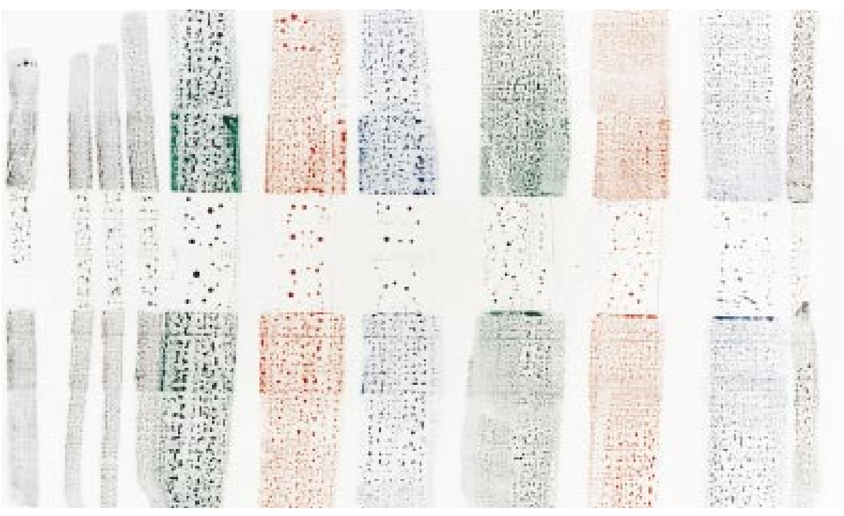
Good-bye, Taki 183

When polymer chemists want to fashion a material to which nothing will stick, they often design a molecule in which a fluorine atom takes the place normally occupied by that of hydrogen in the ubiquitous hydrocarbons. Fluorine forms a tight covalent bond with carbon, keeping it from combining with atoms from other molecules.

The best known of these antistick concoctions is polytetrafluoroethylene, aka Teflon. Dow Chemical has come up with a new fluorocarbon formulation whose abhesive (as opposed to adhesive) properties may be better than ordinary Teflon and that may be of use where its frying-pan cousin may not. The fluorochemical might, for example, serve not only as a protective coating for subway walls but as a surface that guards against barnacles on ships, dirt on wallpaper and ice on aircraft.

Unlike hard-to-apply Teflon, the Dow coating can be painted onto a surface with a brush or spray gun. Cured at a relatively tepid 100 degrees Celsius for an hour, it forms a hard, clear finish. Teflon must often be baked on metal at temperatures above 250 degrees C. Once cured, the hardness of the finish left by the Dow polymer resembles an automobile lacquer. Less porous than Teflon, the fluorocarbon stops contaminants from getting caught below the surface. It also prevents the surface from being "wetted," or covered, by common solvents used in paints, dyes or inks, a quality of use in the war against graffiti. A felt-tip marker leaves tiny ink beads that stick poorly on a surface treated with the Dow chemical but remain in place on Teflon. Still, Teflon is likely to retain its longtime affiliation with scrambled eggs. The Dow fluorochemical does not tolerate heat as well as Teflon does. It does, however, reside in a mostly water-based solution and so emits little of the organic vapors that might warrant scrutiny from environmental regulators.

Dow's creation, which has yet to garner a trade name, is manufactured by making a polymer that contains both fluorocarbon and hydrocarbon molecules. The fluorocarbons on these long polymer chains show a marked distaste for water.



***STICK TEST** demonstrates how the ink from a felt-tip marker forms into tiny beads on Dow Chemical's fluorinated polymer* **(horizontal band)**, *but the marker dye remains on the Teflon substrate on both the top and bottom.*

Tuck shared his idea with McCaw, who suggested that the network be set up primarily to carry data and video instead of regular telephone calls. Earlier this year Calling Communications became Teledesic, with McCaw as chairman and Tuck as vice chairman. It was McCaw who convinced Gates to come on board as an investor. Kinship Partners II, in which Tuck is a principal, holds a $1.5-million stake in Teledesic. Tuck has remained fascinated for years with combining satellites and entrepreneurialism. He is best known as founder of a company called Magellan, which is now one of the leading suppliers of the receivers for the Navstar global positioning system, a network of 24 satellites that provides airplanes, boats, automobiles and backpackers with precise location information.

Tuck tries to make Teledesic's seemingly insurmountable costs look somewhat manageable. They can, he says, be addressed through economies of scale realized from building hundreds of satellites. Teledesic, in fact, will utilize methods developed for manufacturing the Brilliant Pebble satellites proposed for Star Wars. It will also tap that program's investigations into automated control of large satellite networks.

The shift to small satellites reflects the broader trend toward miniaturization in the computer industry. W. Russell Daggatt, a lawyer and the president of Teledesic, contrasts the Telepebbles to the huge communications satellites in geostationary orbit at a height of 22,300 miles. "I make the analogy to the computer industry that evolved from big mainframes," Daggatt says. "If you were thinking of sending up 840 of those powerful mainframes, people would think you were crazy. But if you were thinking of sending up 840 notebook computers, it is a different matter altogether."

Daggatt denies that Teledesic wants to monopolize markets in developing nations. "Our total system capacity on a worldwide basis is infinitesimal compared with the need," he asserts.

The brouhaha surrounding Teledesic has served as free publicity for a spate of other ventures that have targeted the lower reaches of outer space. Motorola and partners plan to spend $3.3 billion to send up Iridium, a 66-satellite constellation. Loral, Qualcomm and nine partners envisage spending $1.8 billion on Globalstar to launch 48 satellites. These projects are just two of a number of plans for orbiting networks that would give portable telephone users the ability to place calls from virtually anywhere around the globe.

By using much higher frequencies, Teledesic, which is not a mobile service, is able to provide the network capacity needed for video and high-speed data communications as well as for ordinary telephone calls. It can support more than 20,000 simultaneous digital connections, each carrying 1.54 million bits per second of video or data.

But even Motorola's and Loral's more modest schemes may not escape unscathed. Financing and construction of a multisatellite network must be completed before even the first call is switched. In contrast, fiber-optic or cellular telephone networks can be built piece by piece. Herschel Shosteck, a Maryland economist who has studied the wireless market, believes diminishing costs and advancing technology for land-based cellular networks will make satellite telephone links uncompetitive.

Even on the earth, some grand cyberspace alliances have faltered, brought down by unpalatable financial and regulatory realities. Nevertheless, the field is full of dreamers. And if they succeed in building, can the rest of us afford not to come? —*Gary Stix*

They tend to float to the surface of an aqueous solution, as if they were tiny molecular buoys. The hydrocarbon molecules that remain submerged on the lower part of a chain contain ions that bind to a surface underneath. The result is a perfect yin-yang combination: abhesive side up, adhesive down. These ionic groups also bind to another polymer, an oxazoline, that joins each buoy chain and helps to attach them to a surface. This cross-linkage fixes the chains rigidly in place.

Donald L. Schmidt, a Dow polymer chemist, along with a few colleagues, first devised the material seven years ago. The company will now leave it to others to find commercial uses for the product. "We're not a coating company," says Benjamin M. DeKoven, a project collaborator. Last September, Dow signed a nonexclusive license with 3M, which is investigating a range of applications from finishes for kitchen countertops to a potential addition to its Scotchgard line of fabric coatings. 3M is also working on getting the chemical to cure at room temperature so that it can be sold as an ordinary self-drying finish like paint.

Keeping the writing off the wall may take a while longer. That means that Cool Boogie Dancer, Taki 183 and even good old Kilroy may still be around to leave a memento or two at the millennial New Year's celebration. —*Gary Stix*

Requiem for *Alpha?*

If Rep. Brown loses his game of chicken, the space station dies

The planned international space station has fallen into dire political difficulties that may lead to its demise, a result of budgets below the expected levels and apprehensions about Russia's large role in the program. In March, Representative George E. Brown, Jr., of California, the influential chairman of the House Committee on Science, Space and Technology, announced his "regretful" (and perhaps Machiavellian) "decision" to oppose the space station unless two conditions are met. The first is that the Clinton administration's proposed budget for the National Aeronautics and Space Administration in 1995—$14.3 billion—survive its appropriations battles unscathed. That stipulation alone would require unusual cooperation among legislators. But Brown's second condition for supporting the station puts him in a direct battle of wills—and perhaps wits—with the White House.

Brown says he will oppose the space station, now provisionally known as *International Space Station Alpha,* unless funding for NASA after 1995 is increased above the administration's published plans. Those plans call for only a small raise by 1999, to $14.6 billion. Brown fears that, without the increases he seeks, the space station will "drain the health, vitality and future of the rest of NASA."

