SCIENTIFIC AMERICAN

APRIL 1994
$3.95

The dilemmas of prostate cancer.

The real culprit in U.S. economic ills.

Watching the Mind at work.

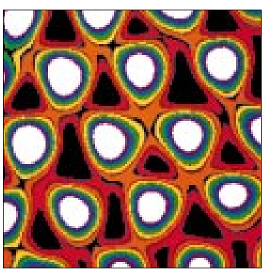
Ancient Peruvian *mask and headdress offer clues about a mysterious pre-Incan civilization.*

SCIENTIFIC AMERICAN
April 1994 Volume 270 Number 4

44 

Trade, Jobs and Wages
Paul R. Krugman and Robert Z. Lawrence

The sources of U.S. economic malaise are here, not abroad. Manufacturing as a percentage of GDP declines because consumers are buying more services and fewer goods. Manufacturing jobs vanish because machines replace workers; wages stagnate because productivity has slowed. These trends would persist even in the absence of foreign competition and the rise of a global economy.

50 

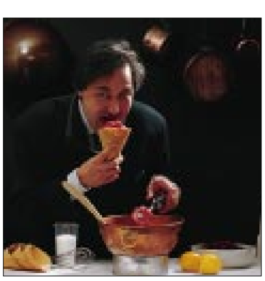
Charge and Spin Density Waves
Stuart Brown and George Grüner

In certain metals the lattice can affect the charge or spin of the electrons so that the particles organize themselves into crystalline arrays. When voltage is applied, the electrons, like the members of a marching band, all move together, maintaining their relative positions. Such systems provide an opportunity to study self-organized criticality, which is also evidenced in earthquakes and avalanches.

58 

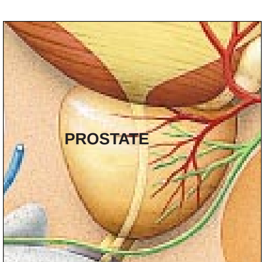
Visualizing the Mind
Marcus E. Raichle

In the hands of neurobiologists, MRI and PET imaging are revealing the physiological processes in the brain that underlie the mind. Monitored by scanning devices, subjects recall a word or generate a verb. As they perform such tasks, the flow of blood to various parts of the brain changes as each becomes engaged. The research presents some of the first images of the human mind at work.

66 

Chemistry and Physics in the Kitchen
Nicholas Kurti and Hervé This-Benckhard

Cuisine, *haute* and *bas,* has evolved for centuries in the form of an ephemeral art. Yet the materials are humble biological ones that respond in predictable ways to one another and to changes in temperature and pressure as time passes. Why should not the knowledge embodied in chemistry and physics be brought into the kitchen, where it can serve instinct and inspiration?

72 

The Dilemmas of Prostate Cancer
Marc B. Garnick

A common cause of death among men, this cancer has been detected with increasing frequency in recent years. Questions about the effectiveness of therapy cloud decisions about treating the illness. Older patients, if left untreated for small tumors, may die of other causes. But even if treatment is curative, many men—young and old—face impotence and incontinence as a result of therapy.

82 

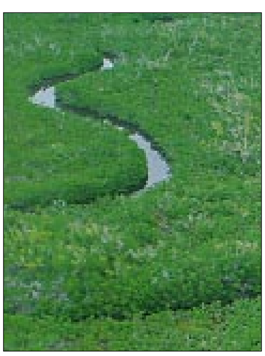
Precious Metal Objects of the Middle Sicán
Izumi Shimada and Jo Ann Griffin

Two aspects separate this fieldwork from other studies of the Sicán in Peru: the tomb had not been plundered, and an experienced goldsmith was among the workers. As a result, the find sheds clear light on a great native American culture, and the techniques used by the Sicán masters have deftly been reconstructed.

90 

The Pioneer Mission to Venus
Janet G. Luhmann, James B. Pollack and Lawrence Colin

Over 14 years, the multiple components of *Pioneer Venus* gathered vast stores of information about our sister planet. Engineering resourcefulness came together with scientific creativity in a synergy that lifted understanding of Venus in particular and planetary physics in general to unexcelled levels of sophistication.

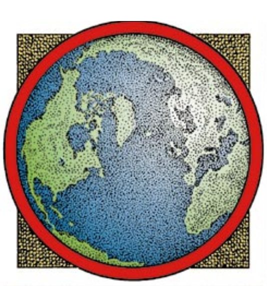
98 

TRENDS IN BIOLOGICAL RESTORATION

Nurturing Nature
Marguerite Holloway, staff writer

Florida's Everglades are serving as testing ground—and battlefield—for an epic attempt to restore an environment damaged by human activity. As conservationists, officials and commercial interests square off, the practitioners of the fledgling technology of biological restoration attempt to bring back a wounded ecology. Can they succeed here or elsewhere? How can success be measured?

DEPARTMENTS

14 

Science and the Citizen

Perry for Defense.... Epidemic endometriosis.... Bang! You're alive.... Quantum computer.... Hedgehog genes.... Pinning the flux.... Scientific silliness.... PROFILE: Edward O. Wilson revisits sociobiology.

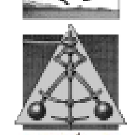
110 

Science and Business

Japan Inc.'s listening posts.... Texas Instruments does it with mirrors.... Softwars.... Cheap solar cells.... Architecture that shakes quakes.... THE ANALYTICAL ECONOMIST: Hospital profits continue to cool.

10 

Letters to the Editors
April foulers bag 1993's howlers.... Highfield's name clarified.

12 

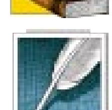
50 and 100 Years Ago
1894: Electricity at home.
1944: No war dividend.

120 

The Amateur Scientist
Professor Kurti and Monsieur This present the scientific soufflé.

124 

Book Reviews
Threads of the urban fabric.... Art of hard copy.... Star photographer.

128 

Essay: *Anne Eisenberg*
Emoticons and other artifacts of the new Epistolary Age.

THE COVER photograph shows for the first time a recently reassembled Sicán mask and headdress. Because of extensive looting of the Sicán tombs, no other assemblage of this type is known to have survived the melting-pot fate of most of the stolen artifacts. The Sicáns flourished before the Incas, from A.D. 700 to 1300, in northern Peru. They produced huge numbers of gold objects, many showing remarkable technological and aesthetic sophistication (see “Precious Metal Objects of the Middle Sicán,” by Izumi Shimada and Jo Ann Griffin, page 82).

THE ILLUSTRATIONS

Cover photograph by Yutaka Yoshii

Page	Source
44–45	AP/World Wide Photos
46	Culver Pictures, Inc. (*top*), Dimitry Schidlovsky (*bottom*)
47	Peter Yates/SABA
48	Dimitry Schidlovsky
49	Comstock, Inc./ Jim Pickerell
51	Robert V. Coleman and C. Gray Slough, University of Virginia
52–55	Jared Schneidman/Jared Schneidman Design
56	Jared Schneidman/JSD (*top*), Comstock, Inc./ Georg Gerster (*bottom*)
58–59	Jonathan D. Cohen, Carnegie Mellon University
60–62	Marcus E. Raichle
63	Guilbert Gates
64	Rodolfo R. Llinas, New York University Medical Center
66–67	Steve Murez/Black Star
68	Andrew Paul Leonard/APL Microscopic (*left*), Dana Burns-Pizer (*right*)
69	Dana Burns-Pizer
70	Paulette and André Lacour, INRI
71	Steve Murez/Black Star
73	Michael Grecco/Sygma
74	Tomo Narashima
75	Johnny Johnson
76	Dimitry Schidlovsky
78	Johnny Johnson
80–81	Dimitry Schidlovsky
83	Yutaka Yoshii (*photograph*)
84	Izumi Shimada (*left, top right* and *bottom right*), Yutaka Yoshii (*center right*)
85	Izumi Shimada
86	Yutaka Yoshii (*top*), Jo Ann Griffin (*bottom*)
87	César Samillán (*drawing*), Jo Ann Griffin (*top right* and *bottom*), Yutaka Yoshii (*middle*)
88	Yutaka Yoshii (*top*), Izumi Shimada (*bottom*)
89	Yutaka Yoshii
91	George Retseck
92	National Aeronautics and Space Administration (*top*), Tomo Narashima (*bottom*)
93	NASA (*left*), Jared Schneidman/JSD (*right*)
94	Tomo Narashima
95	A.I.F. Stewart, University of Colorado
96–97	Tomo Narashima
98–99	Marguerite Holloway
100	Johnny Johnson (*top*), Marguerite Holloway (*bottom*)
101	South Florida Water Management District
102	Patricia J. Wynne
103	James Arnovsky/Zuma
104	Patricia J. Wynne
106	Andre F. Clewell, Hall Branch Restoration Project
108	Ken Sherman
120–123	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

Science Marches On

I am requesting that you attempt the following experiment. You will need the cooperation of a whole town of people. Have everybody remove their wedding rings. I would be interested in what happens to human sexuality and reproduction when nobody wears a ring.

You might be inclined to respond that nothing would happen. But gold is an unusual element in terms of electron conductivity, and the ring, because of the bipedal form of humans, is in close proximity to the sex organs.

GEORGE SILIS
Cleveland, Ohio

My theory, and I will offer proof, is that the late, great Howard Hughes was a time traveler. Every business venture that Mr. Hughes undertook involved high technology and advanced applications. Where did he get his insight? The answer: from the future! Why was Mr. Hughes such a recluse? The answer: he was back from the future and did not want to be revealed.

CHRISTOPHER J. RONAN
U.S. Air Force

For some time, I have been chagrined at the bumbling of physicists. The entire field needs a new beginning. I can offer the following help: Space has no dimensions and no fabric. Space cannot be warped. Space is a state of nothingness. I repeat, space cannot be warped.

My girlfriend says you guys are going to pass this around, saying, “Hey, Charlie, check out *this* quack.”

JOHN NICHOLS
Carson, Calif.

The names of most scientific disciplines end in the suffix “-ology.” I propose a new name for those amazing bits of discovery to which we react by saying, “Wow!” We could call it bygology.

DONALD M. SWAN
Old Saybrook, Conn.

NO! You haven’t heard from me, and you shall not until you have signed a contract. The cost to *Scientific American* is now $1.00 per character space for any article. For more than 25 years, I have sent articles to you, all of which were rejected. The Grand Unified Field is now wide open. The ancient mathematics has been recovered, and I cannot tell you how amazing it is. It depends entirely on prime numbers, of which you have not the slightest comprehension. There is no need for trigonometry or calculus. They are now dead subjects (as dead as *Scientific American* is going to be).

Be sure to include a retainer check in your next letter.

BEN IVERSON
Tigard, Ore.

Something We Said?

Cancel my subscription at once.

I thought you might have changed. But you never do, do you? Is there no end to your Stalinist suckups? Of course not. Whether it is your sniff-and-sneer approach to reporting economics or your toadying to the eco-statist line, the garbage never stops. In the name of science, you commit these abominations every issue without fail. You are the damned of the earth. Yours is the guilt beyond forgiveness.

JOHN L. QUEL
Bellevue, Wash.

Thanks to “Red-Banner Burger,” by Gary Stix [“Science and Business,” SCIENTIFIC AMERICAN, June 1993], I am up-to-date on your attempts to restrict me to a “choice of a hamburger well done or just plain burned.” You could not have made it any more clear that your objective is to kill your readers.

I sincerely hope you and your associates at *Scientific American* will be among the first and most enthusiastic users of the latest poisonous meat product—irradiated chickens. I will be delighted to dance at your funerals.

ROBERT G. HUENEMANN
La Honda, Calif.

Nobody’s Perfekt

The excellent article “Current Events,” by Philip Yam [SCIENTIFIC AMERICAN, December 1993], mentions “a two-horsepower motor—strong enough to power the cooling fan in a desktop computer.” Do desktop computers now really need fans the size of those for central air conditioning?

ROBERT NEUBOLD
York Haven, Pa.

The editors reply:
The fan also cools our more overheated comparisons.

Have you actually used “inputted” as the past tense of a verb? Yes, in the caption of the figure on page 150 of the January issue. I am upset. I am appalled. I am horrified.

I am out putted.

DAVID C. CALHOUN
Seattle, Wash.

The illustration on page 98 of the February issue says the strong force “couples quarks to form proteins and neutrons.” Do you favor meat or dairy products as sources of quark proteins?

BRUCE C. ALLEN
Cleveland, Ohio

The editors reply:
We prefer crow or, better still, irradiated chickens.

As the coauthor of *The Private Lives of Albert Einstein,* I would like to distance myself from Peter Highfield (no relation), who was portrayed as a tabloid hack in “Keyhole View of a Genius,” by Fred Guterl [“Profile,” SCIENTIFIC AMERICAN, January]. If Paul Carter and I had wanted to put Einstein in the worst possible light, we would not have sent the draft manuscript to three Einstein scholars and Einstein’s granddaughter. We do not in any way suggest there were “shades of Woody Allen” in Einstein’s relationship with his stepdaughter! I remain confident that our book contains the most authentic depiction of Einstein the man, thanks to our use of more than 1,100 references and the kind help we received from the Einstein Papers Project at Boston University and Evelyn Einstein.

ROGER HIGHFIELD
The Daily Telegraph
London, England

The editors reply:
Dr. Highfield is correct: his name is Roger.

50 AND 100 YEARS AGO

APRIL 1944

"In seven years the Armour Foundation has grown from a name and a good idea into one of the most important institutions of its kind in the United States. The 'good idea' was to provide an industrial research service for the particular benefit of small business. Today the foundation is 100 percent devoted to war products. But its director is free to make some predictions about the future. One of these concerns 'radio cookery,' an outgrowth of diathermy and 'artificial fever' treatment. He prophesies that we shall have electronic cooking as a generally accepted commercial practice, but doubts its use in the home because of the hazard of high voltage. One commercial company has already perfected a thermal radio hamburger and hot-dog vending machine. The customer drops a coin into a slot, and after half a minute a radio-cooked morsel pops out."

"Recently, Dr. James Hillier of RCA Laboratories announced the preliminary development of a fundamental tool to which he gave the name electron microanalyzer. Its function, he said, was the elemental analysis of extremely small areas within electron microscope specimens. With this instrument, the user can study a specimen already so microscopic that it must be magnified thousands of times in order that its details may be seen at all. It is possible to select one local area or perhaps a particle no larger than $1/100{,}000$ inch in diameter and as small in weight as $1/1{,}000{,}000{,}000{,}000{,}000$ gram, and determine exactly which chemical elements that one sub-microscopic area or particle contains."

"Although the war has been responsible for many new inventions, it has added little to the world's store of fundamental knowledge, Dr. Frank B. Jewett, vice-president of the American Telephone and Telegraph Company, recently told members of the New York University Institute on Post-War Reconstruction. Progress in certain fields of scientific knowledge, he said, has been offset by a virtual cessation of research work in others that are not considered essential to the war effort."

SCIENTIFIC AMERICAN

APRIL 1894

"The use of electricity for household purposes has hardly got beyond the experimental stage, save in the department of lighting; but enough has been done to show what a transformation may be worked when it is possible to have houses heated by it. Then the mere turning of a switch will suffice. With regard to cooking, there are numerous appliances already devised, and only waiting for the cheapening of the current to be widely taken advantage of. A New York lady is said to have so contrived matters that she can, before getting out of bed, start a fire in the kitchen by turning on the current; and when she comes downstairs finds the kettle boiling and the place comfortably warmed.—*Chambers's Journal.*"

"Mr. Lester Ward, in a lecture recently delivered before the Anthropological Society of Washington, showed that the work of Ramon y Cajal and others indicated that protoplasm is not merely the physical basis of life, but is the physical basis of mind also. In his words, 'the prevailing fashion among scientific men of emphasizing the mystery of mind is unnecessary and illogical, since mind is no more a mystery than matter.' "

"Dr. J. M. Macfarlane has recently discovered that leaf blades of the Dionæa—the Venus flytrap—will not respond to a single touch. There must be a second stimulus before an attempt at closing is made. But even here the stimuli must have an interlude of nearly a minute. If the two stimuli follow closely, no response follows. Here may be the advantage of the interlude: it offers a way of discovering whether that which alights on the leaf's surface is eatable or not. A piece of gravel might rebound—might make two stimuli close after one another. An insect would wait a short time to collect its senses, and formulate some plan of escape before struggling to get free. The discovery of Dr. Macfarlane is probably the most wonderful of all wonderful things that have been discovered in the behavior of plants.—*The Independent.*"

"The Midwinter Fair, an international exhibition, opened on January 27, 1894, and occupies about 160 acres of Golden Gate Park, San Francisco. The Manufactures and Liberal Arts Building, shown in the drawing, is probably the first building to attract the eye. The great blue dome and golden lantern glistens against the intense blue of the semitropical sky like an immense jewel, while a peculiar suggestion of age is given by the grayish-green tiles of the roof. This building is the largest structure at the Fair. In this great building thirty-eight nations have exhibits. The United States is well represented."

The Manufactures and Liberal Arts Building at the Midwinter Fair

Cool Man, Hot Job

William Perry takes on the challenge of military reform

JEFFREY MARKOWITZ *Sygma*

STEALTH MANDARIN: Secretary of Defense William J. Perry will have to fight budget pressures and bureaucracies to sustain research for the military of next century.

First Les Aspin buckled, then Bobby Inman blew up on takeoff. When the dust settled, there was Bill Perry. His initial demurrer notwithstanding, the view inside the Beltway is that this quiet-spoken technocrat, scholar and businessman seems in some important ways to have emerged as secretary of defense at just the right time.

Although not given to bull-in-the-woods bellowing, William J. Perry has for years been a tenacious advocate of the electronic battlefield. His original rationale had more to do with preventing Soviet forces from overrunning their numerically inferior NATO counterparts than with fighting brushfire wars. But high-technology weaponry used for General H. Norman Schwarzkopf's Desert Storm permitted an intoxicating (perhaps dangerously so) victory.

That lesson of history will stand Perry in good stead as he goes on point in the corridors of power. The disappearance of the Soviet Union and its satellites as credible military threats and the consequent demand for a "peace dividend" have led to steadily decreasing defense expenditures in recent years. Budgets are down 35 percent in real terms from their peak in 1985, and the administration's proposed defense budget for 1995, at $263.7 billion, continues that trend. But the research and development component, at $39.5 billion, represents a 4 percent increase. Basic research, which amounts to $1.23 billion within that total, is also slated for a small increase.

Despite his studied blandness, Perry will fight hard for military research. This commitment has earned him the respect of Pentagon brass, as well as of defense contractors who are already staggering from the impact of cutbacks. They are well aware that as undersecretary of defense for research, engineering and acquisition under President Jimmy Carter in the 1970s, Perry championed stealth technologies and precision-guided munitions. At the time, many in the military favored matching brute force with brute force.

Not all Perry's technological fixes have been triumphs: his critics point out that he gave the thumbs-up to the expensive and controversial MX missile, as well as to the canceled Aquila Remotely Piloted Vehicle. Moreover, notes Kosta Tsipis, a defense analyst at the Massachusetts Institute of Technology, the mild-mannered Perry lacks the political clout he might need to defend budgets in bruising cabinet battles.

Yet Perry enjoys the advantage of knowing the defense business from the inside. He prospered as the founder of a defense electronics firm and as a technology investment adviser. "He was always a believer in technological superiority," Tsipis says. "I think he'll try to maintain military research and development generally, but because of pressures on the budget he'll put more into basic research."