Because Brown is a long-term supporter of the space program, his views will carry considerable weight if there is a tight vote. Last year authorization for the station was approved by a margin of just one vote, and the opponents, led by Representative Tim Roemer of Indiana, are hoping they will succeed in canceling the program in 1994.

NASA estimates that the remaining cost of bringing the station to full operation in 2002 is $17.4 billion. According to Marcia S. Smith of the Congressional Research Service, $11.2 billion has already been spent over the past decade on successive designs and redesigns. *Alpha,* unveiled in late 1993, represented the fifth and biggest redesign since the space station program began in 1984. Now it is being modified to accommodate Russian components.

Three weeks after his ultimatum, the Congressional Budget Office handed Brown ammunition in the form of a report that paints a gloomy view of NASA's future. The budget office analysts doubt the agency's ability to improve efficiency. They make it clear that if the space station is to go forward—and budgets do not increase above current levels—plans for unmanned space exploration will have to be scaled back.

Disclosure earlier this spring of a dispute between NASA and the space station's prime contractor, Boeing, has muddied the waters further. The two sides were about $800 million apart at the beginning of April. Both expressed optimism that the gap could be bridged, but the fracas strengthens the hand of *Alpha's* opponents.

A more substantial cause for alarm among station supporters is the growing misgivings about Russia's political course. The purpose of bringing the U.S.'s former cold war adversary into the game was to save money and to persuade Russia to comply with restrictions on exports of missile technology. But the strategy may yet backfire. According to plans NASA was publicizing in April, Russia would launch seven of the first 11 station assembly missions. Moreover, Russia would provide the propulsion module that is essential to keep *Alpha* in orbit.

The heavy reliance on Russian hardware has disturbed some longtime station supporters, including Brown and Representative F. James Sensenbrenner of Wisconsin. Sensenbrenner now says he will vote against the program unless NASA can produce a credible backup plan so that the U.S., together with Europe, Canada and Japan, could complete *Alpha* without Russian help. Furthermore, Smith observes, the amount that Russian involvement will save has shrunk. It appears to be no more than $2 billion, half the figure announced last year by NASA administrator Daniel S. Goldin.

The space station has survived eight previous attempts to cancel it, and it may survive another. Some developments are in its favor: in late March the program, which is now being run by a new management team at Johnson Space Center in Houston, successfully passed two hurdles when it completed a systems design review as well as an inspection by a panel chaired by Charles M. Vest, president of the Massachusetts Institute of Technology. Vest's panel outlined options for the space station last summer. Louis J. Lanzerotti, a physicist at AT&T Bell Laboratories and a member of the Vest panel, says he and other panel members were impressed by the management team and by the attention that NASA is giving to how the station will ultimately be used.

The dream team may have taken to the field too late. Several well-placed congressional aides believe sentiment has now turned against *Alpha*. The new chairman of the House appropriations committee, who will hold decisive power over the program, is Representative David R. Obey of Wisconsin. Obey lambasted the space station in 1993 as a "flying turkey." And a prominent station supporter, Representative Jim Bacchus of Florida, has said he will retire, which has increased uncertainty still more.

If President Bill Clinton were to mount a full-court press for the space station—and if there are no more alarming developments in Russia—the program might yet survive. But John H. Gibbons, Clinton's science adviser, told an appropriations subcommittee earlier this year that if he had to make a choice for additional budget cuts, he would "rather look harder at NASA than look at funds for the National Science Foundation." If Gibbons's remark is a clue to the administration's priorities, then *Alpha* may follow the route to oblivion the Superconducting Super Collider took a year ago. —*Tim Beardsley*


NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

INTERNATIONAL SPACE STATION ALPHA ***is depicted in its latest incarnation, with elements from the U.S., Europe, Canada, Japan and Russia.***

Immuno-Logistics

Moving vaccines from the lab to the bush and the street

The six major vaccines administered every year to some 110 million of the world's children who are fortunate enough to get them—whether in Mogadishu or inner-city Miami—cost more to deliver than to make. The United Nations Children's Fund, for example, spends a total of $1.50 per child on the vaccines produced for diphtheria, pertussis, tetanus, poliomyelitis, measles and tuberculosis. That amount is a tenth of what a government then has to disburse for labor, transportation, training and refrigeration to get these vaccines to infants and young children.

Streamlining the way vaccinations are administered would bring about more savings than would measures to lower the cost of these commodity vaccines. "It's an almost perfect use for the science and technology coming out of the biotechnology industry," says Anthony Robbins, who is helping to develop a Clinton administration program to immunize all children born in the U.S. against a battery of diseases.

Such optimism colors the planning of the Children's Vaccine Initiative (CVI), an effort by leading international organization programs to spur development of a supervaccine that could make immunizations more efficient and less costly while extending coverage to individuals exposed to such endemic diseases as malaria and shigellosis. The CVI has set the ambitious goal of devising a vaccine that can be trickled into a baby's mouth in a single dose shortly after birth. This ideal vaccine should re-

TIMOTHY ROSS *J.B. Pictures*

MALARIA VACCINES, such as this experimental one about to be administered to Colombian children, might eventually be added to the regimen of childhood vaccinations. That goal would help realize the vision of the international Children's Vaccine Initiative.

sist spoilage in the heat of the tropics. It should also incorporate additional disease protection—malaria and hepatitis B, for example. And, of course, it should be cheap.

The CVI vision so far is just that. Its annual budget for research, development and other expenses totals a mere $4 or $5 million, despite fund-raising plans to bring in more than $300 million. But even in its embryonic state, the program, shepherded since 1990 by the U.N. Children's Fund, the U.N. Development Program, the Rockefeller Foundation, the World Bank and the World Health Organization, has provided focus to disparate laboratory activities.

The Benezech-Simpson Company in Vion, France, and the Pasteur Institute found that a vaccine made by combining a live but weakened poliovirus with deuterium oxide (heavy water) could last for a week at tropical temperatures without refrigeration. "Deuterium forms stronger bonds to the virus than does the hydrogen in water normally used in this process," says Philip K. Russell, a professor at the School of Hygiene and Public Health at Johns Hopkins University and an adviser to the CVI. "The more rigid structure presumably slows down the degradation of nucleic acids and proteins in the virus."

The Walter Reed Army Institute of Research and its partner, Virogenetics, have developed a unique vaccine against malaria. Most trials with malaria vaccines try to guard against a single stage of the life cycle of the parasite that causes the disease: either when it is first injected into the human bloodstream through the proboscis of a mosquito, or once it changes and multiplies in the liver, or when it reenters the blood from the liver. This new vaccine should supply a complete defense.

It works by bioengineering into the vaccinia virus the genes for seven malarial antigens, one or more of which correspond to a life stage of the parasite. The technique of studding a single virus with genes for various antigens is also being explored to induce immunity against several diseases at once. The virus serves as a vector that, once injected, allows the genes to express antigens that provoke an immune response.

OraVax, a firm in Cambridge, Mass., is collaborating with the University of Maryland's Center for Vaccine Development on human trials of an oral vaccine that uses an attenuated strain of *Shigella* to promote immunization against the dysentery-causing bacterium.

CVI's vision will still need molding with hard currency; many small biotechnology research companies lack the financial resources to develop a vaccine for the commercial market. For their part, major pharmaceutical companies may show scant interest in research and development on these new agents.

Government help may be needed. A National Academy of Sciences report, issued last year, recommended establishment of a National Vaccine Authority. The authority would be empowered to subsidize purchasing or even assist in development of vaccines spurned by the private sector. If the government could take a vaccine through late-stage clinical trials, the costs of development would diminish for a large pharmaceu-

tical producer. A company might then be willing to take over the manufacturing of an otherwise unprofitable product. "There's a problem making the leap between a good research idea and something licensable," says Jerald C. Sadoff, director of the division of communicable diseases and immunology at the Walter Reed institute.

But pharmaceutical companies have not embraced the idea of Uncle Sam as vaccine developer. "The government does not have the people, the facilities or the researchers to do it," says R. Gordon Douglas, Jr., president of Merck & Co.'s vaccine division. Douglas prefers the option of vending technology to developing countries. Merck's sale of the intellectual know-how for making hepatitis B vaccine to the government of China exemplifies the process, he says.

Vaccines may never prove as enticing as drugs for pharmaceutical makers. They can be more expensive to develop, and they may produce less income—repeated refills not required. Governments and international aid agencies also cannot afford pharmacy prices.

By one estimate, the annual worldwide vaccine market is about $2 billion. At $1.2 billion in revenues and climbing, Prozac, that existential elixir, might overtake the sale of potions that save the lives of millions of children every year. —*Gary Stix*

Little Winners

Bioprospectors reap rewards from Third World microbes

Strolling through a Costa Rican bazaar, you trip over a sick dog. Effluent drips from its inflamed eyes. Do you hop away in alarm—or collect the fluid? Apparently the latter, if you happen to be a bioprospector. The microbe *Acremonium kiliense,* deposited with a governmental agency for possible patent claim, was gleaned from just that source.

Since penicillin achieved commercial success near the end of World War II, drug manufacturers have been scouring the nooks and crannies of the earth for microorganisms that produce medically active chemicals. Successes have been manifold. Microbes in Spanish soil harbored an anticholesterol agent marketed as the financial blockbuster Mevacor. Others from Argentina, Venezuela and elsewhere have yielded the streptomycins and chloromycetins of corner drugstores.

Such pharmaceutical treasures used to be literally free for the taking. Now developing nations, waking up to the fortunes that they have unknowingly exported, are asking for some returns.

To bioprospectors, microbes have a particular advantage over plants: they are small. A fistful of soil or a flaskful of sewage may contain thousands of organisms that can easily be cultured in a laboratory. Besides, one-celled creatures, which like plants and insects are much more diverse in tropical countries, have traditionally attracted less political attention. Most nations did not really care if within their boundaries you clipped the hair of a spiny rat or rooted in zebra dung.

Now the Convention on Biological Diversity has changed all that. Although microbes may flourish in tropical countries, Pat Roy Mooney of the Rural Advancement Foundation International in Ottawa, Canada, estimates that 86 percent of microbial culture collections are in the industrialized world. The American Type Culture Collection (ATCC)—where all microbes under U.S. patent application, and many more, are deposited—lists more than 3,000 bacteria and fungi collected in nations such as Panama, Ethiopia and India. When the convention came into force in December 1993, these microorganisms became in effect the exclusive property of the depositor. The country of origin will have an uphill battle claiming any royalties that a microbe may earn.

Increasingly, Southern Hemisphere nations are feeling that they never quite figured out the value of what they gave away. Besides the microbes, the 68 percent of seeds in Northern gene banks that were originally bred by farmers in developing nations—as well as plants yielding a quarter of the world's prescription drugs—now belong to industrialized countries. Comments Amir Jamal, former ambassador to the United Nations from Tanzania: "We're realizing rather late in the day that we sold ourselves short."

Traditionally, biological resources have been considered "the common heritage of mankind." In practice, that has meant that industrialized countries took freely of those resources—and sold them back to the Third World as commercial products. As a result, nations barely decades past colonialism are drawing some uneasy parallels with their earlier history. Moreover, biotechnology is a strategic science for the next century—with developed countries already in the lead.

The biodiversity convention, while ruling that biological resources are the sovereign property of a nation—on the same footing as coal or oil—also requires that these resources be made available to prospectors. But guidelines for such access will take some time to draft. "Most Third World countries don't quite know what, or how, to control," Jamal says. Nor is there any consensus as to the intrinsic worth of a bioprospector's gleanings.

"Until you find an active chemical entity, this stuff has no value," points out David J. Newman of the National Cancer Institute, which screens soil, sponges and other natural substances for antitumor and antiviral agents. On the other hand, Walter V. Reid of the World Resources Institute in Washington, D.C., calculates that if 3 percent royalties are paid (typical for samples of unknown activity) and if one assumes that a chemical has a one-in-40,000 chance of becoming a drug worth $10 million, a soil sample with 100 new chemicals could be worth about $500 to the source country. Yet companies are not about to pay such amounts up front.

Microbes, at least, are relatively free of the intellectual-property questions that plague work on plants and seeds originally identified or bred by native cultures. Those companies that find a useful microorganism and develop a marketable product—a process often costing millions of dollars and lasting over a decade—can take credit for the discovery. Sometimes, though, the credit has to be shared. A strain of fungus reportedly used by Brazilian farmers to kill fire ants, which cause hundreds of millions of dollars in crop damage in the U.S., was patented by a scientist at the University of Florida.

While the question of how much a developing country should get for its raw materials is debated, drug manufacturers are sifting through the soil in their own neighborhoods so many times that they keep turning up the same microbes. Adding incentive to collecting in tropical climes, the Food and Drug Administration has made it harder for "me-too" drugs to be approved. These are medications that are very similar to a popular one, like the many variations of penicillin on the market. Radically new organisms are needed.

Those prospectors who yearn for jungle adventure may be disappointed. Microbes are as easy to pick up as diarrhea. (The source of one acquisition in the ATCC is listed as "stool of an Iowa man who had recently been in Bangladesh.") But until the terms of exchange between genetically rich countries and technically rich countries are settled, the more ethically inclined are unsure of how to proceed. "Meanwhile," says James D. McChesney of the University of Mississippi, "valuable microbial resources are disappearing before we get a chance to look at them." Tropical forests, after all, are even richer than tropical sewers.

—Madhusree Mukerjee

RAGHUBIR SINGH

CALCUTTA COURTYARD: How many patents can you spot in this picture?

For Sale: One Country, As Is

When new leaders came into power in eastern Europe and the states of the former Soviet Union, they found themselves saddled with an unwanted inheritance: tens of thousands of businesses, from shops to steel mills that encompassed entire cities. Central planning and government control of these myriad enterprises had led to the economic disaster that swept communist governments out of power. Yet transferring these state-owned companies to private hands is proving to be nearly as much of a headache as was trying to run them. “It’s not just a puff of smoke and a wave of the wand, and the companies belong to the citizens,” rues Anthony A. Repa, an economic adviser to the Polish government.

The problems of privatization are simple to state but tricky to solve. Most citizens have no money with which to purchase corporate shares, and many enterprises are in such perilous financial condition that few people would want to buy them. Even solvent businesses come encumbered with now irrelevant assets—apartment complexes and sports stadiums among them—that make valuation difficult.

Countries have taken various approaches to shedding state property, according to economists at the World Bank. In the territories of the former East Germany, a federal agency manages businesses and properties while trying to arrange their sale or return to former owners. In Russia and in the Czech and Slovak republics, in contrast, citizens have received coupons with which they can buy shares in state-held concerns. In Poland, privatization has followed several strategies, including focusing on foreign investors and on investment funds that act as proxies for citizens with coupons. Most of the emerging market nations have also taken a more informal approach to privatizing small businesses by selling them to their managers or by auctioning them.

Coupon schemes are popular because they avoid many problems that arise when state property is sold for cash. Nemat Shafik of the World Bank’s Central Europe Department lists some of these drawbacks in a recent report on the Czech experience: foreigners can easily outbid citizens; the few people who have money can acquire the lion’s share of formerly public assets; and the low price paid for firms because of general lack of capital distorts future market patterns. But coupons entail a different problem. Unlike cash sales, they generate no government revenue and no capital for modernizing a firm.

At the same time, unless special precautions are taken, coupon privatization may diffuse corporate ownership so widely that effective governance is impossible. In Russia, some kind of governance has been preserved by reserving a percentage of shares in each enterprise for managers and workers. The Czechs and Slovaks took another approach, according to Shafik. They encouraged the development of mutual funds to which citizens signed over coupons. In theory, the funds can keep a closer watch on each company than can individual shareholders.

The problems of privatization are simple to state but tricky to solve.

In Poland, this form of privatization, when it is implemented, will be somewhat more intricate. Citizens’ coupons buy shares only in investment funds; the funds, in turn, purchase companies and manage them. The share prices rise and fall depending on the market’s assessment of the companies they invest in, thus providing incentives for careful oversight. Only about 450 Polish enterprises, out of a total of about 9,000, will be subject to mass privatization, Repa points out. Other companies are being groomed for foreign sale or have been acquired by their managers in leveraged buyouts.

Although such subtle manipulations of market structure may help provide effective oversight of companies and give the public access to the capitalist dream of shared ownership, they do not necessarily address the problems of actual restructuring needed for survival. Almost all these businesses need capital for new investment, and many are burdened by the debts that piled up in the old days of central planning.

Firms that cannot survive in their present form may be subjected to what is called asset liberation (on Wall Street it is known as “rape, loot, pillage”). During this process, auditors strip out property or businesses that might fetch a good price on their own and discard the remainder. This tactic is explained by the rationale that it is far cheaper for the economy to shut a plant down and pay workers for two or three years until they can find other jobs than it is to pay people to lose money indefinitely.