Other defense analysts agree that Perry will honor President Bill Clinton's commitment to a strong military by nurturing research that might yield the game-changing technologies of next century. Research is much cheaper than late-stage weapons development. Advances in computer simulation mean it is now possible to learn a lot about the performance of a weapon without going to the expense of building it, notes Albert R. C. Westwood, a researcher at Sandia National Laboratories. Perry well understands the significance of those advances, Westwood says. Furthermore, Perry can be expected to support international arms-control treaties as an economic route to military security.

Perry has also taken it on himself to improve efficiency all round. The Department of Defense, bowing to the inevitable, is now reviewing the roles of the 68 laboratories that it runs, and the Department of Energy has impaneled a blue-ribbon commission to look into the future of the national labs that it maintains. Undersecretary of Defense for Acquisition and Technology John M. Deutch acknowledged at a recent conference that "significant problems" mean the defense labs "inevitably will have to be downsized."

Perry has other schemes in the works to get the most bang for the research buck. "Perry is a very innovative thinker," says Stanley V. Jaskolski, chief technology officer at Eaton Corporation. Jaskolski cites the example of "the Perry initiative," a plan Perry started as deputy secretary of defense last year. In essence, the initiative swaps U.S. weapons know-how for access to Japanese technology. Japanese manufacturers are licensed to produce U.S.-designed weapons that embody advanced technology in exchange for rights to Japanese technologies that U.S. companies would like access to. Something similar "could be done in other countries," Jaskolski notes. "In principle, this could be universally applied."

The Perry initiative is a logical extension of his campaign to promote "dual use" technologies, which can be profitable in both military and civilian settings. That program brings him face-to-face with the most redoubtable dragon in the Pentagon's cave: the military procurement system. The heart of that system is the "milspec."

Numerous commissions and reports have stated the case for simplifying or abolishing milspecs, the elaborately detailed technical requirements that the Pentagon habitually lays down for purchases of everything from jet fighters to ashtrays. Milspecs prevent the Pentagon from buying at civilian prices in the civilian marketplace: they were responsible for the celebrated $640 toilet seat and the $435 hammer. Milspecs also impede, Perry has said, the diffusion of technologies resulting from military research into the private sector.

In 1992 Perry chaired a task force of the Carnegie Commission on Science, Technology and Government, whose report makes his views clear. The report observes that in 1991, 40 percent of the military acquisition budget was spent on management and control personnel. In civilian commerce the equivalent figure is between 5 and 15 percent. The solution that the task force advocated is as simple as it is radical: "The reform of the defense acquisition system must have as its principal thrust the integration of the country's defense industry and commercial industry to create a single industrial base. The critical ingredient of adaptation to commercial practice is conversion from a regulation-based system to a market-based system." Now Perry should be in a better position to achieve that conversion.

Simplifying milspecs "is extremely wise" because such burdensome requirements often have the perverse effect, according to Westwood, of delaying technological improvements. Westwood was previously a vice president for research and technology at Martin Marietta Aerospace, where he frequently encountered the problems milspecs create. Milspecs "don't allow you to say if there's a better material—they cut you off from real improvements," Westwood complains.

Moreover, Perry's credo of dual-use technology has become a buzzword in defense circles. A trickle of defense contractors starting to work on civilian projects has become a flood in the past couple of years, encouraged by defense department programs that support such shifts. "What we need is a healthy manufacturing infrastructure that is convertible," argues John Cassidy, vice president for research at United Technologies.

Few in Congress would disagree. Yet fights over which weapons systems to cancel and which industries to support will be bloody. "Perry is going to be in a very dynamic position to recraft the military of the future," reflects one high-ranking congressional aide. "It's going to be a hot job." —*Tim Beardsley*

Quantum Computing Creeps Closer to Reality

More than a decade ago a small group of physicists, among them Richard P. Feynman, began wondering whether it would be possible to harness quantum effects for computation. Until recently, such investigations have been highly abstract and mathematical. Now Seth Lloyd, a researcher at Los Alamos National Laboratory, has proposed in *Science* how a so-called quantum computer might actually be built.

Lloyd points out that in one sense "everything, including conventional computers, and you and me, is quantum mechanical," since all matter obeys the laws of physics. One feature distinguishing quantum computers from conventional ones, Lloyd explains, is the way they store information. Conventional computers use electrical charge or its absence to represent 0's or 1's used in the binary language of data storage.

In a quantum machine, information would be represented by the energy levels of individual particles or clusters of particles, which according to quantum mechanics occupy discrete states; the ground, or "down," state could signify a 0 and the excited, "up" state a 1. Lloyd says such computers could be made out of materials with identical, repeating units that behave quantum mechanically, including long organic molecules, or polymers; arrays of quantum dots, which are clusters of atoms with precisely controllable electronic properties; and crystals. "Something as simple as a salt crystal might do," he states.

Input is supplied by pulses of light or radio waves, which would nudge the atoms, molecules or quantum dots into energy levels representing, say, a particular number. More pulses of light would cause the system to carry out a computation and disgorge an answer. Because quantum systems are notoriously susceptible to disruption from external effects, an error-correction program would monitor the progress of a computation and put it back on track when it goes awry.

Such a computer would be much smaller and faster than any current model, Lloyd contends. It could also perform certain tasks beyond the range of any classical device by exploiting a bizarre quantum effect known as superposition. Under certain precisely controlled conditions, a particle can briefly inhabit a "superposed" energy state that is, in a sense, both down and up. It has a 50–50 probability of "collapsing" into one state or the other.

Computers that can store information in a superposed form, Lloyd suggests, could generate truly random numbers, a task that has proved fiendishly difficult for classical computers. They could thus solve certain problems with a probabilistic element—such as those involving quantum mechanics—more accurately than can conventional machines.

Rolf Landauer of the IBM Thomas J. Watson Research Center, an authority on the limits of computing, has "a number of reservations" regarding Lloyd's scheme. Landauer argues, for example, that Lloyd's error-correction method will destroy the very superposition that he seeks (for reasons related to the fact that mere observation of a quantum system alters it). Yet Lloyd's work is still "a step forward," Landauer says. "He's given us something to evaluate in more detail." —*John Horgan*

Super *Sonic*

A gene named for a video game guides development

The shape of a hand is as comfortingly familiar as, well, the back of one's hand. But to developmental biologists, it is also an enigma. What biochemical sculptor molds the delicate embryonic tissues into limbs and functioning organs during the first weeks of life? Researchers think they have finally found a family of genes that nudge embryonic cells toward their proper destiny. One of these genes is a real overachiever: in vertebrate organisms, it organizes the central nervous system, defines the orientation of limbs and specifies where fingers and toes should grow. Its discoverers have whimsically—and appropriately—dubbed this gene *Sonic hedgehog,* after the hyperactive hero of a popular video game.

Cliff Tabin of Harvard Medical School, Andrew P. McMahon of Harvard University and Philip W. Ingham of the Imperial Cancer Research Fund in Oxford, England, lead the three laboratories that recently brought *Sonic* into the spotlight through a set of papers in *Cell.* Their demonstration that *Sonic* induces dramatic changes in embryos, Tabin explains, "opens the door. It's a great start for looking at signaling events early in embryogenesis."

Before the advent of molecular biology, embryologists usually resorted to the Frankenstein-like measure of cutting small bits of tissue out of embryos and grafting them into new positions to see what the results might be. Crude though those experiments might seem today, they yielded important clues. Workers found that during critical periods in development, some blocks of cells organize extensive changes in their neighbors. For example, cells in the zone of polarizing activity (ZPA) along the posterior edge of a limb bud somehow dictate how the limb should be oriented and where digits should form. Removing the ZPA prevents the limb from forming; moving the ZPA can change the limb's orientation. Embryonic structures called the notochord and the neural floor plate were found to serve a similar patterning function in the development of the spine and central nervous system.

Embryologists theorized that cells in the ZPA and other patterning centers release a morphogen, or signaling molecule. Nearby cells presumably interpreted the gradient of morphogen as positional information and differentiated accordingly. For the past 20 years, much of developmental biology has been a largely frustrating quest for those morphogens. "In the whole of vertebrate embryology, there isn't yet a single unequivocally identified morphogen," notes Lewis Wolpert, a pioneer in the study of limb development.

Two years ago the cloning of a gene called *hedgehog* in fruit flies presented a new opportunity. *Hedgehog* takes its name from the appearance of the mutant flies that lack it: they become short-lived embryos whose bottom surfaces are covered with spiky hairs. Tabin, McMahon and Ingham decided independently to look for a similar gene in vertebrates but soon began collaborating.

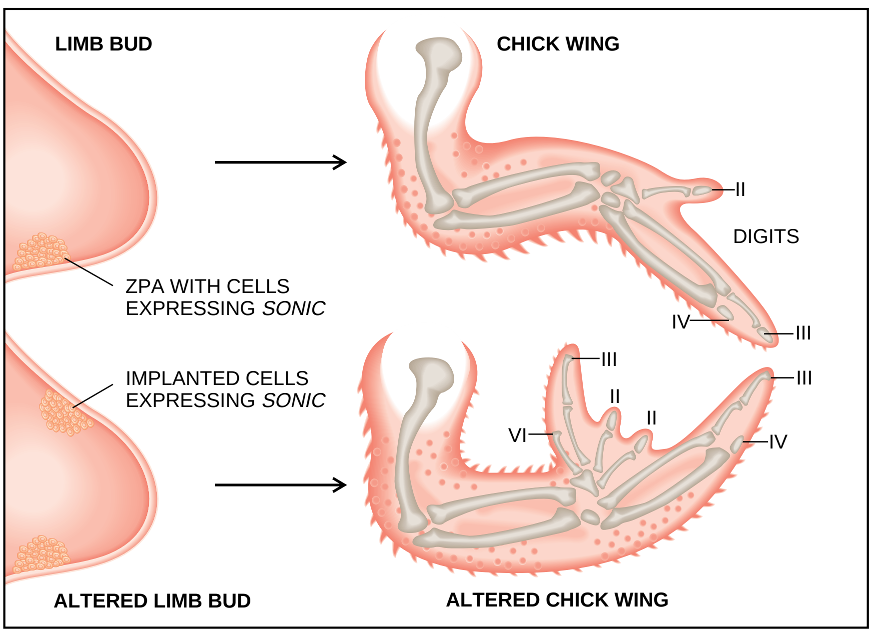
Using copies of the insect gene as probes, the investigators found four related *hedgehog* genes in vertebrates. These genes appear to make a family of structurally unique signaling proteins. The researchers named three of these genes after species of living hedgehogs: *Desert, Indian* and *Moonrat.*

But it is the fourth *hedgehog* gene, *Sonic,* that has so far proved most dazzling. In mice, chicks and zebrafish, cells in the ZPA, floor plate and notochord activate *Sonic* at precisely the times when they are shaping nearby structures. Moreover, when the researchers inserted genetically engineered cells that expressed *Sonic* into embryos, those cells served as new patterning centers. As such, they could change the orientation of limbs or create odd "mirror image" deformities. "Generally, you don't expect to find any single factor that mediates several different important signaling interactions," McMahon observes. "So it was a big surprise."

Tabin emphasizes that although Sonic protein is a primary developmental signal, it may not be a morphogen "in the classical sense." No one yet knows whether Sonic tells the limb bud how to grow by diffusing out of the ZPA and forming a concentration gradient. "It could just as easily be something that signals the adjacent set of cells and starts a cascade of other information signals from them," he explains.

Tabin, McMahon and Ingham are now looking for the receptor molecules to which Sonic binds: the locations will clarify which cells are the direct targets of the protein. The workers are also interested in determining which "upstream" signals tell cells to express *Sonic.* And then there are the other *hedgehog* genes to decipher: *Desert* seems to limit its activity to the reproductive system, whereas early tests suggest that *Indian* guides the differentiation of structures later in development.

"Excitement!" is Wolpert's reaction to the *hedgehog* revelations. More thoughtfully, he adds, "My joke is that we should enjoy it while we can. We went through a similar excitement for retinoic acid. The history of these things is that they turn out to be more complicated than one thinks." So the current success of *Sonic hedgehog* just kicks the developmental problem up to a new level of difficulty. Game over? Not by a long shot. —*John Rennie*



***MIRROR-IMAGE DEFORMITY** involving the duplication of digits occurs in embryonic limbs that contain both a normal zone of polarizing activity (ZPA) and an implant of cells expressing the patterning gene* **Sonic hedgehog.**

Bang! You're Alive

An unusual trio wins support for "nonlethal" weapons

In weapons laboratories, the Pentagon and even the justice department, a new buzzword is breeding: nonlethality. The basic idea is that soldiers and police, if only to maintain good public relations, often want not to kill their opponents but merely to disable them or their weapons. Bosnia, Somalia and Waco come to mind.

Of course, nonlethal weapons, ranging from radar jammers to rubber bullets, have long been in use. Federal researchers are now investigating a broader array of devices. These include laser rifles that temporarily blind the enemy or his optical-sensing gear; low-frequency "infrasound" generators powerful enough to trigger nausea or diarrhea; explosives that emit electronics-disrupting pulses of electromagnetic radiation; "stickums" and "slickums," chemicals that make roads or runways impassably gluey or slippery; and biological agents that can chew up crops or other strategic resources.

Steven Aftergood of the Federation of American Scientists, who has been tracking the nonlethal defense program, calls some of its components, and proponents, "weird." He notes that the program has been linked to nuclear weapons, which are hardly nonlethal, and to "mind control" devices; moreover, three of the most prominent advocates of nonlethality share an interest in psychic phenomena.

"Everything everyone says has to be treated with skepticism," Aftergood warns. But he adds, "This is a real program. Lots of money is being invested, mostly on a classified basis" by the departments of energy and defense and, to a lesser extent, the justice department. Aftergood has called for opening up the program to more public scrutiny.

At a conference on nonlethality held in Laurel, Md., last November, one of the few unclassified talks was given by Edward Teller of Lawrence Livermore National Laboratory, who is known as the father of the hydrogen bomb. Teller revealed that Livermore researchers were studying the feasibility of a miniature rocket "guided so accurately that it will fly down the muzzle of a gun, make a little pop, destroy the gun, not the gunner."

Teller urged that the nonlethality concept be stretched a bit to accommodate "small nuclear explosives" with yields equivalent to 100 tons of conventional high explosives, or roughly 1 percent that of the bomb that destroyed Hiroshima. With these explosives placed on "smart" rockets, the U.S. could force, say, North Korea to shut down its military facilities, Teller said. "We shall tell the enemy, the North Koreans, that if we find that people continue to go into these places, then at an unannounced moment they will be bombed." As long as the Koreans obey the U.S., in other words, the nuclear weapons remain nonlethal.

The chairman of the conference was John B. Alexander, who heads the nonlethal defense program at Los Alamos National Laboratory and has been called (albeit by a publicist at Los Alamos) "the father of nonlethal defense." For these efforts, Alexander was recently honored as an "Aerospace Laureate" by the respected journal *Aviation Week & Space Technology.* The citation did not mention that last year Alexander organized a meeting on "Treatment and Research of Experienced Anomalous Trauma," at which attendees discussed alien abductions, ritual abuse and near-death experiences.

STICKY FOAM engulfs a mannequin in a test at Sandia National Laboratories. Sticky-foam guns are among the nonlethal weapons being considered by the Department of Justice and other agencies.

Alexander is a former U.S. Army colonel and self-described "hard-core mercenary" with a doctorate in thanatology, the study of death. In 1980 he wrote an article for a defense journal on possible military applications of psychic powers. He emphasizes that he does not think paranormal techniques should be part of the nonlethal program, because such an association might be looked on askance by funding agencies.

Two other advocates of nonlethality, Janet E. Morris and her husband, Christopher C. Morris, are science-fiction writers and self-educated national security experts associated with the U.S. Global Strategy Council and the better-known Center for Strategic and International Studies, think tanks in Washington, D.C. Like Alexander, they have shown an interest in paranormal phenomena, including remote viewing (in which one supposedly "sees" distant scenes) and what they call "the effect of mind on probability."

The Morrises have been involved in promoting a "psycho-correction" technology, developed by a Russian scientist, that is intended to influence subjects by means of subliminal messages embedded in sound or in visual images. The Morrises say their intention is not to make the device part of the nonlethal arsenal but to make the U.S. aware of its dangers so that countermeasures can be developed.

Last year the Morrises organized meetings in which the technology was demonstrated for U.S. scientists and officials by its Russian inventor. A health official who observed the demonstration and requested anonymity described the demonstration as "hocus-pocus," adding that previous studies have shown subliminal-suggestion techniques to be ineffective. Nevertheless, the Federal Bureau of Investigation reportedly considered using the technique to convince cult leader David Koresh to surrender before he and his followers were immolated last year.

In the late 1980s Janet Morris was introduced to Alexander by a mutual acquaintance, Richard Groller, a former intelligence officer. Morris, Alexander and Groller then teamed up to write *The Warrior's Edge: Front-Line Strategies for Victory on the Corporate Battlefield.* Published in 1990, the book tells corporate climbers how psychic powers can help them rise to the top.

Since then, the Morrises and Alexander have had a falling-out, with each side accusing the other of hogging credit for the concept of nonlethality. "Alexander was ripping off our ideas and sending them up the chain of command," Janet Morris says. The Morrises also charge Alexander and other military officials with trying to keep the nonlethal program under wraps by classifying it.

Alexander acknowledges that many aspects of nonlethal research are indeed classified—including the budget of the program he oversees at Los Alamos (a laboratory spokesperson would say only that the figure is "in the millions"). But he contends that although he is in favor of relaxing restrictions, the decision is not his to make. He also denies coopting ideas from the Morrises. He asserts that he wrote a paper on nonlethality five years ago and that the basic concept had been discussed by defense analysts since at least 1972. "Claiming to have invented this concept is analogous to claiming to have developed civil rights," he declares. —*John Horgan*

An Epidemic Ignored

Endometriosis linked to dioxin and immunologic dysfunction

Endometriosis is a mysterious disease. Often misdiagnosed, its symptoms are varied, its cause obscure, its cure unknown. But some of the secrets of this illness, which afflicts 10 percent of women in their childbearing years—about 5.5 million people in the U.S. and Canada—are being unraveled. A report has linked the illness to dioxin exposure; other research suggests that immune dysfunction plays a role. "This is a pivotal time for the study of endometriosis," says Sherry E. Rier, an immunologist at the University of South Florida who led the team that conducted the work on dioxin.

The discoveries coincide with the recognition that the prevalence of endometriosis may be rising and becoming more common in young women. "The public health impact of this disease is enormous," says Jeff Boyd, a molecular geneticist at the National Institute of Environmental Health Sciences. "It affects millions and millions of people, but it does not garner the resources that more lethal diseases do, even though they affect a lot fewer people."

Endometriosis is a disease of renegade cells. Tissue from the uterine lining proliferates in other areas of the body, such as the bladder, intestine or, in rare cases, the lung. How these cells reach the distant organs remains unknown. One theory holds that rather than draining out of the body, menstrual blood flows backward into the fallopian tubes and moves on from there. Regardless of where they end up, endometrial cells continue to respond to the hormonal pulses of the menstrual cycle. When estrogen levels increase, the cells act as the uterine lining does, by building up; when progesterone rises, they slough off, causing internal bleeding. This shedding can cause great pain.