Asset liberation is, however, a complex undertaking. The commingling of government and business produced innumerable company towns throughout eastern Europe and the former Soviet Union. In these sites an electric power company might own not only generators and oil refineries but housing for its employees, day care centers and casinos. Some of these adjuncts can be sold for a profit; others must be taken over by a state or municipality if the underlying company is to be sold profitably.

Ironically, remnants of the political liberalization that marked the early stages of the transition to market economies may be hindering the last stages of privatization. In their final years, central authorities ceded ownership of many enterprises to provincial or municipal governments. Towns and cities are even more strapped for resources than are national governments, so they are unwilling to bear the brunt of closings that would eventually benefit the economy as a whole.

In one instance, the economic transition seems to be occurring most successfully independently of state-led efforts. The private sector in Poland has been growing rapidly, contributing about 50 percent of the gross national product last year. “Poland is looking potentially like the economic tiger of the region,” Repa notes. “The growth of the private sector has been spontaneous. It has not been the result of privatization, but the result of a couple of million new businesses.” In fact, Repa explains, the private sector is encouraged by ongoing inefficiencies in the vestigial state-run sector: “In a competitive market, it gives them an advantage.” Whether jealous, unfrocked apparatchiks will permit that advantage to persist remains to be seen.

—Paul Wallich and Marguerite Holloway

Genetically Altering *Escherichia coli*

Cows that produce pharmaceuticals in their milk and plants that resist pests are only two of the myriad benefits promised by genetic engineering [see "Grading the Gene Tests," by John Rennie, page 88]. Whereas endowing the genomes of cows and plants with desirable traits can be difficult, genetically manipulating bacteria is fairly straightforward. You can insert a gene for resistance to penicillin into the bacterium *Escherichia coli* by following the steps outlined here. Similar transformations often occur naturally, as in hospitals where antibiotic-resistant bacteria sometimes proliferate.

Because *E. coli* is already present in your gut, there is little to worry about. But remember to adhere to the sterile procedures described. They will ensure that you do not inadvertently ingest the stuff—and that it is *E. coli,* and not something else, that is blossoming on your petri dishes.

The essential trick to manipulating *E. coli* genetically is to get the creature's single cell to think that a foreign gene is one of its own. In this case, we will graft the penicillin-resistance gene onto a plasmid from *E. coli.* Plasmids are circular loops of DNA that exist in cells independently of chromosomes. After being altered, the plasmids are able to reenter the cell and thus help us sneak our gene into the bacterium. They then replicate within the cell and produce proteins, including the ones that will bring about our desired trait. Furthermore, plasmids are propagated through successive generations. Thus, the gene is transmitted to all offspring, conferring its resistant properties to the entire colony of *E. coli* grown from the altered individuals.

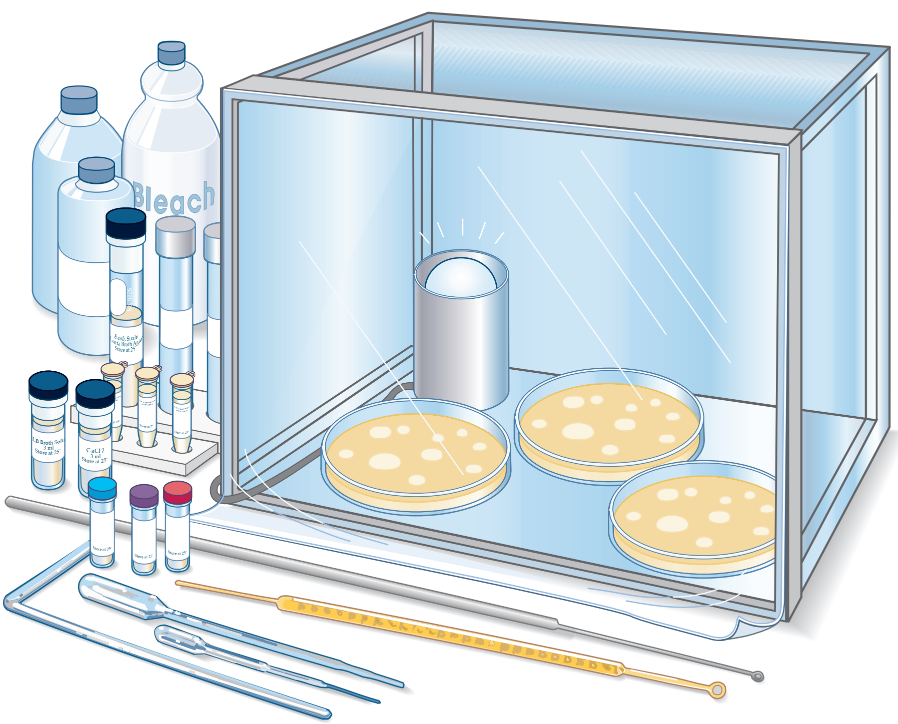
In this experiment, we will construct three types of recombinant plasmids, called pAMP, pKAN and pAMP/pKAN, by a process called ligation. By injecting these into *E. coli,* we will confer resistance for ampicillin (pAMP), for kanamycin (pKAN) or for both ampicillin and kanamycin (pAMP/pKAN) to the bacterium. To see the results of our engineering, we will then grow colonies of genetically altered *E. coli* on plates treated with ampicillin and kanamycin.

For the materials needed in this experiment, I suggest purchasing a DNA recombination kit. In addition, you will need an aquarium (or some container of similar volume), a Bunsen burner and a few odds and ends listed in the box on the top of page 110.

First, we need to build an incubator in which to grow our *E. coli* colonies. A 20-gallon glass aquarium will do nicely; if you do not have one, use any container of the same volume, such as a cardboard box. Place the aquarium on its side with the open end facing you. Tape a piece of plastic to the top of the aquarium so that it drapes down, covering the open end. You will need to lift the plastic up and out of the way to work inside the aquarium.

The optimal temperature for *E. coli*'s growth is that of the human body, 98.6 degrees Fahrenheit—not surprising, given where it lives. To warm the incubator, put in a standard incandescent lamp enclosed in a can or small pail. I needed a 75-watt bulb to heat the incubator to 92 degrees F (this is less than the optimal temperature but works fine). Start out using a low-power bulb—say, 40 watts—and measure the temperature after 12 to 24 hours; increase the wattage if necessary. If the incubator becomes hotter than 98.6 degrees F, lower the wattage—a bit cooler is better than a bit hotter. If changing the bulb does not do the job, insert a light dimmer and use it to adjust the power to the lamp and consequently the temperature.

It is a good idea to have the incubator ready when the kit arrives. Open the kit and refrigerate the culture plates upside down, along with the vials of plasmid pAMP, pKAN and calcium chloride (needed for conditioning the bacteria). Freeze the vial of ligase/ligation buffer/ATP. The other materials can be stored at room temperature. Just before starting the experiment, wipe the incubator and the work area with a 10 percent bleach solution (made by adding one part of Clorox liquid bleach to nine



ESCHERICHIA COLI ***blossom on petri dishes in an incubating aquarium. Culture tubes, pipettes and other materials await the experimenter.***

JOHN IOVINE has published numerous books on science and electronics. He currently teaches a workshop on holography for novices at the Art Lab on Staten Island, N.Y.

Supplies for Genetic Alteration Experiment

E-Z Gene Splicer DNA Recombination and Transformation Kit. PN# 21-1160; $35.67

The kit includes:

- One vial of plasmid pAMP
- One vial of plasmid pKAN
- Cultured *E. coli*
- Three LB plates
- Two LB/AMP plates
- Two LB/KAN plates
- Two LB/AMP/KAN plates
- Three vials of ligase/ligation
- One vial of calcium chloride
- Twelve needle-nose pipettes
- Eighteen one-milliliter sterile transfer pipettes
- Four 15-milliliter sterile culture tubes
- Five sterile inoculating loops
- Glass cell spreader
- Manual

Available from:

Carolina Biological Supply Company
2700 York Road, Burlington, NC 27215 (910) 584-0381

In addition to the materials provided in the kit, you will need:

- Alcohol lamp or Bunsen burner
- Antibacterial soap
- Aquarium or box
- Beakers or bowls (two)
- Crushed ice
- Distilled water (small quantity)
- Ethanol alcohol (70 to 95 percent)
- Felt-tip marker for labeling
- Household bleach, such as Clorox (small quantity)
- Sheet of plastic for covering aquarium
- Thermometer

parts of water) or a disinfectant, such as Lysol. Also disinfect after each session and wash your hands with an antibacterial soap both before and after. Keep your work area spotless and disinfect all the hardware, such as tubes, pipettes and transfer loops, by placing them in the bleach solution after use.

Our next step is to incubate and grow *E. coli* bacteria on a culture plate. The culture plates contain Luria broth, or LB, which provides nutrients on which the bacteria thrive. We will need these *E. coli* colonies for the rest of the experiment. Light the alcohol lamp or Bunsen burner. Take one culture plate labeled "LB" and mark on its bottom "*E. coli.*" Write the date on it, too.

Select the inoculating wire loop from the kit and sterilize it by putting it into the flame of the lamp or the burner. Allow the wire to get red-hot, then remove it from the flame and hold it for a few seconds to cool. Do not put the loop down—that will contaminate it.

Place the vial of *E. coli* culture in your other hand and remove its cap with the little finger of the hand holding the inoculating loop. With the cap removed, briefly pass the mouth of the vial through the flame to sterilize it.

Lift the top of the LB plate marked "*E. coli*" just enough to insert the inoculating loop. Push the loop into the side of the jelly on the plate to cool it. Next drag the loop a few times through the *E. coli* culture in the vial. Remove the loop, pass the mouth of the vial through the flame again and recap it.

Drag the loop across the jellylike surface, making a Z shape in one quadrant of the plate. Replace the lid.

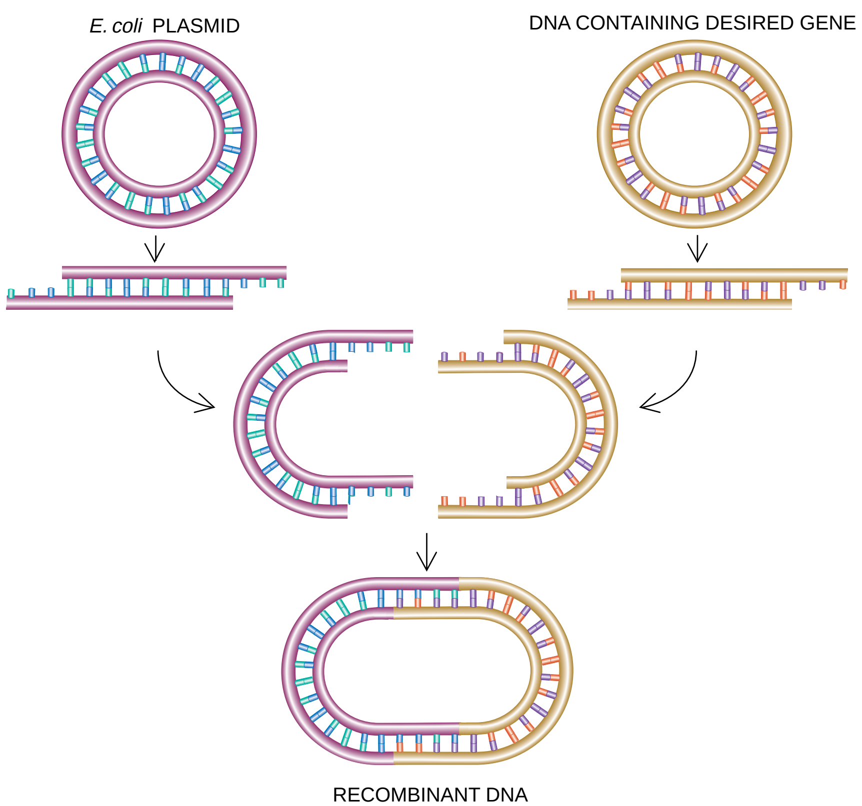
Turn the culture plate 90 degrees, reheat the loop and cool it by stabbing at the gel away from the first streak. Then pass the loop once through the first streak and make another zigzag shape. Repeat the turning, reflaming and streaking another two times so that there are four zigzag shapes, one in each quadrant of the culture plate. Replace the lid on the plate.

Reflame the loop before putting it down, so as not to contaminate the work space. Make this a habit.

Place the smeared plate upside down in the incubator to prevent condensation that may collect on the lid from falling into the *E. coli* colonies. Incubate the plate for 12 to 24 hours—no more. Then remove it from the incubator and allow the colonies to grow for one or two days at room temperature.

The next step is to link antibiotic-resistant DNA fragments with the *E. coli* plasmids. The vials of pAMP and pKAN contain the indicated DNA fragments as well as the plasmids; a reagent called ligase inserts the DNA fragments and reseals the plasmid loops. (ATP in the ligation solution provides energy for the joining reaction.) The procedure actually makes many different types of hybrid molecules; however, only those that are properly formed will be maintained and expressed in the cells.

Take the three vials containing 20 microliters each of ligase/ligation buffer/ATP from the freezer. Label one vial "+pAMP/KAN," another "+pAMP" and the last one "+pKAN." Take out the tubes labeled "pAMP" and "pKAN." Using a sterile needle-nose pipette—one for each reagent—add 10 microliters of pAMP and 10 microliters of pKAN to the +pAMP/KAN vial; 10 microliters of pAMP and 10 microliters of distilled water to the +pAMP vial; and 10 micro-



RECOMBINANT DNA is made when a plasmid from* E. coli *and a circular piece of DNA are cut into strips by enzymes and then rejoined. The ends are sealed by ligase; only those recombinants that are properly oriented will survive in a cell.

liters of pKAN and 10 microliters of distilled water to the +pKAN vial. Close the tops of the vials and gently tap their bottoms on the table to mix the reagents. Incubate the vials at room temperature for 12 to 24 hours. Now the +pAMP/pKAN vial contains plasmids for resistance to both ampicillin and kanamycin, the +pAMP vial only those for resistance to ampicillin and the +pKAN vial only those for resistance to kanamycin.

Now we must make the *E. coli* cells "competent" to absorb the recombinant plasmids. This requires suspending the *E. coli* cells in a cold calcium chloride solution and subjecting them to a brief heat shock at 107.6 degrees F. Just how DNA is absorbed by competent *E. coli* cells is not known.

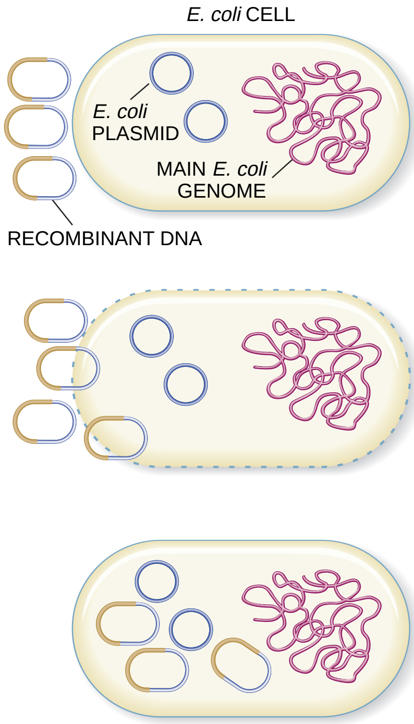
Prepare a water bath for heat-shocking the bacteria; you will need the bath just once for 90 seconds. You can use an aquarium heater to bring the water in a container to 108 degrees F (a couple of degrees more or less should not matter). In a pinch, just run in tap water, adjusting the temperature.

Get four sterile 15-milliliter tubes. Label them "+pAMP/KAN," "–pAMP/KAN," "+pAMP" and "+pKAN," respectively. With a sterile transfer pipette, add 250 microliters of cold calcium chloride to each tube. Place the tubes in a beaker or bowl with crushed ice.

Pick up one or two colonies of *E. coli* from the LB starter plate using a sterile plastic inoculating loop. Be careful not to take any gel from the plate. Immerse the loop in the calcium chloride solution in the +pAMP/KAN tube and tap against the side of the tube to dislodge the cell mass. Suspend the cells in the solution by repeatedly pipetting in and out with a sterile transfer pipette. Return the +pAMP/KAN tube to the ice. Transfer cell colonies to the other three tubes in the ice in the same way.

For each of the following steps, use a a fresh needle-nose pipette. Transfer 10 microliters of ligated plasmid +pAMP/KAN from the appropriate vial to the +pAMP/KAN culture tube. Add 10 microliters of ligated +pAMP to the +pAMP culture tube and 10 microliters of ligated +pKAN to the +pKAN culture tube. Do not transfer any material into the –pAMP/KAN culture tube; this last tube should contain only unaltered *E. coli.* Place all the tubes back in the ice and let them sit for 15 minutes.