The discomfort caused by endometriosis has often been considered an unfortunate but untreatable aspect of women's biology: excruciating menstrual periods are just some women's lot. For that reason, physicians frequently did not recognize the disease until it was severe, often requiring the removal of uterus and ovaries. The Endometriosis Association, a Milwaukee-based organization, reports that 70 percent of women diagnosed with endometriosis were initially told by their doctors that there was no physical reason for their pain. Black women were told a slightly different story: 40 percent of those suffering intense pelvic pain that proved to be endometriosis were told they had a sexually transmitted disease.

Identifying endometriosis has become easier in the past decade because it is more widely recognized and because laparoscopy—the insertion of a tiny viewing tube into the abdomen—facilitates seeing the growths. But until recently, the enigma of its etiology seemed impenetrable. The disease was associated with many variables, including immune disorders such as lupus, with cancer, with the use of intrauterine devices (IUDs) and, most consistently, with infertility.

Between 30 and 40 percent of women who are treated for infertility have endometriosis, although it is not clear which condition, if either, causes the other. For many years, women's careers were deemed responsible. Researchers announced that delaying childbirth was the problem: the more periods a woman has in her life, the more susceptible she is. But "it is very easy to demolish that argument," comments Mary Lou Ballweg, president and executive director of the Endometriosis Association. Ballweg says many women experience their first symptoms in their teens: 41 percent of women diagnosed with endometriosis had symptoms before the age of 20. "I don't think we are going to want to tell 13-year-olds to go out and get pregnant as a form of prevention," she adds.

Today "endometriosis appears to be more of an immunologic than a reproductive disorder," Ballweg explains. "And when you look at the dioxin literature, everything starts falling into place." Dioxins are pollutants created in certain industrial processes; the most potent of the 75 kinds is 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, or TCDD. The link between TCDD and endometriosis was made last November, when Rier reported in *Fundamental and Applied Toxicology* that 79 percent of the females in a rhesus monkey colony exposed to dioxin developed endometriosis. The monkeys were exposed 15 years ago and subsequently monitored.

After three of the monkeys were found to have widespread endometriosis, the rest of the colony was examined. The prevalence and severity of the disease correlated with exposure: 43 percent of the animals who received five parts per trillion (ppt) of dioxin developed moderate to severe endometriosis, as did 71 percent of those exposed to 25 ppt. (An average person has about seven ppt of TCDD; the people most contaminated in the 1976 industrial accident in Seveso, Italy, had 56,000 ppt in their blood.)

Other findings support the dioxin connection. In 1992 German researchers announced that women with high blood levels of polychlorinated biphenyls (PCBs), compounds related to dioxins, have a greater than normal incidence of endometriosis. Scientists at the Department of Health and Welfare in Canada have also found that many female rhesus monkeys exposed to PCBs developed endometriosis. These data have not yet been published.

After Rier's findings, Boyd and his colleagues began evaluating blood levels of dioxins and 200 related compounds in women with endometriosis. Brenda Eskenazi of the University of California at Berkeley and Paolo Mocarelli of the University of Milan will study dioxin-exposed women in Seveso. "There have been a lot of studies of occupationally exposed males, and we really need some on women," notes Linda Birnbaum, a toxicologist at the Environmental Protection Agency, who studies rodent models of the disease.

The dioxin findings are intriguing because researchers are increasingly convinced that the pollutant acts like a hormone, often mimicking estrogen, and disturbs the immune system. Scientists have observed immunologic dysfunction in animals exposed to the contaminant. At least one researcher has reported similar disturbances in children born to dioxin-exposed mothers, although these data have not yet been peer-reviewed. The mechanisms of such interactions remain hidden for now, but it is evident that "these systems do not function alone," Rier says. "Changes in the endocrine system cause changes in the immune system."

In addition to finding a correlation between dioxin and endometriosis, Rier found immunologic changes in the monkeys that reflect those seen in people. Women with endometriosis often have very aggressive macrophages, a type of immune system cell, in the peritoneum. These macrophages secrete cytokines and growth factors that can irritate endometrial cells. Rier cautions that the monkey data are preliminary and that no one knows if changes in the immune system result from dioxin or from endometriosis. Nevertheless, the research creates excitement. "The whole issue of the immune system is fascinating; it is the right track," concurs David L. Olive, a reproductive endocrinologist at the Yale University School of Medicine.

Another interesting aspect of the immune system work may clarify the relation between endometriosis and infertility. Bruce A. Lessey, a reproductive endocrinologist at the University of North Carolina at Chapel Hill, has identified a receptor, called beta-3, for a cell-adhesion molecule absent in women with endometriosis. These molecules have many functions, including a role in immune response. Lessey also found beta-3 to be missing in some infertile women. Beyond potentially serving as a means of identifying and treating endometriosis and infertility, information about beta-3 "could be used to make a contraceptive," Lessey exclaims.

Taken together, the dioxin and immunologic research indicates that a fuller understanding of endometriosis may not be far-off. In this context, the suggested rise in incidence could be ominous. Environmental distribution of dioxin and its cousins has been spreading. Given "that dioxin is an endocrine disrupter and that there is a tight linkage between the immune system and endometriosis, it is not inconceivable that incidence is increasing and that the age of onset is decreasing," Birnbaum notes. —*Marguerite Holloway*

Silly Season

A brace of nutty events: read only if suffering cabin fever

The observation that things are not always what they seem may be particularly true in the realm of scientific inquiry. History, or at least its first draft in print, often needs to be corrected. Consider the case of *Fannia scalaris.* For 30 years, experts believed this species of fly had long gone unchanged. One specimen, preserved in amber from the Baltic region some 38 million years ago, gave *F. scalaris* its reputation. Entomologist Willi Hennig first examined the fossil in 1966. Astonishingly, he reported that the prehistoric housefly was identical in every way to common latrine flies of the 19th century. Such an evolutionary feat was widely celebrated and hardly questioned—until last fall, that is.

While poring over the collection of 12,500 fossil insects at the Natural History Museum in London, Andrew Ross, a scholar of ancient bugs, noticed something peculiar about the prized piece of amber. "A crack appeared around the fly," Ross says. "I realized something was very wrong." Indeed, the specimen showed the handiwork of an unknown, fly-by-night forger.

On further inspection, Ross discovered that the genuine Baltic amber had been split, carved, adorned with a passing pest, filled with an amber resin and glued back together to form a nearly flawless fake. "The entomologists are pleased because you shouldn't find such an advanced fly living that long ago," Ross says about reactions to his revelation. "I don't think anyone had really looked at the specimen after Hennig," he adds. A simple case of mistaken identity.

But some matters are less transparent than ancient amber. Reports widely circulated in the popular press have suggested that anyone can increase his or her I.Q. by listening to Mozart. This supposed quick fix is false. The confusion began after Frances H. Rauscher of the University of California at Irvine wrote a letter to *Nature* last fall. In her correspondence, Rauscher discussed a correlation she had observed between enhancement of spatial reasoning abilities and the act of listening to music.

She tested 36 volunteers in the following manner: Each participant listened to 10 minutes of Mozart's Sonata for Two Pianos in D Major and then answered questions taken from the Stanford-Binet intelligence test that gauge spatial reasoning abilities. The exercise was repeated two more times, using different listening conditions—a spoken voice and silence—and different questions. Rauscher's data showed that I.Q. scores based on spatial abilities alone were on average eight to nine points higher for those tests the subjects took after listening to Mozart.

Rauscher explains that her work is based on the premise that listening to music and performing a spatial task prime the same neural firing patterns. But that's just a guess. "We've heard a lot of skepticism, including from ourselves," she says. "It's disturbing that there have been so many misinterpretations claiming things we never said."

Nevertheless, some inaccuracies will stay on the books. Take, for example, the big bang, a name given to the theory that suggests that our universe began in a great explosion. Astronomy writer Timothy Ferris argued in an essay in *Sky & Telescope* magazine last August that the term was misleading. He noted that whatever happened back then, before time and space, truly was not big and probably did not go bang. Moreover, about 40 years ago, Fred Hoyle, who tenaciously supports a steady state theory of the universe, in-

vented the term to mock the whole idea.

Following Ferris's lead, *Sky & Telescope* sponsored a contest challenging anyone to come up with a more accurate catchphrase to describe the event. At the January meeting of the American Astronomical Society, the editors announced that there were no winners. Although they had received 13,099 entries, not one could match the wit and fame of the big bang name.

The three judges—Ferris, Carl Sagan of Cornell University and ABC's Hugh Downs—tossed out *What Happens If I Press This Button?, Jurassic Quark* and *You're Never Going to Get It All Back in There Again.* They discarded acronyms such as *NICK (Nature's Initial Cosmic Kickstart) of Time, SAGAN (Scientific Apprehension of God's Awesome Nature)* and *Big TOE (Theory of Everything).*

Downs and Ferris each picked a few favorites, but their semifinalists did not match. Sagan did not like any of the proposals. "The idea of space-time and matter expanding together and not 'into' anything may be permanently beyond reach in the universe of short and lucid phrases," Sagan said last summer. So even if *F. scalaris* is a fraud and Mozart won't make you smart, in the end, it will still be the big bang in the beginning. —*Kristin Leutwyler*

Super Progress

Defects pin intrusive magnetic fields in superconductors

Researchers trying to make useful products from high-temperature superconductors are hindered by the materials' reluctance to carry a current without resistance in strong magnetic fields. These fields, produced externally or by the resistanceless flow itself, appear in such envisioned uses as motors and generators. Recently workers have achieved significant breakthroughs in taming the disruptive effects.

Magnetic fields hamper current flow by penetrating into a superconductor as discrete bundles of flux called vortices. When these vortices move about, they disperse energy and impede the flow. Keeping the current moving without dissipation means anchoring the flux lines. One approach creates traps, usually by bombarding samples with heavy ions. If the ions—atoms stripped of their electrons—are sufficiently massive, they leave columnar tracks that hold the flux lines in place.

Unfortunately, these ions travel only about 20 to 50 microns through the substance. They have difficulty passing through the silver cladding that envelops commercially made superconducting wire tapes. "The heavy ions are out of juice by the time they reach the superconductor, so they cannot make tracks," says Lia Krusin-Elbaum of the IBM Thomas J. Watson Research Center.

To sidestep that problem, Krusin-Elbaum and her collaborators from six institutions accelerated much less massive particles: protons. High energy is more easily imparted to protons, which penetrate farther into a sample than do ions. Although protons are too light to create pinning defects, the researchers postulated that they might induce fission in the material. The process would send out as by-products heavy ions that would create pinning sites.

The experiment bore out the predictions. Using equipment at Los Alamos National Laboratory, the workers accelerated protons to 800 million electron volts and directed them at a bismuth-based superconducting tape. At such energies, the protons could whisk through more than half a meter of material. Electron micrographs indicated that the protons caused some of the bismuth to fission into heavy ions of xenon and krypton. Speeding out of the material, the heavy elements left columnar defects in the superconductor. As a result, the superconductor could carry at 30 kelvins up to 1,000 times more current than before.

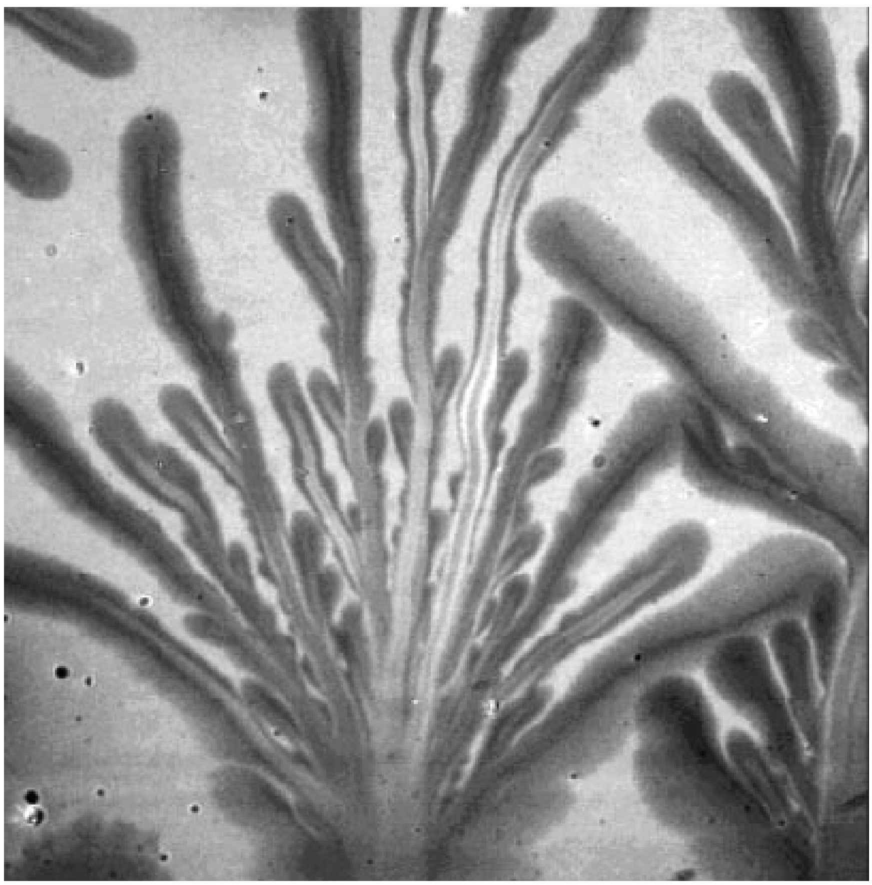
Fission produces tracks called splayed columnar defects. Such defects are a particularly effective means of pinning flux. They induce the magnetic vortices to become intertwined with one another. "Because everything is crossed and tangled up by the splay, the most strongly pinned flux lines will hold the more weakly held ones," explains David R. Nelson, one of the architects of the idea of splayed defects. The experiment with proton irradiation suggests that substances riddled with splayed columnar defects are three times as effective in holding flux lines than are parallel columns.

Pinning with protons may not require the use of a big accelerator. Conceivably, much less energetic protons—and hence much smaller devices—may be able to induce fission. "Thirty million electron volts should be more than enough to make the proton penetrate the charge barrier of the nucleus," Nelson observes.

Findings from Carlos A. Duran, Peter L. Gammel and David J. Bishop of AT&T Bell Laboratories suggest that flux pinners may have another tool to explore besides fission. By using polarized light and a special kind of magnetic coating laid on top of a superconductor, Bishop's group has produced novel optical images of a magnetic field intruding into a superconductor.

The images revealed a surprise. Rather than entering as a unified front, a weak magnetic field branches in, much the way a river running down a mountain produces rivulets. Each branch point appears to mark an area resistant to the intrusion. "The images show that we don't understand in detail flux penetration," Bishop says. "The patterns would suggest new strategies for pinning the flux." Making superconducting wires in layers, for instance, could help. Like ripstop fabric, the layers would arrest the branching and limit the penetration of the magnetic field into the superconductor.

Of course, widespread applications of the high-temperature materials will depend as much on durability and handling as they do on resisting magnetic fields. But in terms of holding the lines, investigators seem closer than ever to victory. *—Philip Yam*



DAVID J. BISHOP *AT&T Bell Laboratories*

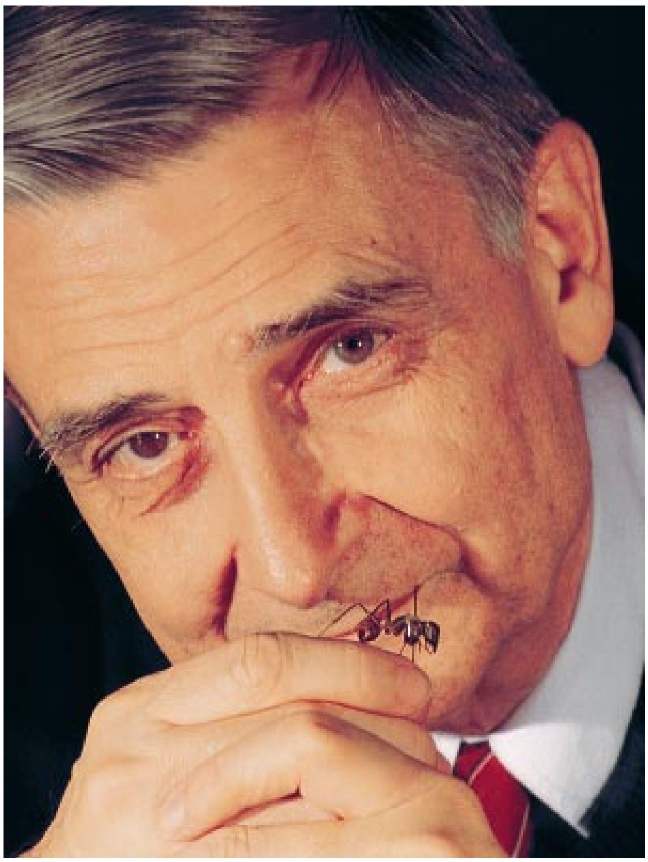
DENDRITES OF MAGNETIC FIELD 120 gauss strong penetrate from the side of a niobium superconducting thin film. The fingers average about 60 microns long.

Revisiting Old Battlefields

I've gone from politically very incorrect in the '70s to politically very correct in the '90s," muses Edward O. Wilson. "And I have to add that I wanted neither distinction, because I don't even think in terms of what's correct or incorrect." Wilson, or at least his public persona, has undergone a remarkable transformation over the past two decades. Today he is widely known and admired for his passionate defense of the biota, most recently in his 1992 best seller *The Diversity of Life.* In the late 1970s, however, Wilson was reviled by some scientists and political activists for his espousal of sociobiology, whose premise is that just as the social behavior of ants can be understood by examining their genetic underpinnings, so can that of humans.

When I first meet Wilson in his office at Harvard University's Museum of Comparative Zoology, I have a hard time imagining him at the center of any controversy. The 64-year-old Baird Professor of Science seems too gracious, even eager to please, and he keeps talking about ants—not surprisingly, since he is the world's leading authority on them. This is a man who once wrote that "ants gave me everything, and to them I will always return, like a shaman reconsecrating the tribal totem."

When I ask if science has anything more to learn about the tiny creatures, Wilson cries, "We're only just beginning!" He is now embarked on a survey of *Pheidole,* a genus thought to include more than 2,000 species of ants, most of which have never been described or even named. "I guess with that same urge that makes men in their middle age decide that at last they are going to row across the Atlantic in a rowboat or join a group to climb K2, I decided that I would take on *Pheidole.*"


JASON GOLTZ

LORD OF THE ANTS: they "gave me everything," says Wilson, shown here with a giant carpenter ant of Borneo.

Wilson's "grand goal" is to make *Pheidole* a benchmark of sorts for biologists seeking to monitor biodiversity. Drawing on Harvard's vast collection of ants, he has been generating descriptions and painstaking pencil drawings of each species of *Pheidole.* "It probably looks crushingly dull to you," Wilson apologizes as we flip through his drawings. He confesses that when he peers through his microscope at a previously unknown species, he has "the sensation of maybe looking upon—I don't want to get too poetic—of looking upon the face of creation."