After the ice incubation, it is time to heat-shock the *E. coli* cells. Remove all the tubes from the ice and immediately immerse them in the 108-degree F water bath for 90 seconds. Then return all the tubes directly to the ice. Keep them there for three to four minutes.



ANTIBIOTIC-RESISTANT DNA hitches a ride into an* E. coli *cell by hiding in an altered plasmid. Plasmids cannot enter the cell* (top) *until it is made receptive* (middle). *On entering, the recombinant DNA is accepted as part of the genetic material of the cell* (bottom). *It can produce proteins and also propagate.

Using a sterile transfer pipette, add 250 microliters of Luria broth to each tube, to give your bacteria something to eat. Gently tap the tubes with your finger to mix in the broth and incubate the tubes at 98.6 degrees F for three to six hours. With luck, the *E. coli* should now be sufficiently shocked to absorb the plasmids in their environment.

Finally, we can check to see if our *E. coli* have indeed acquired resistance to ampicillin and kanamycin. Some of the culture plates in the kit already contain the antibiotics to be pitted against our *E. coli* and are marked as such. The plates marked "LB" contain only Luria broth; label one such plate "+LB" and the other "–LB"—on these plates we will grow the altered and unaltered *E. coli,* respectively. Label one LB/AMP/KAN plate "+pAMP/KAN"—to this plate, containing both ampicillin and kanamycin, we will be adding the *E. coli* cells resistant (we hope) to both antibiotics. Label the other LB/AMP/KAN plate "–pAMP/KAN"—only unaltered *E. coli* bacteria are to be added here. Label one LB/AMP plate "+pAMP" and the other "+pKAN"—to these (ampicillin-treated) plates we will be adding *E. coli* resistant to ampicillin and kanamycin, respectively. Label one LB/KAN plate "+pKAN" and the other "+pAMP."

Add 100 microliters of the cell suspension from the –pAMP/KAN culture tube to the –LB plate and also to the –LB/AMP/KAN plate, using a sterile transfer pipette. Before smearing the cells over the surface of the gel, sterilize the glass spreader. Dip the spreader in the ethanol alcohol and ignite the alcohol using the Bunsen burner or alcohol lamp. After the alcohol burns off, the spreader is ready. Use it to distribute the *E. coli* cells evenly over the gel.

With another sterile transfer pipette—use one for each culture—add 100 microliters of the culture from the +pAMP/KAN tube to the +LB plate and also to the +LB/AMP/KAN plate. Spread the cell suspension as before, sterilizing the glass rod each time. From the +pAMP culture tube, add 100 microliters each of the cell suspension to the +pAMP and +pKAN plates, then spread. From the +pKAN culture tube, add 100 microliters of the cell suspension to the +pAMP and +pKAN plates, then spread.

Allow the plates to gel for 10 minutes, then stack them and tape them together. Place the plates upside down in the 98.6-degree F incubator. Incubate them for 12 to 24 hours. Note that if the colonies are overincubated, they will overgrow and become indistinguishable.

Now you are ready to see the results of your experiment. On the +LB and –LB plates, both the transformed and the natural *E. coli* grow well. For the LB/AMP/KAN plates—containing both ampicillin and kanamycin—there is a growth on the one labeled "+AMP/KAN" and none on the "–AMP/KAN," showing that only the *E. coli* transformed with both antibiotic-resistant genes can grow. The LB/AMP plates illustrate that the gene for pAMP, and not that for pKAN, confers resistance to ampicillin. The LB/KAN plates illustrate that the gene for pKAN, and not that for pAMP, confers resistance to kanamycin.

By measuring the growth of the colonies, you can see that the ligation of two genes, pAMP and pKAN, is more difficult than a single ligation: the +pAMP/+pKAN colonies are sparser (five to 50 colonies) than are the +pAMP or the +pKAN colonies. Also, ligation of the pKAN gene is more difficult (50 to 500 colonies) than that of the pAMP gene (100 to 1,000 colonies).

Now that you have genetically altered bacteria, you can see that in principle they are quite simple to produce. In practice, genetic engineering can be somewhat problematic. For instance, cows and plants will need much bigger incubators.

BOOK REVIEWS by Philip Morrison

Albert *en Famille*

EINSTEIN LIVED HERE, by Abraham Pais. Oxford University Press, 1994 ($25).

Not even Albert Einstein remains unrevised these iconoclastic days, the attackers by now more than a little shoddy. Thus, very welcome is this book by the most thorough and judicious of physicist-biographers, who was himself Einstein's young companion during nine Princeton years of friendly colloquy on all the world. Abraham Pais has just assembled a wonderfully readable look (no math) at that great man and at ourselves viewing him. He offers a four-page "mini-briefing," but no more, on Einstein's physics in this short, often riveting collection of personal but always well-supported studies. For example, a dozen pages sample just one file from the voluminous Einstein archives, the one Einstein called "Die Komische Mappe," perhaps best translated as the "curiosity file." In that Dickensian richness of letters to the famous man are the touching, the enraged and the funny: "My thirteen… talking cats have just told me what the fourth dimension is."

Timeliest of the short pieces is the opener. It examines Einstein *en famille,* with the facts that constrain a new "minor industry" that has arisen around a Serbian biography of Mileva Marić, Einstein's first wife and mother of his three children. The book, *In the Shadow of Albert Einstein,* promotes the claim that it is Mileva, Albert's fellow student of physics, to whom we really owe the relativity that arrived in the wonderful year 1905. Pais explains how a misunderstanding of one Russian citation may have been the taproot of the wistful ascription so long after the fact. In 1905 Einstein signed his startling manuscripts with his new wife's family name added after a hyphen, then a custom in Switzerland. But there was only one author. Mileva herself never made any claim whatever for a share of Albert's fame during all three decades of open estrangement.

The delightful letters between the young pair—more are yet to be published—are full of pet names, largely out of Albert's native Swabian dialect. Their first child, always called Lieserl, was born out of wedlock somewhere near Mileva's home in Serbia, sometime around the turn of 1901 to 1902. The couple were married a year later in Bern. The infant Lieserl never lived with them there; "nothing I have seen indicates that Einstein ever even saw her." In 1903 Mileva went east again to visit her family; she was pregnant with their first son, Hans Albert. Einstein wrote her from Bern in concern over some report about Lieserl's coming down with scarlet fever. No later reference to the daughter is known in all the correspondence, and serious efforts on the spot have turned up no other documentation.

The couple became estranged step by step over 10 years. After Einstein became professor in Berlin, Mileva, ailing and melancholic, took her two boys off to Zurich, where she would live from 1914 until her death. She and Einstein were first formally separated, then divorced in 1918. Little doubt remains that the first marriage had been darkened by Albert's relationship with his cousin, Elsa. Three years older than Albert, she became his second wife in 1919, bringing her two daughters with her. It was something of a mismatch; the scholar who rejected all poses came to resemble "a bohemian guest in a middle-class home"; Elsa's apartment with its fine carpets, furniture and pictures was far better suited to her and her girls than to her new husband. "I know for a fact that in the early 1920s Einstein developed a strong attachment to a young woman," an extramarital affair that endured for a couple of years. It ap-


SECOND NEWS SECTION

For News, Subscriptions or Advertising Call ATlantic 6100.

SATURDAY MORNING,

Pittsburgh Post-Gazette

DECEMBER 29, 1934.

SPORTS, FINANCIAL, CLASSIFIED SECTION

This WORLD of OURS

Constantly Expressive—Einstein Warms Up to His Subject

ATOM ENERGY HOPE IS SPIKED BY EINSTEIN

Efforts at Loosing Vast Force Is Called Fruitless.

SAVANT TALKS HERE

Now Indicates Doubt Of Relativity Theory He Made Famous.

EMILIO SEGRÈ VISUAL ARCHIVES, AMERICAN INSTITUTE OF PHYSICS

ALBERT EINSTEIN talks to reporters about atomic energy. From the December 29, 1934, **Pittsburgh Post-Gazette.**

pears, moreover, not to have been the last of such affairs, as more than one new book will eagerly magnify for you.

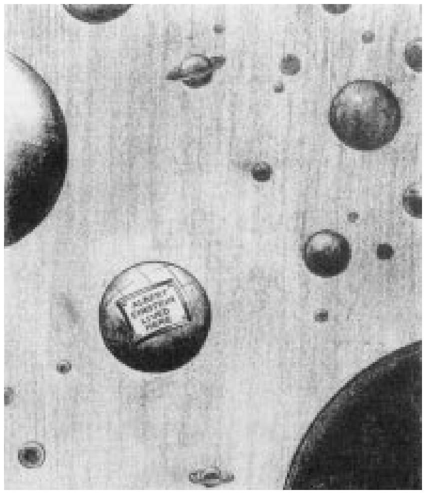
Elsa died in Princeton in 1936. "I... live like a bear in a cave," Einstein wrote then, the more because of "the death of my comrade, who was more attached to people [than I]." Einstein had insight into his marriages; he had not been a good husband, and he knew it. To live in "lasting harmony with a woman" was "an undertaking in which I twice failed rather disgracefully."

The first son, Hans Albert, became a successful professional engineer in Europe, then served from 1947 to 1971 as a prominent professor of hydraulic engineering at the University of California at Berkeley. Deeply attached to the mother with whom he had grown up, Hans was not always on good terms with his father: "Probably the only project he ever gave up on was me. He tried to give me advice, but...I was too stubborn." Near the end of his father's life, he visited in reconciliation. Hans died of a heart attack at age 69 and lies buried on Martha's Vineyard; the grave marker recalls "a life devoted to his students, research, nature, and music."

The younger son, Eduard, called Tete, born to Mileva and Albert in 1910, suffered from schizophrenia, his illness clearly visible only after he had finished high school admirably. Thereafter he lived precariously in and out of institutions, until in 1965 he died in a psychiatric center near Zurich. Poor Mileva had died in 1948 after long ill health. Afflicted for years by Eduard's tragedy, she became lonely, even paranoid, dying with a small fortune in Swiss francs stuffed into her mattress. One Einstein dwells quietly in Bern still, Hans Albert's son, Bernhard.

Almost half the book is an engaging inquiry into the lifelong interaction between Einstein and the media, where "the show goes on." Professor Pais seeks to answer how Einstein became so singularly famous that for decades his mere presence ensured an intolerable crowd of gazers. He was a prisoner of that disturbance and could hardly leave his Princeton neighborhood in any public way. There is no doubt the turmoil began right after World War I, when the British mounted an eclipse expedition to seek Einstein's predicted deflection of starlight by the sun. The effort glowed worldwide as a sign of healing after a terrible war and as a promise of a new order.

An air of mystery about Einstein's "space-warp"—the very words seemed at once simple and paradoxical—soon surrounded the man. In 1921 he told a Dutch newspaper: "I am sure that it is



CARTOON by Herblock appeared in the **Washington Post** *several days after Einstein's death on April 18, 1955.*

the mystery of non-understanding that appeals to them." A reader sees the point but is not quite persuaded. Anyway, it wasn't his photogenic looks: in the 1920s Einstein was still a "friendly...pot-bellied gent, dressed in a... bourgeois way." No sweatshirts yet. Among the two dozen images reproduced, many are fresh and striking: his first childhood photograph, the 1903 wedding picture, the two young sons.

That formula become a logo, $E = mc^2$, has a more understandable rise, the consequence of a pedagogic invention by physicist Harry Smythe as World War II ended. He cited it on the first pages of his celebrated official 1945 report on the secret Manhattan Project, setting out Einstein's relation as the key to the release of nuclear energy. Good physics, though quite misleading history, the idea penetrated beyond any teacher's dream, until now, alas, it is less an explanation than a mantra.

One firm judgment demands repeating. Einstein was deeply honest, unpretentious and caring for the future, but he was never politically naive. His candor looks the better as time passes, bringing hope for compassion and reason here where Albert Einstein lived.

The Whale as Hot-Oil Balloon

AIR AND WATER: THE BIOLOGY AND PHYSICS OF LIFE'S MEDIA, by Mark W. Denny. Princeton University Press, 1993 ($39.50).

In 1988 this Stanford biologist, a talented reverse-engineer of life, from spiderwebs to surf-swept kelp forests, was enlisted to lecture his fellow zoologists on the distinctions between water and air. Original, straight-talking and enthusiastic, Denny has built out of that experience an unusual book for biology students and any serious readers who have even a rusty grasp of freshman math.

"Every day I find a new...biological example of physics in action." The content and heft of his book therefore duly recall a pretty full physics text. The early chapters open with units, Newton's laws, stress and energy. The next few elaborate on fluid pressure and flow. Diffusion, less commonplace, receives a good account. Then we read on to heat, sound, optics, waves. There are plenty of equations, though not much math, for intermediate steps are fully presented. Graphs and tables of the most varied data abound.

All the results have one purpose, a quantitative understanding of the ways of life in two fluids. Biologist Denny intends his readers to build coherent physical intuition; formulas are indispensable but insufficiently nourishing for that. The book tilts to clarity and to examples; derivations that are physical are given pride of place, others omitted. The wave equation is absent; in its stead a careful physical derivation of sound speed is offered. Random walk and the physics of diffusive transport are treated without the heavy statistical weapons.

Life is many-faceted, its systems diverse. That keeps interest high and assures the width of the account. Relativity, quarks and other extensions of physics far beyond everyday experience are forgone; you can't have it all in one book. But what you do have is a well-organized process that analyzes in context about 100 substantial issues of how and why life works as it does in water and in air, as a hopeful designer might address them. The action is in scale, form, motion, energy, field, detection and simple chemistry. Informational issues are offstage; necessity is here well ahead of chance. We are among physiologists, newly wide-ranging ones. A sample of topics follows. Denny works them out in first approximation, no number crunching, unabashedly willing to invoke the spherical organism.

Since water-dwellers match water density rather well, could they adjust to float or sink by raising or lowering body temperature slightly, on the model of the hot-air balloon? A thin-walled sea "balloon" would need only a fraction of a degree rise to equal the density change its airborne counterpart gains only by heating internal air through a scorching 100 degrees Celsius. But the energy cost is high, losses strong within dense water. So there are no living hot-air or hot-water balloons. There is one hot-oil balloon: surprisingly, it is the huge sperm whale. Its unique spermaceti organ holds two or three tons of a special lipid mix, whose density falls as its temperature rises 10 or 20 times as much per degree as does the density of water. Fifty-ton Moby Dick can adjust buoyancy at will by 100 pounds or so, simply by heating or cooling slightly his headful of sperm oil, a task for which he has several means in place.

Corals deep in the reef feel strong buoyant forces and little gravitational load. Yet redwoods grow 10 times as high as corals. The answer is dynamical: steady flow produces forces of lift and drag on solid bodies (well worked out here), but there are also reaction forces from accelerating flow of ambient fluid. Reaction forces limit the length of a rooted column in proportion both to the fluid density and to its maximum acceleration. Accelerations in the surf reach measured values near 40 gravities. If trees endured the same reaction limit, they could withstand superhurricane accelerations of 3,000 gravities. Plainly these never occur in the atmosphere. Trees are not limited in height as corals are by major unsteady flow but mainly by their instability to buckling under their own weight. Talk of this with any sailor!

A physicist reader finds this work evokes pride for his science and envy for the biologists who raised these new problems and then found out how to solve them simply. It deserves many readers, and even more it would grace the shelves of long-term browsers.

Of Flies and Men

HUMAN LONGEVITY, by David W. E. Smith. Oxford University Press, 1993 ($35).

Pathologist and molecular biologist, David W. E. Smith of Northwestern University has for most of a decade centered his wide interests on human longevity. He synthesizes his conclusions—and his doubts—in this brief, firm and eye-opening book. The facts of life and death are never straightforward, even when they are accounted for numerically. Their meaning is treated widely, within medicine and human pathology, in the biology of other forms of life, in the sciences of human behavior and in human evolution. Five chapters discuss these topics, one focused on the intrinsic and ubiquitous difference between longevity in men and in women. Everywhere (save in South Asia and Iran), the men die younger even though women are sicker, perhaps to a degree from medical inattention. The three final chapters treat evolution and offer a forecast of the future.

Sooner or later we animals all die, but that is not so for all individual living cells. The fission of bacteria proceeds indefinitely two for one. Then, the saying goes, "with sex came death." Cells like our own show clonal senescence: a cell and its progeny do not continually divide. They need a new deal, some meiotic dance. Many human cell lines have been cultivated in tissue culture until growth stopped, after 50 doublings more or less. But "cancer cells are immortal." Tumor cell lines have been maintained for decades, still doubling every day or two.

What dies is an organism, an interdependent society of cells, human or fruit fly. The little fly passes egg and larval stage, to reproduce and die after a few weeks as adult. The causes of fly death are not fully known, but often it is the cells that line the active digestive tract that stop working, clogged by a pigment derived from massive membrane breakdown. (In mammals, unlike the flies, the digestive system cells are capable of replacement.) Then the fly neurons, never plentiful, drop gradually out of action. "Old flies die constipated and largely immobilized."