I first detect a martial spirit glinting through this boyish charm when he shows me the leafcutter ant farm sprawling across a counter in his office. The scrawny little specimens scurrying across the surface of the spongelike nest are the workers; the soldiers lurk within. Wilson pulls a plug from the top of the nest and blows into the hole. An instant later several bulked-up behemoths boil to the surface, BB-size heads tossing, mandibles agape. "They can cut through shoe leather," he says, a bit too admiringly. "If you tried to dig into a leafcutter nest, they would gradually dispatch you, like a Chinese torture, by a thousand cuts." He chuckles.

Later, Wilson emphasizes that although he has not written much about sociobiology per se lately, its precepts inform all his work, on biodiversity as well as on ants. Moreover, he still harbors vast ambitions for human sociobiology. He thinks it has the potential to "subsume most of the social sciences and a great deal of philosophy" and bring about profound changes in politics and religion. He scolds Americans for their continued reluctance to confront the role played by genes in shaping human behavior. "This country is so seized by our civic religion, egalitarianism, that it just averts its gaze from anything that would seem to detract from that central ethic we have that everybody is equal, that perfect societies can be built with the goodwill of people." As he delivers this sermon, Wilson's long-boned face, usually so genial, is as stony as a Puritan preacher's.

Ever the biologist, Wilson has described his own career as a series of adaptations to environmental stresses. His father was a federal worker who kept moving from town to town in the Deep South. "Because of the difficulty in social adjustment that resulted from being a perpetual newcomer," Wilson has said, "I took to the woods and fields." At seven he lost most of the vision in his right eye after accidentally stabbing it with the fin of a fish he had yanked from a pond. With acute though myopic vision in his left eye, Wilson focused on animals he could scrutinize from short range, namely, ants.

Wilson pursued his studies at the universities of Alabama and Tennessee and, from 1951 on, at Harvard. He began doing fieldwork in such exotic locales as New Guinea, Fiji and Sri Lanka, discovering ant species that exhibited a fantastic array of social structures. Working in the laboratory, Wilson also helped to show that ants and other social insects exchange information by means of a host of chemical messengers, named pheromones.

Wilson's foray into sociobiology was spurred at least in part by a threat to his scientific tribe. In the late 1950s molecular biologists, exhilarated by their ability to decipher the genetic code, began questioning the value of taxonomy and other whole-animal approaches to biology. Wilson has alleged that James D. Watson, the co-discoverer of DNA, who was then at Harvard, "openly expressed contempt for evolutionary biology, which he saw as a dying vestige that had hung on too long at Harvard." The memory still rankles, especially since taxonomy's status relative to molecular biology may have fallen even further. Wilson deplores that situation. "I think a world biological survey would do more for humanity during the next 20 years than the genome mapping project," he declares.

Wilson responded to the challenge from molecular biologists by broadening his outlook, seeking the rules of behavior governing not only ants but all social animals. That effort culminated in *Sociobiology.* Published in 1975, it was a magisterial survey of social animals, from termites to baboons. Drawing on the vast knowledge he had accumulated in disciplines such as ethology and population genetics, Wilson showed how mating behavior, division of labor and other social phenomena were adaptive responses to evolutionary pressure.

Only in the last chapter did Wilson shift his sights to humans. He argued that warfare, xenophobia, the dominance of males and even our occasional spurts of altruism all spring at least in part from our primordial compulsion to propagate our genes. Wilson has admitted that his style was "deliberately provocative," but he insists that he was not seeking or expecting trouble. "I stumbled into a minefield."

The book was for the most part favorably reviewed. Yet a group of scientists—notably Stephen J. Gould and Richard C. Lewontin, also biologists at Harvard—attacked Wilson for promoting an updated version of social Darwinism and providing a scientific justification for racism, sexism and nationalistic aggression. The criticism peaked at a scientific conference in 1978, when a radical activist dumped a pitcher of water on Wilson's head while shouting, "You're all wet!"

While granting that support for his proposals "was very slim" in the 1970s, Wilson asserts that "a lot more evidence exists today" that human traits can have a genetic basis. To be sure, many scientists, particularly in the U.S., shun the term "sociobiology" because it is still "freighted with political baggage." Nevertheless, disciplines with such "circumlocutory" names as "biocultural studies," "Darwinian psychology" and "evolutionary biological studies of human behavior" are all actually "sprigs" growing from the trunk of sociobiology, according to Wilson.

Ironically, Wilson himself, at the very end of *Sociobiology,* revealed some trepidation about the fruit that the field might bear. "When we have progressed enough to explain ourselves in these mechanistic terms," he wrote, "and the social sciences come to full flower, the result might be hard to accept." Wilson acknowledges that he finished the book "in a slight depression" caused by his fear that a complete sociobiological theory would destroy our illusions and end, in a sense, our capacity for intellectual and spiritual growth.

He worked his way out of that impasse by determining that at least two enterprises represented "unending frontiers." One was the human mind, which has been and is still being shaped by the complex interaction between culture and genes. "I saw that here was an immense unmapped area of science and human history that we would take forever to explore," he says. "That made me feel much more cheerful." He wrote two books on the topic with Charles J. Lumsden of the University of Toronto: *Genes, Mind and Culture* in 1981 and *Promethean Fire* in 1983.

The other endeavor that Wilson realized could engage humanity forever was the study of biodiversity. "With millions of species, each one with an almost unimaginably complex history and genetic makeup, we would have a source of intellectual and aesthetic enjoyment for generations to come." Wilson thinks this quest may be propelled by "biophilia," a genetically based concern that humans have for other organisms.

He explored this theory in his 1984 book *Biophilia.* While compiling statistics on the abundance of species for the book, however, he fell into another depression. Species, he found, were vanishing at an alarming rate; the diversity he so cherished was in mortal danger. That realization catapulted him into his role as a champion of biodiversity.

Wilson's writings on biodiversity have been praised even by some of his former critics. Gould, in a review in *Nature,* lauded *The Diversity of Life* as "a thoroughly successful mixture of information and prophecy." Yet this embrace was not complete; Gould derided the biophilia theory, arguing that humans show as great a propensity for destruction of life as for preservation of it.

"That has been due more to ignorance in humanity's history than desire to wipe other forms of life off the earth," Wilson retorts. Gould, with whom Wilson is "quite friendly," is "allergic to any idea that human nature has a biological basis, and I must say I believe he is nearly alone in that perception now."

Wilson intends to take up the banner of sociobiology again in two upcoming books. (A self-confessed "workaholic," Wilson has already written or edited 18 books and more than 300 scientific and popular articles.) One is a full-scale autobiography he has nearly completed and hopes will be published by the end of this year. "I am revisiting all the old battlefields," he remarks.

Wilson's next book will address "natural philosophy," a hoary term he has revived to refer to "the still uncharted and relatively vaguely defined region between biology, the social sciences, moral reasoning and the environment." Perhaps the book's most radical theme will be that findings from evolutionary biology can guide us in resolving moral disputes over topics as diverse as the preservation of species or birth control.

Most philosophers and even scientists believe evolutionary biology "cannot be prescriptive," Wilson states. "That is true to a certain extent," he adds, "but my position is that where we can agree on moral precepts is governed very much by our evolutionary history."

Far from promoting fatalism, knowledge of our evolutionary roots should help liberate us from dangerous patterns of behavior, according to Wilson. A society based on sociobiological precepts would allow us to develop a more rational political system, one that encourages the "maximum personal growth" of humans while preserving the environment.

He points out, for example, that evolutionary biology has shown that sexual intercourse promotes parental bonding and so the stability of the entire family. These findings might persuade the Catholic Church, which believes that the primary purpose of sex is procreation, to drop its prohibition against birth control, thereby aiding efforts to curb population growth. Wilson seeks to "build bridges" rather than to initiate yet another controversy with such arguments. "I don't know exactly where I'm going to end up," he says, "but I hope it's not in the midst of another minefield." —*John Horgan*

SCIENTIFIC AMERICAN

Trade, Jobs and Wages

Blaming foreign competition for U.S. economic ills is ineffective. The real problems lie at home

by Paul R. Krugman and Robert Z. Lawrence

The real wage of the average American worker more than doubled between the end of World War II and 1973. Since then, however, those wages have risen only 6 percent. Furthermore, only highly educated workers have seen their compensation rise; the real earnings of blue-collar workers have fallen in most years since 1973.

Why have wages stagnated? A consensus among business and political leaders attributes the problem in large part to the failure of the U.S. to compete effectively in an increasingly integrated world economy. This conventional wisdom holds that foreign competition has eroded the U.S. manufacturing base, washing out the high-paying jobs that a strong manufacturing sector provides. More broadly, the argument goes, the nation's real income has lagged as a result of the inability of many U.S. firms to sell in world markets. And because imports increasingly come from Third

PAUL R. KRUGMAN and ROBERT Z. LAWRENCE teach economics at the Massachusetts Institute of Technology and at Harvard University, respectively. Krugman works primarily on international trade and finance; he is a leading proponent of the view that historical and political factors play at least as strong a role in trade as do underlying national economic characteristics. In 1991 he was awarded the John Bates Clark Medal by the American Economics Association. Lawrence's investigations focus on international trade, with particular attention to its effects on the labor market. He is also a nonresident senior fellow at the Brookings Institution.

World countries with their huge reserves of unskilled labor, the heaviest burden of this foreign competition has ostensibly fallen on less educated American workers.

Many people find such a story extremely persuasive. It links America's undeniable economic difficulties to the obvious fact of global competition. In effect, the U.S. is (in the words of President Bill Clinton) "like a big corporation in the world economy"—and, like many big corporations, it has stumbled in the face of new competitive challenges.

Persuasive though it may be, however, that story is untrue. A growing body of evidence contradicts the popular view that international competition is central to U.S. economic problems. In fact, international factors have played a surprisingly small role in the country's economic difficulties. The manufacturing sector has become a smaller part of the economy, but international trade is not the main cause of that shrinkage. The growth of real income has slowed almost entirely for domestic reasons. And—contrary to what even most economists have believed—recent analyses indicate that growing international trade does not bear significant responsibility even for the declining real wages of less educated U.S. workers.

The fraction of U.S. workers employed in manufacturing has been declining steadily since 1950. So has the share of U.S. output accounted for by value added in manufacturing. (Measurements of "value added" deduct from total sales the cost of raw materials and other inputs that a company buys from other firms.) In 1950 value added in the manufacturing sector accounted for 29.6 percent of gross domestic product (GDP) and 34.2 percent of employment; in 1970 the shares were 25.0 and 27.3 percent, respectively; by 1990 manufacturing had fallen to 18.4 percent of GDP and 17.4 percent of employment.

Before 1970 those who worried about this trend generally blamed it on automation—that is, on rapid growth of productivity in manufacturing. Since then, it has become more common to blame deindustrialization on rising imports; indeed, from 1970 to 1990, imports rose from 11.4 to 38.2 percent of the manufacturing contribution to GDP.

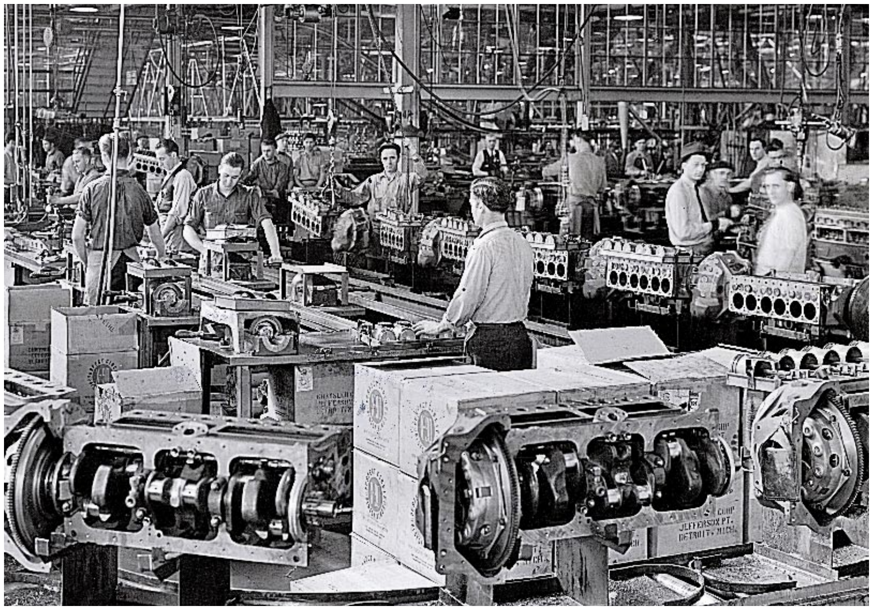
Yet the fact that imports grew while industry shrank does not in itself demonstrate that international competition was responsible. During the same 20 years, manufacturing exports also rose dramatically, from 12.6 to 31.0 percent of value added. Many manufacturing firms may have laid off workers in the face of competition from abroad, but others have added workers to produce for expanding export markets.

To assess the overall impact of growing international trade on the size of the manufacturing sector, we need to estimate the net effect of this simultaneous growth of exports and imports. A dollar of exports adds a dollar to the sales of domestic manufacturers; a dollar of imports, to a first approximation, displaces a dollar of domestic sales. The net impact of trade on domestic manufacturing sales can therefore be measured simply by the manufacturing trade balance—the difference between the total amount of manufactured goods that the U.S. exports and the amount that it imports. (In practice, a dollar of imports may displace slightly less than a dollar of domestic sales because the extra spending may come at the expense of services or other nonmanufacturing sales. The trade balance sets an upper bound on the net effect of trade on manufacturing.)

Undoubtedly, the emergence of persistent trade deficits in manufactured goods has contributed to the declining share of manufacturing in the U.S. economy. The question is how large that contribution has been. In 1970 manufactured exports exceeded imports by 0.2 percent of GDP. Since then, there have been persistent deficits, reaching a maximum of 3.1 percent of GDP in

ATTACKS on imported products, such as this Honda-bashing in Latrobe, Pa., are often motivated by the perception that foreign competition threatens jobs in the U.S. The authors argue that such hostility is misguided because international trade exerts only minor effects on the U.S. labor market.

INCREASING AUTOMATION has permitted U.S. factories to reduce employment while maintaining output, as visible in these photographs of mid-century and modern automobile plants. Evidence suggests that earlier fears that machines would replace people may be closer to the mark than current worries about foreign competition.



1986. By 1990, however, the manufacturing deficit had fallen again, to only 1.3 percent of GDP. The decline in the U.S. manufacturing trade position over those two decades was only 1.5 percent of GDP, less than a quarter of the 6.6 percentage point decline in the share of manufacturing in GDP.

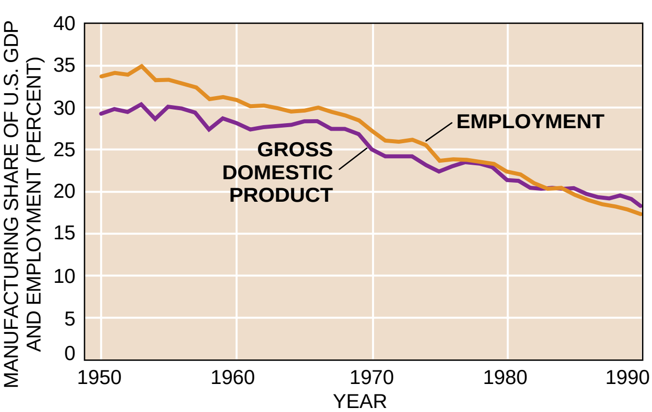
Moreover, the raw value of the trade deficit overstates its actual effect on the manufacturing sector. Trade figures measure sales, but the contribution of manufacturing to GDP is defined by value added in the sector—that is, by sales minus purchases from other sectors. When imports displace a dollar of domestic manufacturing sales, a substantial fraction of that dollar would have been spent on inputs from the service sector, which are not part of manufacturing's contribution to GDP.

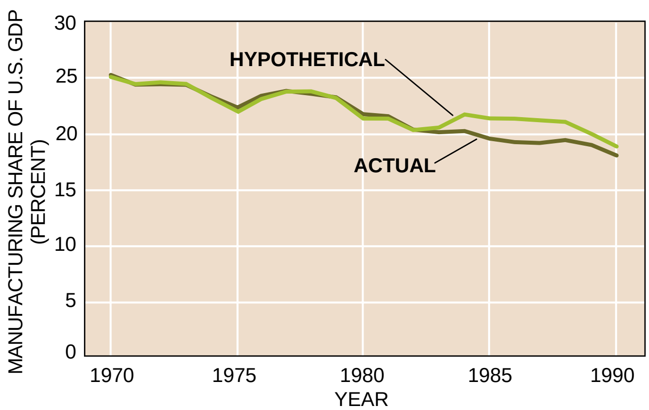
To estimate the true impact of the trade balance on manufacturing, one must correct for this "leakage" to the service sector. Our analysis of data from the U.S. Department of Commerce puts the figure at 40 percent. In other words, each dollar of trade deficit reduces the manufacturing sector's contribution to GDP by only 60 cents. This adjustment strengthens our conclusion: if trade in manufactured goods had been balanced from 1970 to 1990, the downward trend in the size of the manufacturing sector would not have been as steep as it actually was, but most of the deindustrialization would still have taken place. Between 1970 and 1990 manufacturing declined from 25.0 to 18.4 percent of GDP; with balanced trade, the decline would have been from 24.9 to 19.2, about 86 percent as large.

International trade explains only a small part of the decline in the relative importance of manufacturing to the economy. Why, then, has the share of manufacturing declined? The immediate reason is that the composition of domestic spending has shifted away from manufactured goods. In 1970 U.S. residents spent 46 percent of their outlays on goods (manufactured, grown or mined) and 54 percent on services and construction. By 1991 the shares were 40.7 and 59.3 percent, respectively, as people began buying comparatively more health care, travel, entertainment, legal services, fast food and so on. It is hardly surprising, given this shift, that manufacturing has become a less important part of the economy.

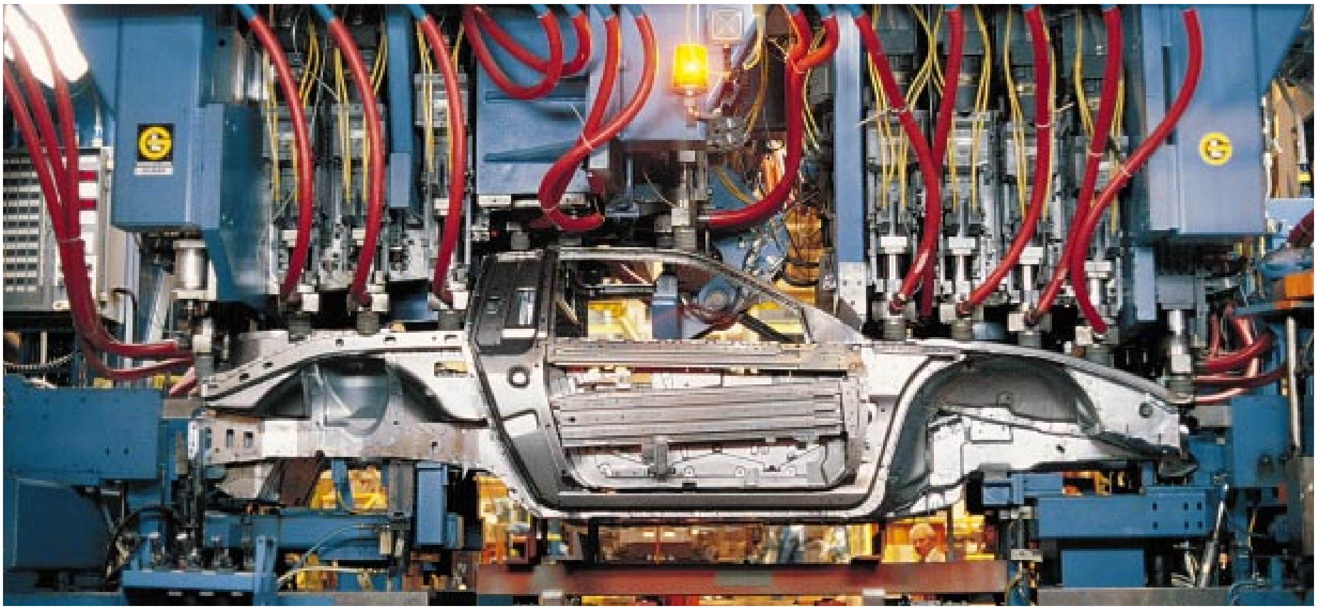
In particular, U.S. residents are spending a smaller fraction of their incomes on goods than they did 20 years ago for a simple reason: goods have become relatively cheaper. Between 1970 and 1990 the price of goods relative to services fell 22.9 percent. The physical ratio of goods to services purchased remained almost constant during that period. Goods have become cheaper primarily because productivity in manufacturing has grown much faster than in services. This growth has been passed on in lower consumer prices.

Ironically, the conventional wisdom here has things almost exactly backward. Policymakers often ascribe the declining share of industrial employment to a lack of manufacturing competitiveness brought on by inadequate productivity growth. In fact, the shrinkage is largely the result of high productivity growth, at least as compared with the service sector. The concern, widely





MANUFACTURING SHARE of gross domestic product has declined during the postwar era. The sector's share of domestic employment has decreased even more rapidly (*left*). Even if the U.S. were not importing more manufactured goods than it exports, however, correcting for trade balance shows that most of the decline would still have taken place (*right*).



voiced during the 1950s and 1960s, that industrial workers would lose their jobs because of automation is closer to the truth than the current preoccupation with a presumed loss of manufacturing jobs because of foreign competition.

Because competition from abroad has played a minor role in the contraction of U.S. manufacturing, loss of jobs in this sector because of foreign competition can bear only a tiny fraction of the blame for the stagnating earnings of U.S. workers. Our data illuminate just how small that fraction is. In 1990, for example, the trade deficit in manufacturing was $73 billion. This deficit reduced manufacturing value added by approximately $42 billion (the other $31 billion represents leakage—goods and services that manufacturers would have purchased from other sectors). Given an average of about $60,000 value added per manufacturing employee, this figure corresponded to approximately 700,000 jobs that would have been held by U.S. workers. In that year, the average manufacturing worker earned about $5,000 more than the average nonmanufacturing worker. Assuming that any loss of manufacturing jobs was made up by a gain of nonmanufacturing jobs—an assumption borne out by the absence of any long-term upward trend in the U.S. unemployment rate—the loss of “good jobs” in manufacturing as a result of international competition corresponded to a loss of $3.5 billion in wages. U.S. national income in 1990 was $5.5 trillion; consequently, the wage loss from deindustrialization in the face of foreign competition was less than 0.07 percent of national income.

Many observers have expressed concern not just about wages lost because of a shrinking manufacturing sector but also about a broader erosion of U.S. real income caused by inability to compete effectively in world markets. But they often fail to make the distinction between the adverse consequences of having slow productivity growth—which would be bad even for an economy that did not have any international trade—and additional adverse effects that might result from productivity growth that lags behind that of other countries.

To see why that distinction is important, consider a world in which productivity (output per worker-hour) increases by the same amount in every nation around the world—say, 3 percent a year. Under these conditions, all other things remaining equal, workers’ real earnings in all countries would tend to rise by 3 percent annually as well. Similarly, if productivity grew at 1 percent a year, so would earnings. (The relation between productivity growth and earnings growth holds regardless of the absolute level of productivity in each nation; only the rate of increase is significant.)

Concerns about international competitiveness, as opposed to low productivity growth, correspond to a situation in which productivity growth in the U.S. falls to 1 percent annually while elsewhere it continues to grow at 3 percent. If real earnings in the U.S. then grow at 1 percent a year, the U.S. does not have anything we could reasonably call a competitive problem, even though it would lag other nations. The rate of earnings growth is exactly the same as it would be if other countries were doing as badly as we are.

The fact that other countries are doing better may hurt U.S. pride, but it does not by itself affect domestic standards. It makes sense to talk of a competitive problem only to the extent that earnings growth falls by more than the decline in productivity growth.

Foreign competition can reduce domestic income by a well-understood mechanism called the terms of trade effect. In export markets, foreign competition can force a decline in the prices of U.S. products relative to those of other nations. That decline typically occurs through a devaluation of the dollar, thereby boosting the price of imports. The net result is a reduction in real earnings because the U.S. must sell its goods more cheaply and pay more for what it buys.

During the past 20 years, the U.S. has indeed experienced a deterioration in its terms of trade. The ratio of U.S. export prices to import prices fell more than 20 percent between 1970 and 1990; in other words, the U.S. had to export 20 percent more to pay for a given quantity of imports in 1990 than it did in 1970. Because the U.S. imported goods whose value was 11.3 percent of its GDP in 1990, these worsened terms of trade reduced national income by about 2 percent.

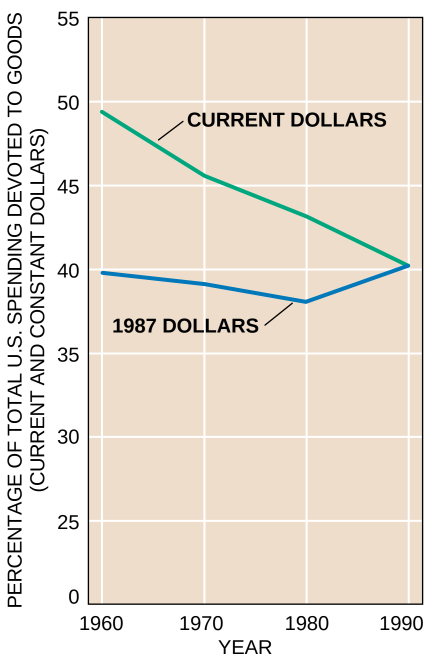
Real earnings grew by about 6 percent during the 1970s and 1980s. Our calculation suggests that avoiding the decline in the terms of trade would have increased that growth to only about 8 percent. Although the effect of foreign competition is measurable, it can by no means account for the stagnation of U.S. earnings.

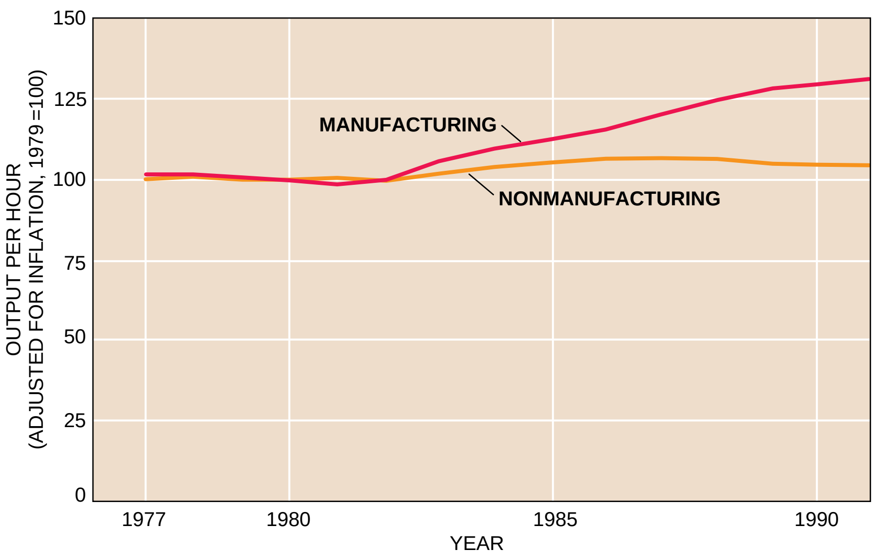
A more direct way of calculating the impact of the terms of trade on real income is to use a measure known as command GNP (gross national product). Real GNP, the conventional standard of economic performance, measures what the output of the economy would be if all prices remained constant. Command GNP is a similar measure in which the value of exports is deflated by the import price index. It measures the quantity of goods and services that the U.S. economy can afford to buy in the world market, as opposed to the volume of goods and services it produces. If the prices of imports rise faster than export prices (as will happen, for example, if the dollar falls precipitously), growth in command GNP will fall behind that of real GNP.

Between 1959 and 1973, when U.S. wages were rising steadily, command GNP per worker-hour did grow slightly faster than real GNP per hour—1.87 percent per year versus 1.85. Between 1973 and 1990, as real wages stagnated, command GNP grew more slowly than output, 0.65 percent versus 0.73. Both these differences, however, are small. The great bulk of the slowdown in command GNP was caused by the slower growth of real GNP per worker—by the purely domestic impact of the decline in productivity growth.

If foreign competition is neither the main villain in the decline of manufacturing nor the root cause of stagnating wages, has it not at least worsened the lot of unskilled labor? Economists have generally been quite sympathetic to the argument that increased integration of global markets has pushed down the real wages of less educated U.S. workers.

Their opinion stems from a familiar concept in the theory of international trade: factor price equalization. When a rich country, where skilled labor is abundant (and where the premium for skill is therefore small), trades with a poor country, where skilled workers are scarce and unskilled workers abundant, the wage rates tend to converge. The pay of skilled workers rises in the rich country and falls in the poor one; that of unskilled workers falls in the rich country and rises in the poor nation.





SHARE OF U.S. DOMESTIC SPENDING going to manufactured goods has declined substantially since 1960, although the volume of goods purchased has not (*left*). Instead goods have simply become cheaper relative to services. Productivity growth in the manufacturing sector has far outpaced such growth in service industries, especially during the past 10 years (*bottom*).

Given the rapid growth of exports from nations such as China and Indonesia, it seems reasonable to suppose that factor price equalization has been a major reason for the growing gap in earnings between skilled and unskilled workers in the U.S. Surprisingly, however, this does not seem to be the case. We have found that increased wage inequality, like the decline of manufacturing and the slowdown in real income growth, is overwhelmingly the consequence of domestic causes.

That conclusion is based on an examination of the evidence in terms of the underlying logic of factor price equalization, first explained in a classic 1941 paper by Wolfgang F. Stolper and Paul A. Samuelson. The principle of comparative advantage suggests that a rich country trading with a poor one will export skill-intensive goods (because it has a comparative abundance of skilled workers) and import labor-intensive products. As a result of this trade, production in the rich country will shift toward skill-intensive sectors and away from labor-intensive ones. That shift, however, raises the demand for skilled workers and reduces that for unskilled workers. If wages are free to rise and fall with changes in the demand for different kinds of labor (as they do for the most part in the U.S.), the real wages of skilled workers will rise, and those of unskilled workers will decline. In a poor country, the opposite will occur.

All other things being equal, the rising wage differential will lead firms in the rich country to cut back on the proportion of skilled workers that they employ and to increase that of unskilled ones. That decision, in turn, mitigates the increased demand for skilled workers. When the dust settles, the wage differential has risen just enough to offset the effects of the change in the industry mix on overall demand for labor. Total employment of both types of labor remains unchanged.

According to Stolper and Samuelson's analysis, a rising relative wage for skilled workers leads all industries to employ a lower ratio of skilled to unskilled workers. Indeed, this reduction is the only way the economy can shift production toward skill-intensive sectors while keeping the overall mix of workers constant.

This analysis carries two clear empirical implications. First, if growing international trade is the main force driving increased wage inequality, the ratio of skilled to unskilled employment should decline in most U.S. industries. Second, employment should increase more rapidly in skill-intensive industries than in those that employ more unskilled labor.

Recent U.S. economic history confounds these predictions. Between 1979 and 1989 the real compensation of white-collar workers rose, whereas that of blue-collar workers fell. Nevertheless, nearly all industries employed an increasing proportion of white-collar workers. Moreover, skill-intensive industries showed at best a slight tendency to grow faster than those in which blue-collar employment was high. (Although economists use many different methods to estimate the average skill level in a given industrial sector, the percentage of blue-collar workers is highly correlated with other measures and easy to estimate.)

Thus, the evidence suggests that factor price equalization was not the driving force behind the growing wage gap. The rise in demand for skilled workers was overwhelmingly caused by changes in demand within each industrial sector, not by a shift of the U.S.'s industrial mix in response to trade. No one can say with certainty what has reduced the relative demand for less skilled workers throughout the economy. Technological change, especially the increased use of computers, is a likely candidate; in any case, globalization cannot have played the dominant role.

It may seem difficult to reconcile the evidence that international competition bears little responsibility for falling wages among unskilled workers with the dramatic rise in manufactured exports from Third World countries. In truth, however, there is little need to do so. Although the surging exports of some developing countries have attracted a great deal of attention, the U.S. continues to buy the bulk of its imports from other advanced countries, whose workers have similar skills and wages. In 1990 the average wages of manufacturing workers among U.S. trading partners (weighted by total bilateral trade) were 88 percent of the U.S. level. Imports (other than oil) from low-wage countries—those where workers earn less than half the U.S. level—were a mere 2.8 percent of GDP.

Finally, increasing low-wage competition from trade with developing nations has been offset by the rise in wages and skill levels among traditional U.S. trading partners. Indeed, imports from low-wage countries were almost as large in 1960 as in 1990—2.2 percent of GDP—because three decades ago Japan and most of Europe fell into that category. In 1960 imports from Japan exerted competitive pressure on labor-intensive industries such as textiles. Today Japan is a high-wage country, and the burden of its competition falls mostly on skill-intensive sectors such as the semiconductor industry.

INFORMATION AND SERVICE industries are taking on the role that manufacturing once held in the U.S. economy.

We have examined the case for the havoc supposedly wrought by foreign competition and found it wanting. Imports are not responsible for the stagnation of U.S. incomes since 1973, nor for deindustrialization, nor for the plight of low-wage workers. That does not mean, however, we believe all is well.

Some of those who have raised the alarm about U.S. competitiveness seem to believe only two positions are possible: either the U.S. has a competitive problem, or else the nation's economy is performing acceptably. We agree that the U.S. economy is doing badly, but we find that international competition explains very little of that poor performance.

The sources of U.S. difficulties are overwhelmingly domestic, and the nation's plight would be much the same even if world markets had not become more integrated. The share of manufacturing in GDP is declining because people are buying relatively fewer goods; manufacturing employment is falling because companies are replacing workers with machines and making more efficient use of those they retain. Wages have stagnated because the rate of productivity growth in the economy as a whole has slowed, and less skilled workers in particular are suffering because a high-technology economy has less and less demand for their services. Our trade with the rest of the world plays at best a small role in each case.

The data underlying our conclusions are neither subtle nor difficult to interpret. The evidence that international trade has had little net impact on the size of the manufacturing sector, in particular, is blatant. The prevalence of contrary views among opinion leaders who believe themselves well informed says something disturbing about the quality of economic discussion in this country.

It is important to get these things right. Improving American economic performance is an arduous task. It will be an impossible one if we start from the misconceived notion that our problem is essentially one of international competitiveness.

FURTHER READING

PROTECTION AND REAL WAGES. W. F. Stolper and P. A. Samuelson in *Review of Economic Studies,* Vol. 9, pages 58–73; November 1941.

MYTHS AND REALITIES OF U.S. COMPETITIVENESS. P. R. Krugman in *Science,* Vol. 254, pages 811–815; November 8, 1991.

UNDERSTANDING RECENT CHANGES IN THE WAGE STRUCTURE. L. Katz in *NBER Reporter,* pages 10–15; Winter 1992/93.

TRADE AND AMERICAN WAGES IN THE 1980s: GIANT SUCKING SOUND OR SMALL HICCUP? R. Z. Lawrence and M. J. Slaughter in *Brookings Papers on Economic Activity: Microeconomics,* Vol. 2, 1993.

PEDDLING PROSPERITY: ECONOMIC SENSE AND NONSENSE IN THE AGE OF DIMINISHED EXPECTATIONS. P. R. Krugman. W. W. Norton Company, 1994.

Charge and Spin Density Waves

Electrons in some metals arrange into crystalline patterns that move in concert, respond peculiarly to applied voltages and show self-organization

by Stuart Brown and George Grüner

On a hot July afternoon the Mall in Washington, D.C., is overrun with sightseers. They move earnestly in zigzag patterns carrying their coolers, bouncing from museum to monument to cafeteria. Most of the streets bordering the lawns are flat, and as many tourists stroll in one direction as in the other. Suddenly a drumroll is heard: a marching band is assembling. On the roads, displacing the confused crowd, are gathering serried ranks of uniformed high school students. Soon the band is mustered in neat rows, hardly disturbed even by a child trying to hide between the trumpeters' legs from a pursuing parent. As the tourists watch, the band starts to play and then marches forward with a clash of cymbals.

The wanderers on the Mall imitate rather closely the behavior of electrons in common metals. On cooling to temperatures close to absolute zero, most metals remain in this state; that is, the electrons continue to wander. But in some metals the electrons organize themselves into regular patterns like the ranks of a marching band.

Such ordered ranks of electrons, otherwise known as charge-density waves, or CDWs for short, were envisaged by the theoretical physicist Rudolf E. Peierls in the early 1930s and discovered in the 1970s. A related phenomenon, spin-density waves, or SDWs, were predicted by Albert W. Overhauser in 1960, while at Ford Motor Company; the waves were also first seen in the 1970s. At one time, CDWs were suggested as being the agent of superconductivity. Today we know that superconductivity has a different origin, one in which the students dance in pairs rather than march; yet the many oddities of the marching bands themselves have kept researchers intrigued for decades. Charge-density waves may even find applications one day as tunable capacitors in electronic circuits and as extremely sensitive detectors of electromagnetic radiation.