The curve of dying gives statistics decisive for fly or human. Follow a sample of our kind and count deaths year by year. In developed lands today, most lives come to an end within a rather narrow band of ages, the rate peaked near the expected average lifetime. Small birds and rodents, as well as "drinking glasses washed by machine in a restaurant," show an utterly contrasting pattern. They die by random causes, death taking a near-constant fraction of survivors left at any age, until they have all gone.

A more complex intermediate pattern begins with a high death rate in early life—infant mortality. The rate then falls steeply, flattens out, to rise again slowly, then more and more rapidly to a wider, earlier peak, until the sample runs out. The change from intermediate to modern form is seen in action in the curves shown for the U.S. population at 20-year intervals since 1900. By 1980 the infant death rate is very much smaller, and the high, narrow end peak dominates. The developed countries, and a few others as well, have entered the late-peak phase. (This is only part of the demographic transition; full population studies must of course take account of births, too.)

It is unlikely that our species ever had a survival pattern as random as that for machine-washed tumblers. A natural maximum age seems to be built in among species that have larger brains, mature late and care for their young, particularly those, like primates, with few multiple births. "The human species is unique," the farthest of all out along that axis. A maximum life span is not merely statistical; humans live twice as long as other large mammals. The present record human age was reached by a man of 120 years (with only a touch of doubt about its validity).

In the U.S. now there are 30,000 centenarians among the 1 percent of us—2.5 million people—who are older than 85 years. These people are real outlyers, not merely there on the tail of a smooth mortality curve but distinct. (One mouse from a particularly well-cared-for strain outlived all his kinfolk, too, by about 20 percent.) It is probable that we have been a long-lived species for 100,000 years, even if death came much earlier to most. Turtles and clams and redwoods belong in another category. Indeed, the author does not mention the curious differences even among mammals in the rate of their life processes; a mouse heart beats during a mouse lifetime about as many times as an elephant's does, but in only a few years. People much outlast them all in total pulse count.

We all know that the chief agents of death are no longer random infections that may enter at any age, but long-delayed functional failures in heart or brain, where many cells cannot be replaced, just as in the fruit-fly gut. Does our end resemble the aged fly's? Hardly—our failures show no steady loss of cell by cell until too few are at work but rather an acute loss of whole clusters of cells whose blood supply has ceased.

If there is a simple picture to be painted, it could be that the constellation we fear—ischemic heart disease, cerebrovascular calamity and the common malignant tumors—may depend on the varied human diet and our complex hormonal management of cholesterol. Wear and tear, essentially incremental functional loss, may bring the end in diverse ways, once the aging feedback responses encounter challenges they can no longer meet.

But the story we know today is not complete. A "scientific rationale may emerge," able not only to delay mortality as we can now often do but even to postpone the senescence that causes the disabilities of aging. That optimism ends, a little unexpectedly, the author's sobering insights. Just where we should look in the long genomic message he cannot tell.

The Art of the Scientific Insult

At seminars and conferences, physicist Wolfgang Pauli was famous for dismissing work he particularly disliked as "*ganz falsch,*" completely false, or, yet more damning, "*not even* false." Pauli was not the only practitioner of the deft put-downs the British call pinking. As a group, scientists show a surprising historical talent for the insult. Beneath the impersonal surface of their formal scientific discourse lies a vast subtext of bloodletting.

Isaac Newton, for instance, was a scathing polemicist who honed his attacks on Robert Hooke and Gottfried Wilhelm Leibniz to razor-sharpness in successive drafts, each of them, according to a biographer, more offensive than the last. "M^r Hook thinks himselfe concerned to reprehend me for laying aside the thoughts of improving Optiques by *Refractions,*" Newton wrote in one salvo. "But he knows well y^t it is not for one man to prescribe Rules to y^e studies of another, especially not without understanding the grounds on w^{ch} he proceeds." When Leibniz published his calculus without acknowledging what he knew of Newton's progress, Newton attacked, sometimes through John Keill (Johann Bernoulli referred to Keill as "Newton's toady"), sometimes through a committee of the Royal Society. He even tucked one beautifully compressed accusation of what was then known as plagiary into the mathematical footnotes of *Commercium epistolicum:* "Thus the method which earlier he [Leibniz] wanted, asked for, received, and understood with difficulty, he discovered forsooth...."

Priority fuels the bitterest assaults, but ignorance, particularly when scientists move outside their area of expertise, also provides a rich source. For instance, the pugnacious Hermann Kolbe, a major 19th-century research chemist with an unfortunate lack of interest in optical activity, used the *Journal für Praktische Chemie* to ridicule Jacobus Henricus van't Hoff, who had clarified the relation between optical activity and molecular structure in a brilliant pamphlet. "A Dr J. H. van't Hoff of the veterinary school at Utrecht, finds, as it seems, no taste for exact chemical investigation," Kolbe wrote. "He has thought it more convenient to mount Pegasus (obviously loaned at the veterinary school) and to proclaim in his *La chimie dans l'éspace* how during his bold flight to the top of the chemical Parnassus, the atoms appeared to him to have grouped themselves throughout universal space." Van't Hoff had his revenge; he reprinted Kolbe's jibes in the second edition of *La chimie,* guaranteeing Kolbe's historical reputation as a fool.

Chauvinism, too, gives impetus to the scientific put-down. The French scientist Marie Jean Pierre Flourens said in his review of *The Origin of Species,* "...[W]hat unclear ideas, what false ideas!... O lucidity! O French stability of mind, where art thou?" And the Swedish chemist Jöns Jakob Berzelius, who took to studying textbooks other than his own when confined to bed by gout, wrote, "The chemists of England live in their own world.... There is a great deal of litigation here about priority in the most petty matters.... one can regard them as one would puppies who stand and snarl over their bones, from which the meat has been gnawed on the continent."

Singularly unprophetic remarks abound in the history of science, from the comments of Charles Darwin's doctor father to his son ("You care for nothing but shooting, dogs, & rat-catching, & you will be a disgrace to yourself & all your family") to those of Albert Einstein's headmaster when asked what profession Einstein should follow ("It doesn't matter; he'll never make a success of anything"). Elihu Thomson, electrical engineer of note, visited Thomas A. Edison's workshop and told the newspapers that he "did not think very highly of the Edison lamp and expected no great future for it." J. Louis Agassiz, Harvard University professor and naturalist, maintained that he would "outlive this mania" of evolution, which he characterized in the *American Journal of Science and Arts* as "a scientific mistake."

Press conferences eventually sped up the tempo of scientific sniping, infighting and backstabbing once conducted in the main through debates, letters, symposia and book reviews. Edison used the newspapers in his well-publicized fight to fend off the new AC (alternating current) distributing systems that were far more powerful than his limited DC (direct current) stations. In what became known as the "War of the Currents," Edison invited reporters and guests in daily throughout 1887 to watch the stray cats and dogs of West Orange, N.J., totter onto tin sheets and be electrocuted by high-tension AC currents. Red-lettered pamphlets warned that "patent pirates" like George Westinghouse were bent on introducing these hazardous currents "into the American home." Despite Westinghouse's constant explanations that AC was received only at low voltages because of Stanley transformers, Edison and his colleagues were so successful in arousing the public that the eponym "to Westinghouse" was suggested as an alternative to "to electrocute."

In the present age of litigation, scientists are usually more circumspect than Edison. They consult attorneys before they talk, as did Steven E. Koonin, then chairman of the nuclear physics division of the American Physical Society, before speaking out on cold fusion at a meeting of that society in 1989. Advised to avoid the "F-word" (fraud), Koonin summed up Stanley Pons and Martin Fleischmann's cold fusion work with the deadly phrase "incompetence and perhaps delusion." Nathan S. Lewis, who also spoke, provided a list of pointed questions the press might want to ask Pons and Fleischmann, adding that if the two scientists were going to have publication by press conference, he would institute peer review by press conference.

Not all put-downs are scathing—some small part of them manages to be affectionate, as was the comment of an Edinburgh professor on an evening spent with Charles Babbage ("It was with the greatest difficulty that I escaped from him at two in the morning after a most delightful evening"). Some of them are even meant to be compliments, although, of course, they also pink. For instance, Pauli, already a master of dis-ing when only a graduate student, attended a seminar given by Einstein and generously acknowledged, "You know, what Professor Einstein says is not so stupid."

ANNE EISENBERG teaches at Polytechnic University in Brooklyn, N.Y.