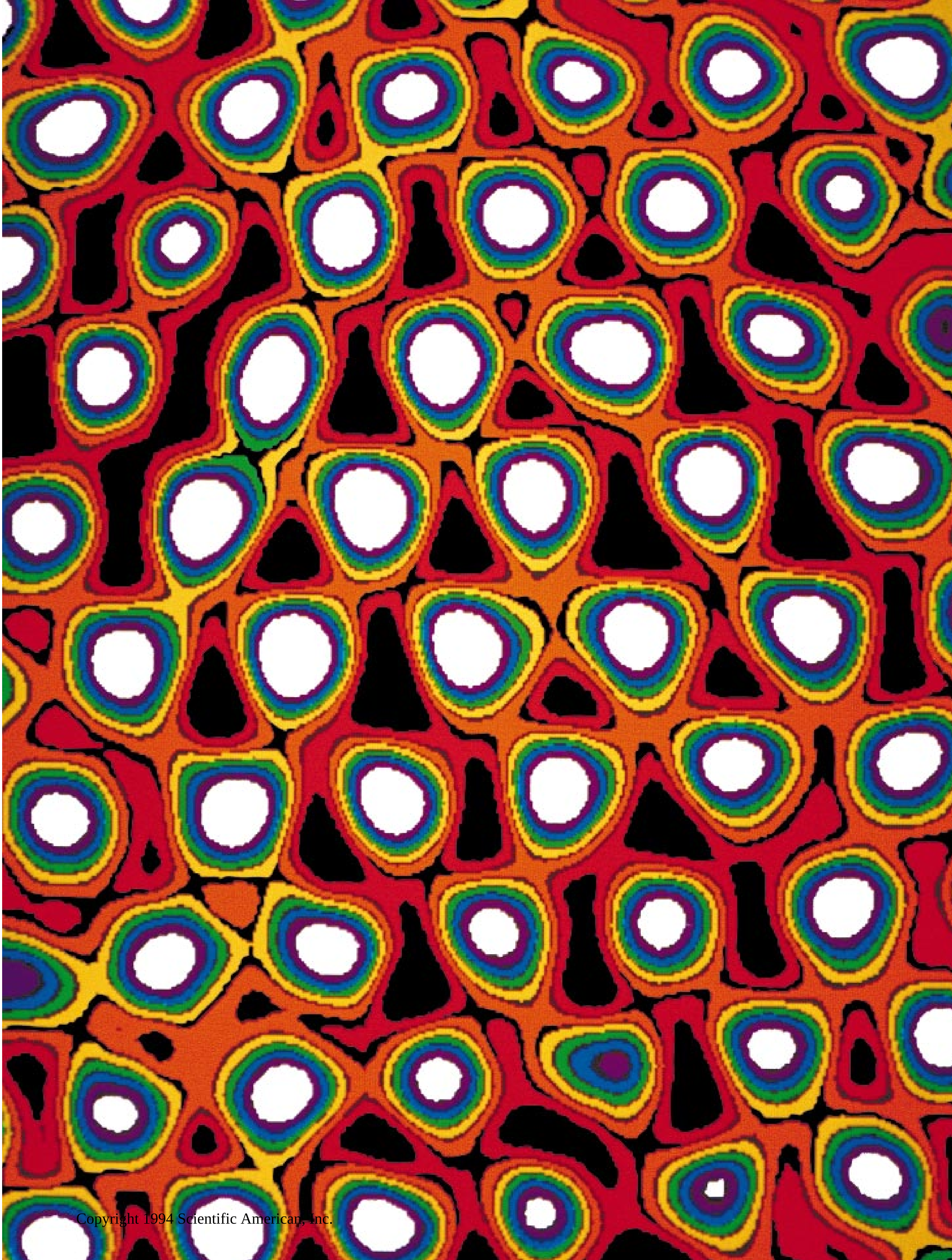
Hook up a battery to the ends of a solid in which a CDW exists and apply a voltage across it. If the voltage is small enough, nothing happens: the shoes of the marchers are stuck to the road with chewing gum. (The sticking is weak, so charge-density waves have a "dielectric constant" several million times that of semiconductors, which allows them to store enormous amounts of charge—hence their potential use as capacitors.) But if you increase the voltage beyond a certain threshold, the shoes suddenly break free, and the band begins to march—there is a large current. The current is not proportional to the voltage, as in ordinary metals obeying Ohm's law; instead it increases vastly with small increases in voltage. Further, a small part of the total current oscillates in time, even if only a constant DC voltage is applied.

And charge-density waves exhibit a "self-organized" response to externally applied forces. In fact, the concept of self-organized criticality grew out of initial work on CDWs. This field attempts to understand the motions of complex systems such as sandpiles or earthquake fault networks. Let's say we dribble sand onto a surface. It piles up into a conical shape; the cone is so steep that often adding a single grain will cause an avalanche. Likewise, tectonic plates perpetually poise themselves on the brink of an earthquake. In some circumstances, charge-density waves so configure themselves that any slight change in an externally applied electric field leads to a drastic change. CDWs are thus a tabletop system in which we can test theories of self-organization.

Why do density waves form? The underlying cause is the interaction between the electrons in a metal. Normally the electrostatic repulsion between the negatively charged electrons is canceled out by the presence of positive ions (atoms that have lost one or more electrons and are therefore positively charged), which form the body of the metal. Then the electrons hardly notice one another. In such a situation, if we picture the electrons as the strolling crowd described earlier, the probability of finding a person—or electron—at any one spot is the same as at any other. So the electrons' charge density is uniform in space. Now suppose the electrons do interact—say by affecting the lattice in which the positive ions are arranged. The lattice can in turn influence the position of a second electron, effectively giving rise to an interaction between the electrons.

STUART BROWN and GEORGE GRÜNER share an interest in the dynamics of driven systems such as charge-density waves and earthquakes. Grüner received his Ph.D. in Budapest in 1971 and worked as a postdoctoral associate at Imperial College, London. He has been professor of physics at the University of California, Los Angeles, since 1981. Brown earned his Ph.D. in 1988 from U.C.L.A. and did postdoctoral research at Los Alamos National Laboratory and the University of Florida. In 1991 he returned to U.C.L.A. as a member of the faculty.

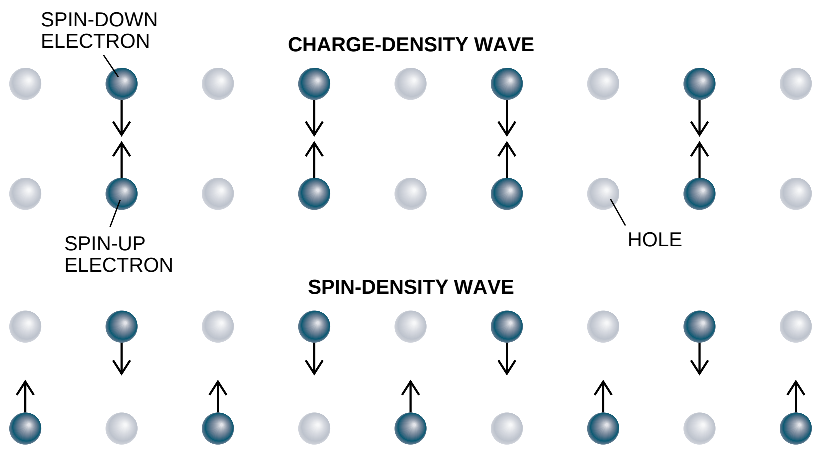
REGULAR PATTERN of charge-density waves in the material tantalum disulfide is revealed by a scanning tunneling microscope. Peaks of high charge density (*white*) are spaced 12 angstroms apart in a hexagonal array. Robert V. Coleman and C. Gray Slough of the University of Virginia provided the scan.



Pairing States

Charge-density waves (*top*) can be thought of as electron-hole pairs as well as electron-electron pairs. (A hole is a vacant quantum-mechanical state; it acts much like a particle.) Looking at the array, we can think of an electron as being paired either with a hole to its right or left or with the electron opposite. (Superconductivity comes from yet another kind of electron-electron pairing.) In spin-density waves (*bottom*), the opposing electron is shifted, so that the total charge density is constant but the spin density goes up and down along the array.

Which of the various pairing states occurs in a given material depends on the strength of the various interactions between the electrons. For instance, if direct electrostatic repulsion between electrons dominates, either a spin-density wave or a spin-parallel (or "triplet") superconducting state is favored. Electrons can also interact by distorting the lattice. The mediation of the lattice leads to attraction between electrons, and either a charge-density wave or a spin-antiparallel (or "singlet") superconducting state results.



The interactions often cause the electrons to become paired up; the pairs subsequently repel one another. Then each pair stays as far away as possible from all other pairs, and an ordered structure like that of the marching band is formed; the charge density becomes bumpy. If we take into account the wave nature of the electrons, a smooth variation of the charge density emerges. This smooth spatial variation of the charge is called a charge-density wave.

In addition to charge, electrons carry around with them something called spin. A spin is a magnetic moment associated with each electron; the moments can assume one of two states, labeled "up" and "down." If electrons with the same spin orientation repel one another, then each up spin wants to have a down spin as a neighbor. The result is a spin-density wave, or SDW. An SDW can be thought of as two CDWs, one for each spin state, superposed, with their peaks in alternate positions. Note that for a charge-density wave the charge varies in space, but not for a spin-density wave.

In general, how the electrons interact—and what kind of quantum-mechanical state is formed—depends on how the electrons' motion is confined. In three dimensions, electrons have the ability to avoid one another by simply moving out of the way. But if they are limited to traveling along a chain of atoms, the electrons cannot avoid one another and tend to interact more strongly. CDWs and SDWs occur mostly in such materials, in which the atoms are lined up in chains. (Many of these materials were first synthesized in the early 1970s.) In some circumstances, the electron pairs attract rather than repel one another, forming a superconducting state.

Chemists often design materials with a chainlike structure; however, they do not have much control over the nature of the electronic interactions. So whether the synthesized substance develops a CDW or an SDW or becomes superconducting cannot be predicted.

Electrons tend to pair at low temperatures. At absolute zero, each electron has its "mate," and the structure is fully ordered. As we warm up the material, some of the pairs become separated; they then induce other pairs to separate. With higher and higher temperatures, more and more pairs are divorced, until the last pair breaks up; above this critical temperature the material has only free electrons and is back to a metal. This process is known as a phase transition, as in the melting of an ice cube. If we reverse the process, cooling the material down from high temperatures, a CDW forms when we cross the phase-transition temperature. The electrons then get stuck. Because small electric fields can no longer dislodge them, and no current flows, the metal changes abruptly to an insulator. This sudden change in electric conductivity in fact signals the formation of a CDW.

Far more direct observations of CDWs have been made using scanning tunneling microscopes, which show the charge density even on atomic scale [*see illustration on preceding page*]. Further, a CDW is accompanied by distortions in the lattice. The distortion pattern, called a superlattice, can be seen by x-ray diffraction: the ions scatter x-rays onto photographic film, displaying a characteristic pattern that reveals their spacing. For example, if the superlattice wavelength is twice that of the lattice wavelength, the x-ray diffraction pattern will show additional spots halfway between the main spots coming from the lattice. (The intensity of the halfway spots relates to the size of the lattice deformation.) The first experiments of this type were performed by Robert Comes and his associates in Paris in the 1970s.

Since a spin-density wave leads neither to a charge fluctuation nor to a lattice distortion, detecting it is much harder. In principle, it could possibly be seen through the magnetic force microscope, an instrument that responds to variations of the spin, but the devices are not yet sensitive enough. The first demonstration of spin-density waves was made by scattering neutrons off chromium. (Neutrons, having spin and no charge, are useful for studying ordered spin structures.) In addition, indirect probes of the magnetic field, such as magnetic resonance—the same technique used in hospitals as a diagnostic tool—are now the only means of sensing the presence of SDWs.

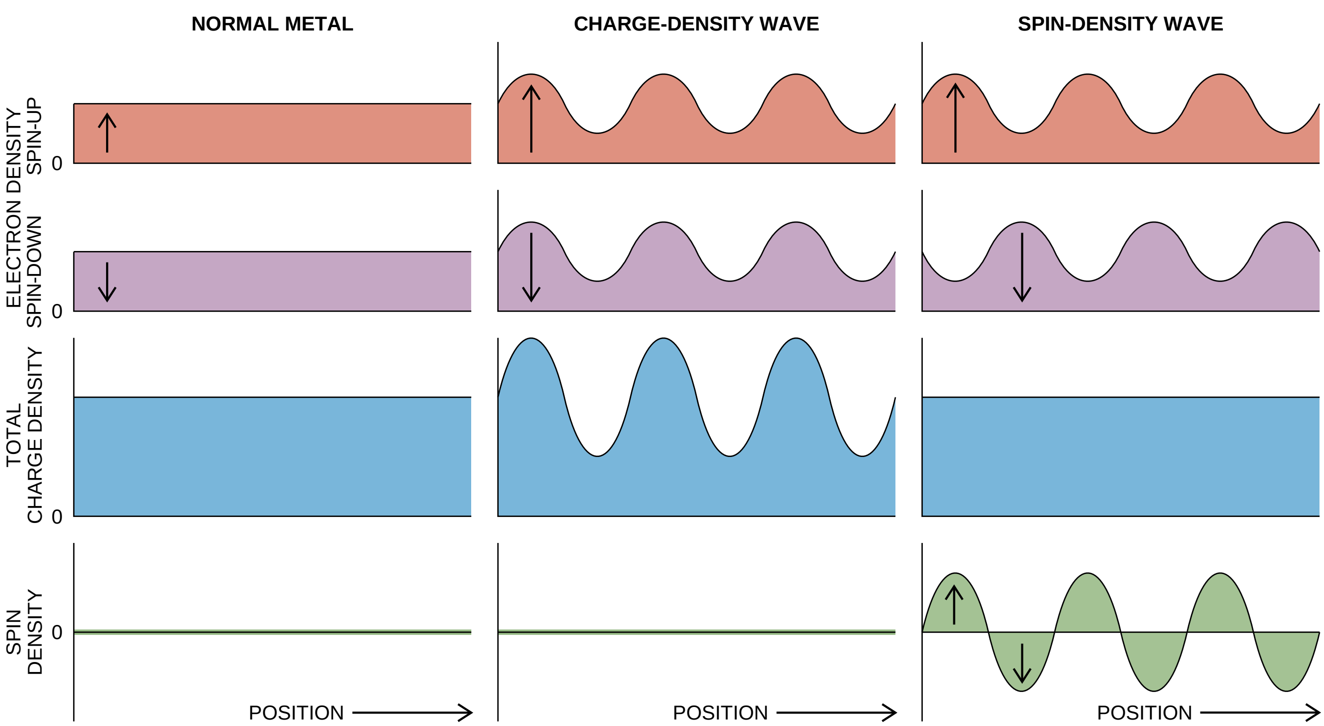
The effects of CDWs and SDWs may also be observed via the motions they perform as a body. These motions can be rather different depending on how the wavelength of the density wave relates to the underlying lattice spacing. The CDW wavelength changes with the number of electrons in the solid: if there are more electrons, the wavelength becomes smaller and, in particular, may not match the original lattice spacing of the ions in any neat way. Then the charge-density wave is said to be "incommensurate" with the original lattice spacing; it floats around unaffected by the lattice until pinned down by a defect. (A defect acts like a pothole—or chewing gum—in the electric potential surface, in which the CDW gets stuck.) But if the charge-density wave and the original lattice spacing are "commensurate" and fit neatly, then, for example, every other student stands in a depression in the road, and it is very hard to get the band moving. For this reason, incommensurate waves are much more intriguing in the varieties of behavior that they display. Commensurate waves—the ones originally envisioned by Peierls—are of largely historical interest.

There are two basic motions that charge-density waves can indulge in as a body, called collective modes. Quantum mechanics allows us to think of these modes as particles, which are then named by the suffix "on." The floating of the crests back and forth and their occasional bunching are a kind of collective mode known as a phason (it involves changing the "phase" of the density wave). For waves that do not fit well with the underlying lattice structure, the floating takes no energy at all (unless a defect pins the wave down), but the bunching takes some. The other way in which CDWs can change is that the crests can get higher. This motion, called an amplitudon, requires a lot of energy. The position and height of the crests may both vary, with variations over shorter distances having higher energies. The energies of these motions were first calculated by Patrick A. Lee, T. Morris Rice and Philip W. Anderson, then at AT&T Bell Laboratories. The SDWs share these collective motions with the CDWs and in addition have a purely magnetic mode that is related to changes in spin orientation. These excitations are called magnons.

The truly dramatic motions occur when we apply an electric field to a solid containing a charge-density wave. A current-voltage relation very different from Ohm's law—in which conductivity is constant—was found in 1986 (in the material niobium triselenide) by Nai-Phuan Ong, Pierre Monceau and Alan M. Portis of the University of California at Berkeley. Since then, some CDW materials have displayed conductivities that vary by several orders of magnitude when very modest electric fields (of less than one volt per centimeter) are applied. We now know that this change in conductivity comes from the depinning and sudden motion of the entire density wave. Even more unusual is the variation of the current as time passes, even when only a constant (DC) voltage is applied. This was first observed by Robert M. Fleming and Charles C. Grimes of AT&T. Our recent measurements, and those of Denis Jérome, Silvia Tomić and others at the University of Paris South and Takashi Sambongi's group at Hokkaido University, have shown that spin-density waves behave much like charge-density waves in the presence of electric fields.

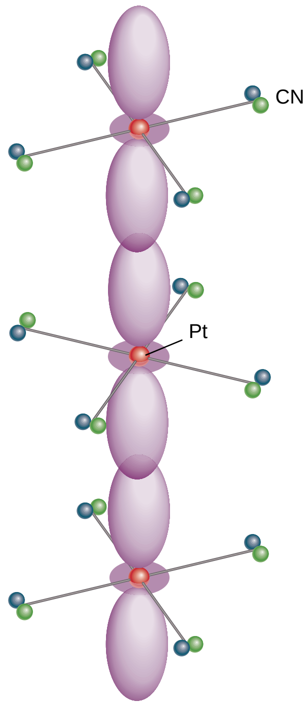
The simplest model that describes the behavior of density waves is called the classical particle model. It was proposed by one of us (Grüner) with Alfred Zawadowski and Paul M. Chaikin, then at the University of California at Los Angeles. The charge-density wave is represented by a single massive particle positioned at its center of mass. The behavior of this particle reflects that of the entire array. When there are no external electric fields, the particle sits on a ribbed surface, like a marble in a cup of an egg tray. This configuration corresponds to the crest of a CDW being stuck at a defect. If we move the CDW, the marble climbs over the edge of the eggcup and falls into the next one, which means that the next crest of the CDW gets stuck at the same defect.

This model allows us to understand much of the versatile behavior of CDWs. The marble is free to move around the bottom of the eggcup and can therefore readjust its position sensitively in re-



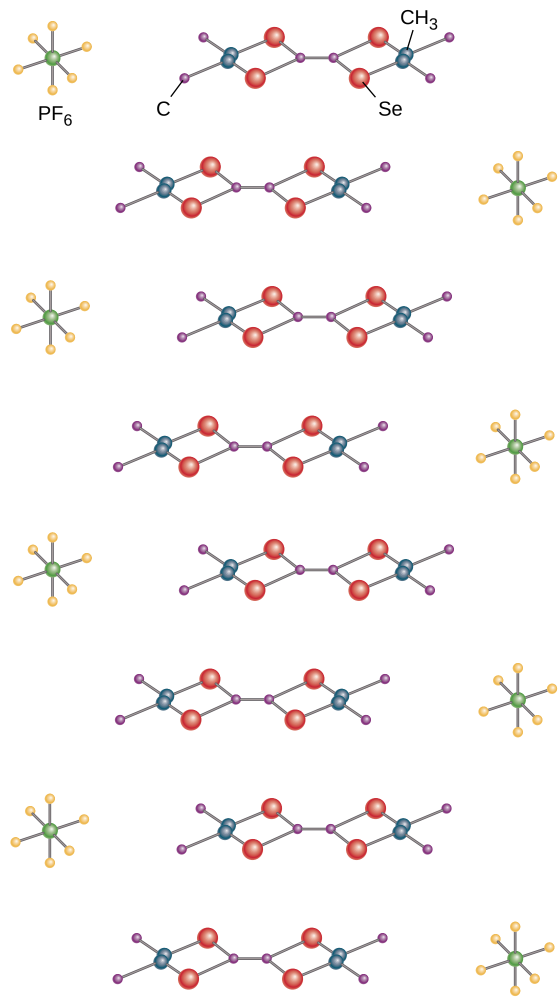
CHARGE AND SPIN DENSITIES of electrons are shown in normal metals, charge-density waves and spin-density waves. The spin-up (*orange*) and spin-down (*purple*) electron densities vary with position within the crystal. They can be summed to yield the total charge density (*blue*); their difference yields the spin density (*green*).

Density-Wave Materials



Two very different types of materials, organic and inorganic, show density waves. The inorganic materials are characterized by chains of transition-metal ions, such as platinum. Within a crystal, each chain is well separated from its neighbors. The electrons move freely along the chains, but the large separation between chains impedes transverse motion, so that the electric conductivity might be from 10 to 1,000 times greater in the chain direction than across. A typical linear-chain compound of the inorganic variety, $K_2Pt(CN)_4Br_{0.3}3H_2O$, or KCP for short, is shown at the left. The vertical lobes (*purple*) show electron orbitals overlapping along the chain. Some other linear-chain compounds having so-called incommensurate charge-density waves are $NbSe_3$, $(TaSe_4)_2I$ and $K_{0.3}MoO_3$.

The other types of CDW materials are grown from flat organic molecules such as the synthetic one, tetramethyltetraselenafulvalene, or TMTSF. Several of these molecules are stacked on one another to form a crystal together with embedded PF_6 ions, as shown at the right. Electrons are free to move up and down the stack but not from one stack to the next; thus, the conductivity is again highly anisotropic.



These organic materials are of particular importance in studying electrons in solids because their properties can be fine-tuned. For example, an entire family of TMTSF-based salts can be created by replacing the PF_6 anion with ReO_4, Br, SCN or AsF_6, among others. Each new salt has slightly different interaction strengths between the electrons; these variations can have profound effects. The crystal $(TMTSF)_2ClO_4$, if cooled slowly down to one kelvin, becomes a superconductor, whereas rapid cooling gives a spin-density wave.

sponse to applied electric fields. Because the marble—that is, the charge-density wave—carries charge, its position affects the electric field within the medium. The marble usually adjusts its position so as to reduce the electric field acting on it. Thus, materials with charge-density waves have a large "dielectric constant," so large that they could be called superdielectrics. Our measurements on both charge- and spin-density waves give values for the dielectric constant more than one million times larger than that of ordinary semiconductors.

What happens if we apply a DC voltage? The egg tray on which the marble lies will tilt. If the tilt—that is, the voltage—is great enough, the marble can roll out of the eggcup and down the egg tray. The marble slows down when it climbs up an edge and speeds up when it falls down one. Consequently, its speed, and the electric current, goes up and down with time. These current oscillations, which we have mentioned earlier, are widely observed. The average current is higher if the tilt in the egg tray—that is, the DC voltage—is higher.

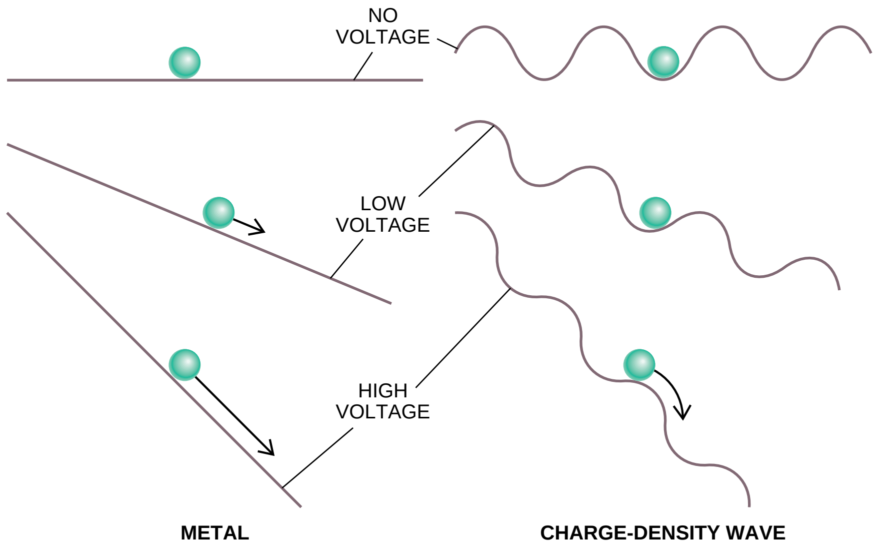
Now suppose that instead of a DC voltage, an AC voltage is applied, in which case the egg tray is rocked back and forth like a seesaw. The marble oscillates back and forth in its cup. This sloshing of the entire density wave scatters light of certain colors, allowing its detection in optical experiments at micron and millimeter wavelengths, as conducted at U.C.L.A. (Conversely, the CDW can sensitively detect electromagnetic radiation.) If we apply both a DC field and an AC field, the former makes the egg tray tilt to one side, whereas the latter makes it jiggle from side to side. Suppose the marble is rolling down the egg tray. If the time the marble takes to go from one eggcup to the next is nearly the same as the time for which the egg tray is tilted "up" by the AC voltage, it will hop between eggcups once each cycle of the AC field. When the marble is hopping down the egg tray with the help of the AC field, augmenting the average tilt of the egg tray by increasing the DC voltage does not change the average current. So if we plot the current versus the DC voltage (in the presence of an AC voltage), we will see the current generally increasing with DC voltage except for certain plateaus where we have a "mode locking" [*see bottom illustration on this page*].

The model we have described, and the equations it implies for the marble's motion, turns out to be applicable in quite diverse situations. For example, it describes a Josephson junction (that between two superconductors), the motion of ions in solids, a pendulum in a gravitational field and certain electronic circuits. Although the equations look simple, they display a variety of solutions, including chaotic behavior.

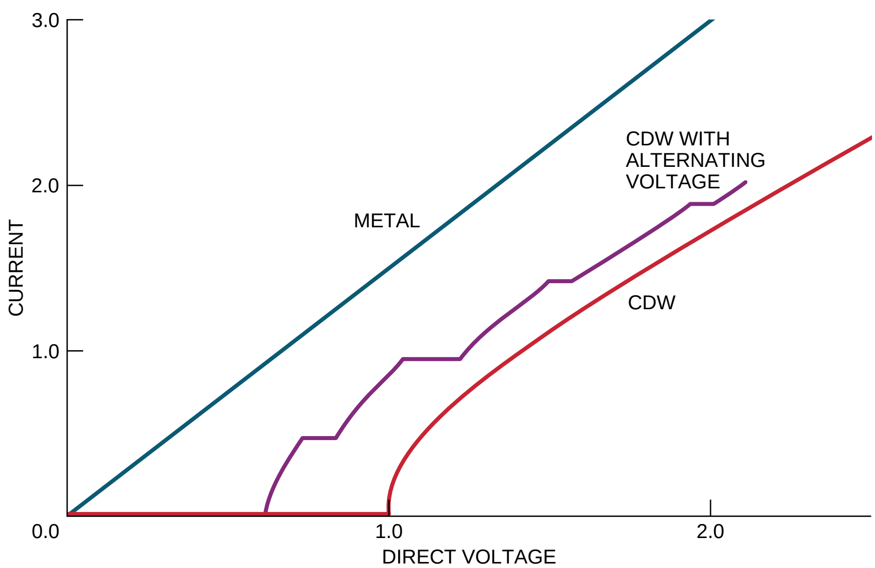
Other behaviors are not so simple to understand. By cooling materials that contain spin-density waves down to almost absolute zero, we find a peculiar phenomenon that has been interpreted as the marble's tunneling through to the next eggcup instead of climbing over its edge. This purely quantum-mechanical effect has since been confirmed by other groups. Tunneling has been predicted by Kazumi Maki of the University of Southern California and by the late John Bardeen of the University of Illinois, but it is too early to tell whether their model applies to our low-temperature experimental findings about SDWs.

Perhaps the most bizarre behavior of all is that of self-organization [see "Self-Organized Criticality," by Per Bak and Kan Chen; SCIENTIFIC AMERICAN, January 1991]. Susan N. Coppersmith and Peter B. Littlewood of AT&T and Kurt A. Wiesenfeld and Per Bak of Brookhaven National Laboratory were the first to deduce this phenomenon, from experiments on CDWs performed by researchers at AT&T and by us at U.C.L.A. Self-organization is a phenomenon that charge-density waves have in common with earthquakes. Just as two tectonic plates rubbing on each other get stuck at ragged edges and then suddenly unstick (with catastrophic consequences), charge-density waves, in the presence of some electric fields, get stuck on defects and suddenly unstick. But the analogy goes deeper. Earthquakes and charge-density waves tend to settle into configurations in which a small disturbance will cause a violent change: they organize themselves into a critical state. The marble balances itself exactly on the thin edge between two eggcups.

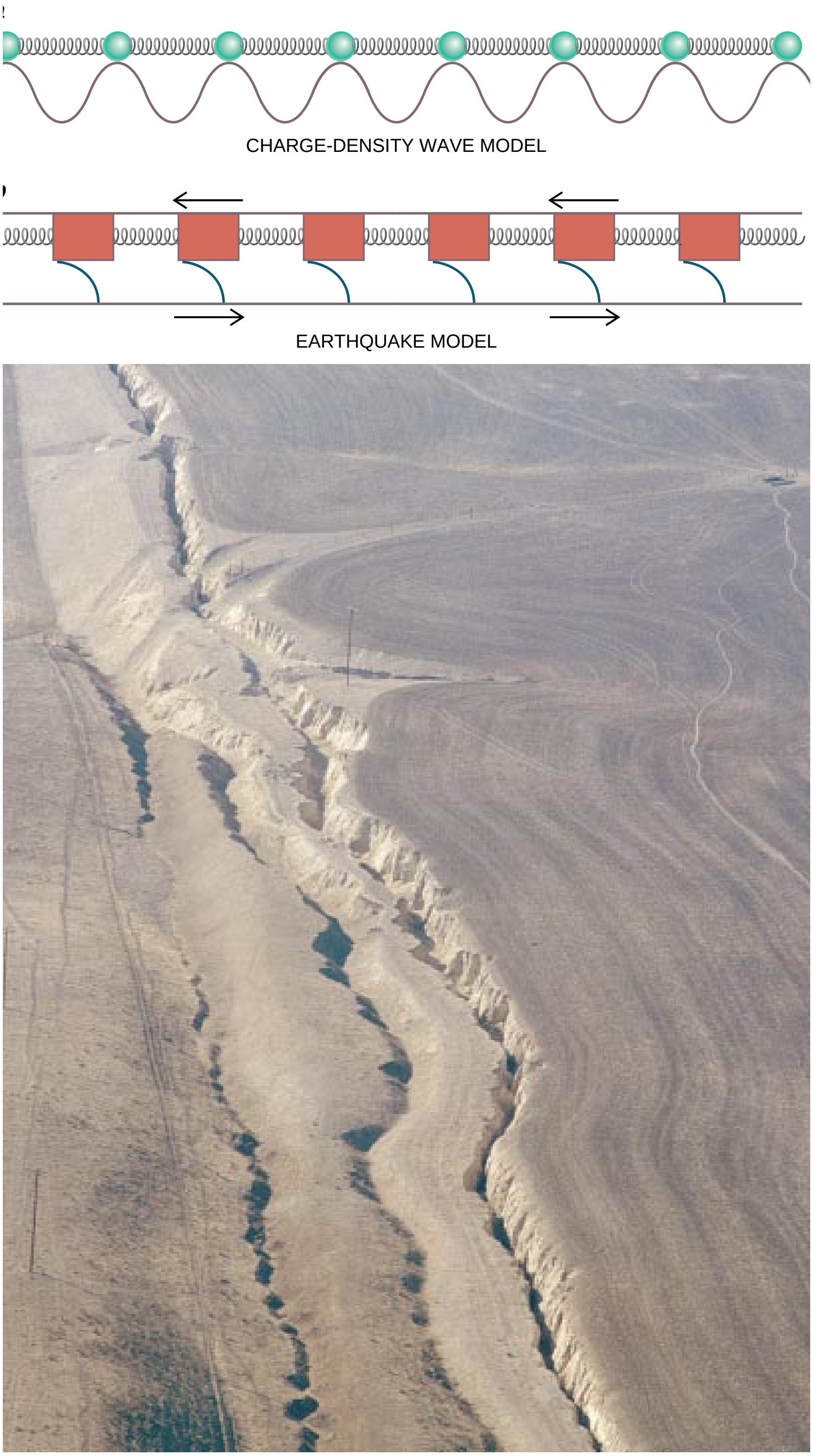
To study the self-organizing behavior, we have to refine our model slightly. Self-organization comes from self-interactions, so our model has to include the push-and-pull between the different regions of the CDW. One marble at the center of mass is no longer enough; we now need a series of marbles attached to their neighbors by springs. This arrangement represents the elasticity of the density wave. Suppose we repeatedly turn on a DC electric field for some time and then turn it off. The marbles move some distance during the "on" time and roll to the bottoms of eggcups when the field is turned off. We would expect them to move farther if the "on" time is longer. But what happens is actually quite different, as was found in simulations by Coppersmith. Just before the field is turned on, the marbles are positioned in neighboring eggcups. When the field is turned off, each is found to be exactly balanced on an edge between two eggcups—no matter how long the field is kept on [*see illustration on next page*]. (After the field is turned off, the marbles roll into eggcups, sometimes to their right and sometimes to their



CHARGED-PARTICLE MODEL shows how current flow in charge-density waves differs from that in normal metals. In a metal (*left*) the particle rests on a flat (electrical potential) surface. If we apply a voltage, the surface tilts, and the particle starts to move: there is a current. For a charge-density wave (*right*), the surface is ribbed. If the applied voltage is low—that is, the tilt is small—the particle changes position only slightly, and there is no current. If the tilt is large enough to get the particle over the barrier, the particle runs down the ribbed surface. Then the current goes up and down in time as the particle climbs over each barrier.



CURRENT VERSUS VOLTAGE is plotted for metals and charge-density waves. In a metal (*blue*) the current increases linearly with voltage. For a charge-density wave, there is no current until the voltage increases to a critical value; only then does the current start to flow (*red*). If in addition to a direct voltage we apply an alternating one, the curve shows plateaus (*purple*). The plateaus correspond to a "mode locking" when the flow of the charge-density wave matches the alternating frequency.



SELF-ORGANIZED BEHAVIOR is compared for models of charge-density waves and earthquake faults. The particles attached by springs represent the position and elasticity of the charge-density wave (*a*). The particles rest on a ribbed surface. If one turns an electric field on and off repeatedly, the marbles are found to sit in the most unstable position possible: each on top of a hill. In the earthquake model (*b*) the blocks are attached to a surface moving sideways with respect to the lower surface. The lower surface has metal strips that drag on the blocks; the blocks are also connected by springs. After some arbitrary (unpredictable) time, the accumulated strain makes the blocks rearrange their positions catastrophically. But the new positions are again unstable. The photograph shows the San Andreas fault in Carrizo Plain, east of San Luis Obispo, Calif.

left.) This bizarre self-regulatory behavior is easier to study in charge-density waves than in earthquakes. It has given the former a particular use in testing complex dynamical theories.

In fact, the density-wave states are probably just the simplest periodic configurations of electrons we can hope to find. Several theories suggest a hierarchy of more complex arrangements. One suggestion came from theoretical physicist Eugene Wigner in 1939. Wigner showed that if the density of electrons is low enough—say in a collection of electrons moving freely in two dimensions—they would settle into a crystalline pattern. Since then, many researchers have searched for "Wigner crystals." In the early 1980s Grimes and Gregory Adams, also at AT&T, showed that electrons deposited on the surface of liquid helium form just such a crystal. Evidence of their presence in solid-state systems has come from groups at Saclay, France, AT&T and elsewhere.

The various properties of density-wave materials have yet to be applied toward enhancing our comfort. Still, plans abound. The dielectric constants of CDW materials, besides being enormous, also change with the electric field; they could be implemented in circuits as tunable capacitors. The strong response of charge-density waves to electromagnetic radiation could make them useful as light detectors; at low temperatures, this sensitivity would ultimately be limited by quantum mechanics. Bardeen, better known for the theory of superconductivity and the invention of the solid-state transistor, worked out the theory of the quantum transport of density waves. Whether quantum detectors such as he envisaged can be built and put to practical use remains to be seen. Right now, we are happy enough just to learn more about the idiosyncracies of charge- and spin-density waves.

FURTHER READING

THE DIFFERENCE BETWEEN ONE-DIMENSIONAL AND THREE-DIMENSIONAL SEMICONDUCTORS. Esther M. Conwell in *Physics Today,* Vol. 38, No. 6, pages 46–53; June 1985.

THE DYNAMICS OF CHARGE-DENSITY WAVES. G. Grüner in *Reviews of Modern Physics,* Vol. 60, No. 4, pages 1129–1181; October 4, 1988.

CHARGE DENSITY WAVES IN SOLIDS. Edited by L. P. Gor'kov and G. Grüner. Elsevier, 1990.

EVIDENCE ACCUMULATES, AT LAST, FOR THE WIGNER CRYSTAL. Anil Khurana in *Physics Today,* Vol. 43, No. 12, pages 17–20; December 1990.

Visualizing the Mind

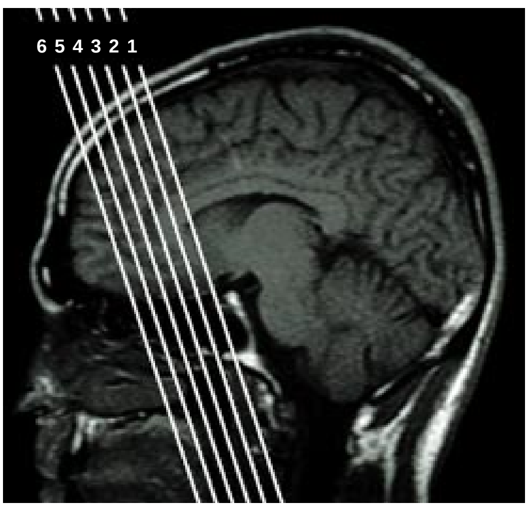
Strategies of cognitive science and techniques of modern brain imaging open a window to the neural systems responsible for thought

by Marcus E. Raichle

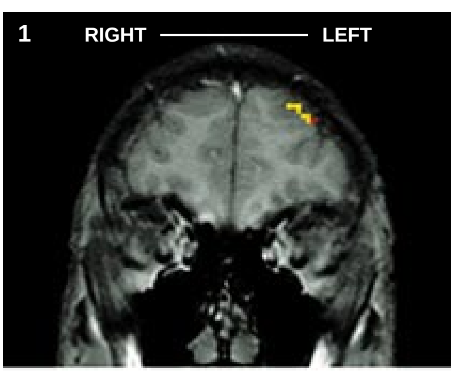
What causes the pity we might feel for the melancholy Dane in *Hamlet* or the chill during a perusal of the *Raven?* Our brains have absorbed from our senses a printed sequence of letters and then converted them into vivid mental experiences and potent emotions. The "black box" description of the brain, however, fails to pinpoint the specific neural processes responsible for such mental actions. While philosophers have for centuries pondered this relation between mind and brain, investigators have only recently been able to explore the connection analytically—to peer inside the black box. The ability stems from developments in imaging technology that the past few years have seen, most notably positron-emission tomography and magnetic resonance imaging. Coupled with powerful computers, these techniques can now capture in real time images of the physiology associated with thought processes. They show how specific regions of the brain "light up" when activities such as reading are performed and how neurons and their elaborate cast of supporting cells organize and coordinate their tasks. The mapping of thought can also act as a tool for neurosurgery and elucidate the neural differences of people crippled by devastating mental illnesses, including depression and schizophrenia.

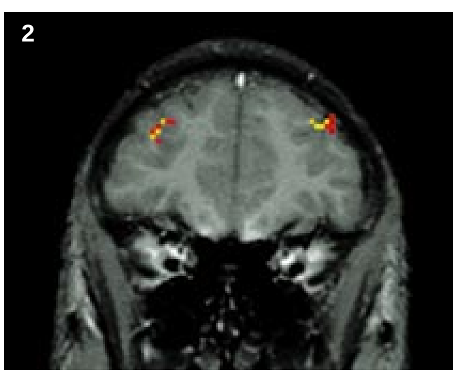
I hasten to point out that the underlying assumptions of current brain mapping are distinct from those held by early phrenologists. They posited that single areas of the brain, often identified by bumps on the skull, uniquely represented specific thought processes and emotions. In contrast, modern thinking posits that networks of neurons residing in strictly localized areas perform thought processes. So just as specific members of a large orchestra perform together in a precise fashion to produce a symphony, a group of localized brain areas performing elementary operations work together to exhibit an observable human behavior. The foundation for such analyses is that complex behaviors can be broken down into a set of constituent mental operations. In order to read, for example, one must recognize that a string of letters is a word; then recognize the meaning of words, phrases or sentences; and finally create mental images.

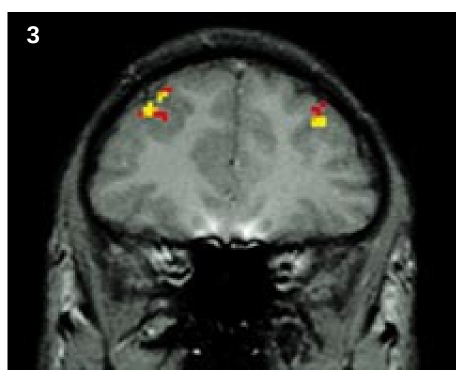
The challenge, of course, is to determine those parts of the brain that are active and those that are dormant during the performance of tasks. In the past, cognitive neuroscientists have relied on studies of laboratory animals and patients with localized brain injuries to gain insight into the brain's functions. Imaging techniques, however, permit us to visualize safely the anatomy and the function of the normal human brain.



ACTIVE NEURAL AREAS from a subject remembering a sequence of letters are mapped by magnetic resonance imaging. The images below represent six slices through the frontal cortex. The slices are identified by numbers in the corners that correspond to those in the scan at the left. Red, orange and yellow represent areas of increasing activity. Jonathan D. Cohen and his colleagues at the University of Pittsburgh and Carnegie Mellon University formed the images.







MARCUS E. RAICHLE is professor of neurology, radiology and neurobiology as well as a senior fellow of the McDonnell Center for Studies of Higher Brain Function at the Washington University School of Medicine in St. Louis. He received his B.S. and M.D. degrees from the University of Washington in Seattle. He began researching brain metabolism and circulation when he was a neurology resident at the New York Hospital–Cornell Medical Center. His current focus is the use of positron-emission tomography and magnetic resonance imaging to study human cognition and emotion.

The modern era of medical imaging began in the early 1970s, when the world was introduced to a remarkable technique called x-ray computed tomography, now known as x-ray CT, or just CT. South African physicist Allan M. Cormack and British engineer Sir Godfrey Hounsfield independently developed its principles. Hounsfield constructed the first CT instrument in England. Both investigators received the Nobel Prize in 1979 for their contributions.

Computed tomography takes advantage of the fact that different tissues absorb varying amounts of x-ray energy. The denser the tissue, the more it absorbs. A highly focused beam of x-rays traversing through the body will exit at a reduced level depending on the tissues and organs through which it passed. A beam of x-rays passed through the body at many different angles through a plane collects sufficient information to reconstruct a picture of the body section. Crucial in the development of x-ray CT was the emergence of clever computing and mathematical techniques to process the vast amount of information necessary to create images themselves. Without the availability of sophisticated computers, the task would have been impossible to accomplish.

X-ray CT had two consequences. First, it changed forever the practice of medicine because it was much superior to standard x-rays. For the first time, investigators could safely and effectively view living human tissue such as the brain with no discomfort to the patient. Standard x-rays revealed only bone and some surrounding soft tissue. Second, it immediately stimulated scientists and engineers to consider alternative ways of creating images of the body's interior using similar mathematical and computer strategies for image reconstruction.

One of the first such groups to be intrigued by the possibilities opened by computed tomography consisted of experts in tissue autoradiography, a method used for many years in animal studies to investigate organ metabolism and blood flow. In tissue autoradiography, a radioactively labeled compound is injected into a vein. After the compound has accumulated in the organ (such as the brain) under interest, the animal is sacrificed and the organ removed for study. The organ is carefully sectioned, and the individual slices are laid on a piece of film sensitive to radioactivity. Much as the film in a camera records a scene as you originally viewed it, this x-ray film records the distribution of radioactively labeled compound in each slice of tissue.

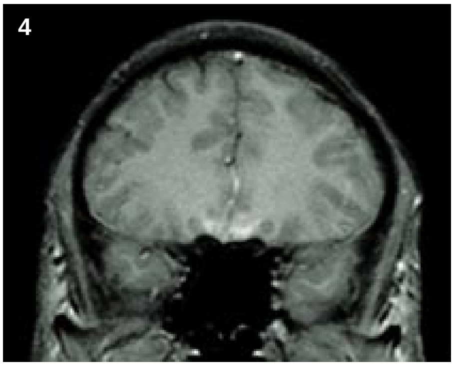
Once the x-ray film is developed, scientists have a picture of the distribution of radioactivity within the organ and hence can deduce the organ's specific functions. The type of information is determined by the radioactive compound injected. A radioactively labeled form of glucose, for example, measures brain metabolism because glucose is the primary source of energy for neurons. Louis Sokoloff of the National Institute of Mental Health introduced this now widely used autoradiographic method in 1977.

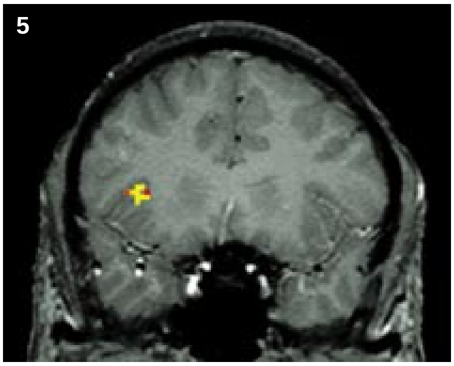
Investigators adept with tissue autoradiography became fascinated when CT was introduced. They suddenly realized that if they could reconstruct the anatomy of an organ by passing an x-ray beam through it, they could also reconstruct the distribution of a previously administered radioisotope. One had simply to measure the emission of radioactivity from the body section. With this realization was born the idea of autoradiography of living human subjects.

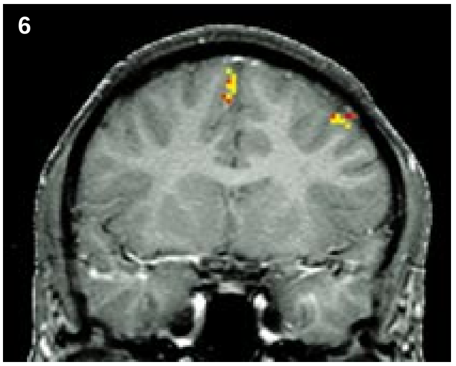
A crucial element in the evolution of human autoradiography was the choice of radioisotope. Workers in the field selected a class of radioisotopes that emit positrons, which resemble electrons except that they carry a positive charge. A positron would almost immediately combine with a nearby electron. They would annihilate each other, emitting two gamma rays in the process. Because each gamma ray travels in nearly opposite directions, devices around the sample would detect the gamma rays and locate their origin. The crucial role of positrons in human autoradiography gave rise to the name positron-emission tomography, or PET [see "Positron-Emission Tomography," by Michel M. Ter-Pogossian, Marcus E. Raichle and Burton E. Sobel; SCIENTIFIC AMERICAN, October 1980].

Throughout the late 1970s and early 1980s, researchers rapidly developed PET to measure various activities in the brain, such as glucose metabolism, oxygen consumption, blood flow and interactions with drugs. Of these variables, blood flow has proved the most reliable indicator of moment-to-moment brain function.

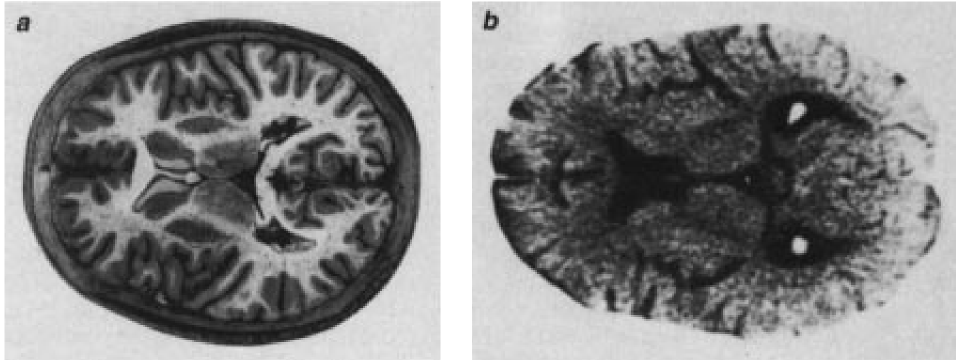
The idea that local blood flow is intimately related to brain function is a surprisingly old one. English physiologists Charles S. Roy and Charles S. Sherrington formally presented the idea in a publication in 1890. They suggested that an "automatic mechanism" regulated the blood supply to the brain. The amount of blood depended on local variations in activity. Although subsequent experiments have amply confirmed the existence of such an automatic mechanism, no one as yet is entirely certain about its exact nature. It obviously remains a challenging area for research.





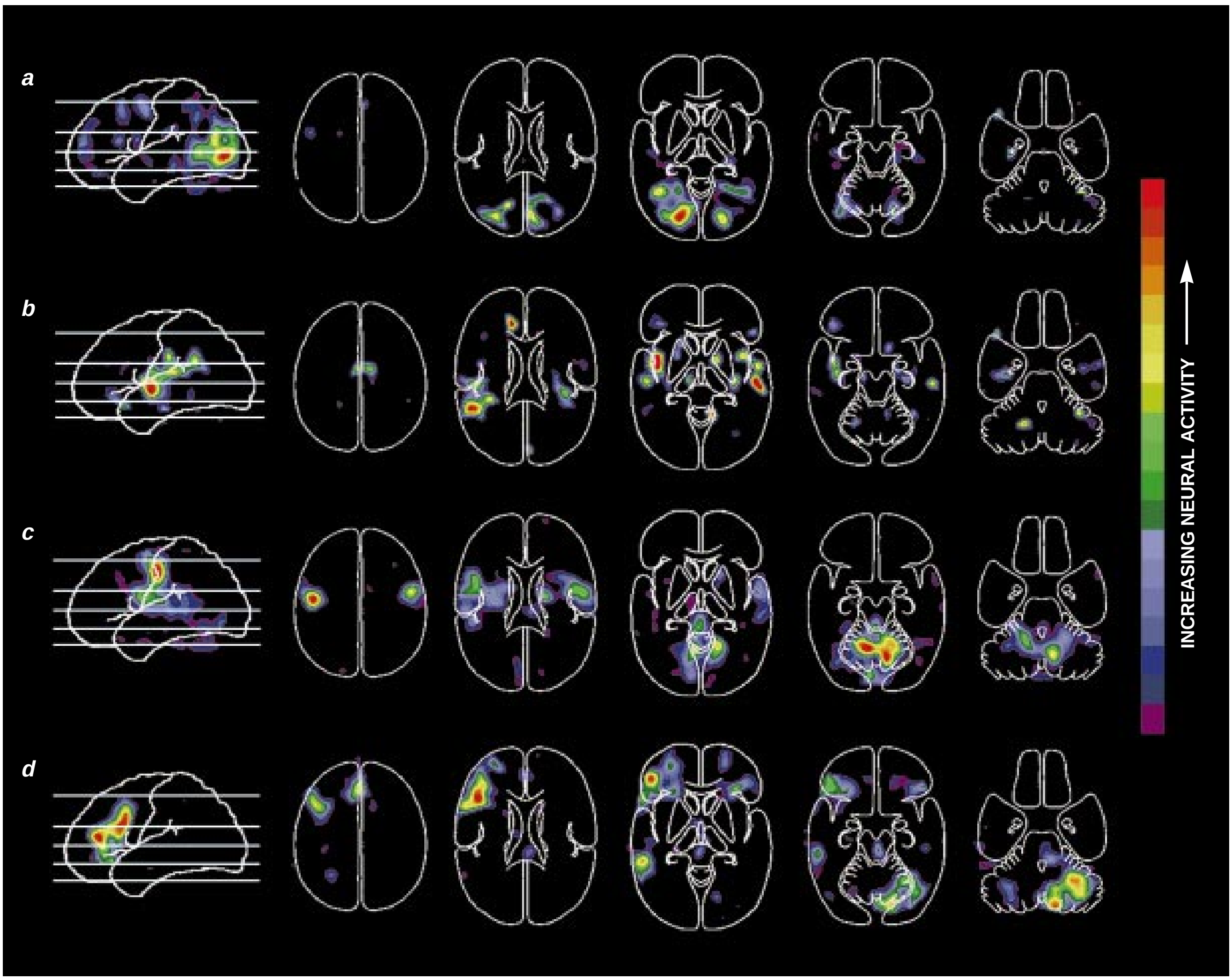


BRAIN SECTION (*a*) is compared with corresponding images of the slice taken by x-ray computed tomography (CT) (*b*), positron-emission tomography (PET) (*c*) and magnetic resonance imaging (MRI) (*d*). CT depicts the features of the brain section, whereas PET shows the amount of neuronal work (the darker the area, the greater the activity). Properly set up, MRI can do both tasks. Here it images structure.

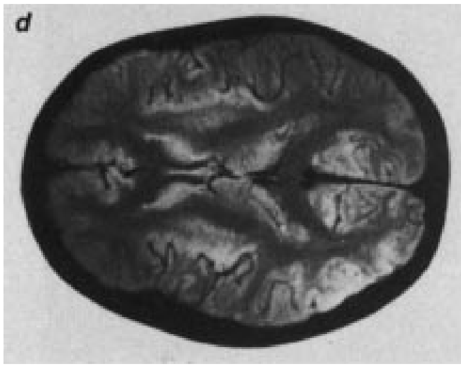


PET measures blood flow in the normal human brain by adapting an autoradiographic technique for laboratory animals developed in the late 1940s by Seymour S. Kety of the National Institute of Mental Health and his colleagues. PET relies on radioactively labeled water—specifically, hydrogen combined with oxygen 15, a radioactive isotope of oxygen. The labeled water emits copious numbers of positrons as it decays (hydrogen isotopes cannot be used, because none emit positrons). The labeled water is administered into a vein in the arm. In just over a minute the radioactive water accumulates in the brain, forming an image of blood flow.

The radioactivity of the water produces no deleterious effects. Oxygen 15 has a half-life of only two minutes; an



PET SCANS show active neural areas. In the far left column the left side of the brain is presented; the next columns show five horizontal layers (the right side faces to the right, with the front to the top). Each row corresponds to the difference between a specific task and the control state of gazing at a dot on a television monitor. When subjects passively view nouns (*a*), the primary visual cortex lights up. When nouns are heard (*b*), the temporal lobes take command. Spoken nouns minus viewed or heard nouns (*c*) reveal motor areas used for speech. Generating verbs (*d*) requires additional neural zones, including those in the left frontal and temporal lobes corresponding roughly to Broca's and Wernicke's areas.



entire sample decays almost completely in about 10 minutes (five half-lives) into a nonradioactive form. The rapid decay substantially reduces the exposure of subjects to the potentially harmful effects of radiation. Moreover, only low doses of the radioactive label are necessary.

The fast decay and small amounts permit many measurements of blood flow to be made in a single experiment. In this way, PET can take multiple pictures of the brain at work. Each picture serves as a snapshot capturing the momentary activity within the brain. Typical PET systems can locate changes in activity with an accuracy of a few millimeters.

A distinct strategy for the functional mapping of neuronal activity by PET has emerged during the past 10 years. This approach extends an idea first introduced to psychology in 1868 by Dutch physiologist Franciscus C. Donders. Donders proposed a general method to measure thought processes based on a simple logic. He subtracted the time needed to respond to a light (with, say, the press of a key) from the time needed to respond to a particular color of light. He found that discriminating color required about 50 milliseconds. In this way, Donders isolated and measured a mental process for the first time.

The current PET strategy is designed to accomplish a similar subtraction but in terms of the brain areas implementing the mental process. In particular, images of blood flow taken before a task is begun are compared with those obtained when the brain is engaged in that task. Investigators refer to these two periods as the control state and the task state. Workers carefully choose each state so as to isolate as best as possible a limited number of mental operations. Subtracting blood-flow measurements made in the control state from each task state indicates those parts of the brain active during a particular task.

To achieve reliable data, workers take the average of responses across many individual subjects or of many experimental trials in the same person. Averaging enables researchers to detect changes in blood flow associated with mental activity that would otherwise be easily confused with spurious shifts resulting from noise.

One of the first assignments in which PET blood-flow mapping has proved useful is in the study of language. The manner in which language skills are acquired and organized in the human brain has been the subject of intense investigation for more than a century. Work began in earnest in 1861, when French physician Pierre Paul Broca described a patient whose damaged left frontal lobe destroyed the ability to speak. (To this day, patients who have frontal lobe damage and have trouble speaking are often referred to as having Broca's aphasia.) Broca's studies of language localization were complemented by Carl Wernicke, a German neurologist. In 1874 Wernicke told of people who had difficulty comprehending language. They harbored damage to the left temporal lobe, a region now usually referred to as Wernicke's area. From these beginnings has emerged a concept of language organization in the human brain: information flows from visual and auditory reception to areas in the left temporal lobe for comprehension and then on to frontal areas for speech production [see "Specializations of the Human Brain," by Norman Geschwind; SCIENTIFIC AMERICAN, September 1979].

All this information was gleaned from brain-damaged patients. Can investigators derive insight about language organization from a healthy brain? In 1988 my colleagues Steven E. Petersen, Michael I. Posner, Peter T. Fox and Mark A. Mintun and I at the Washington University Medical Center began a series of studies to answer just this question. The initial study was based on a PET analysis of a seemingly simple job: speaking an appropriate verb when presented with a common English noun. For example, a subject might see or hear the word "hammer," to which an appropriate response might be "hit."

We chose this assignment because it could be broken down into many components. Each component could separately be analyzed through a careful selection of tasks. The most readily apparent elements include visual and auditory word perception, the organization and execution of word output (speech), and the processes by which the brain retrieves the meanings of words. (Of course, each of these operations can be divided further into several additional subcomponents.)

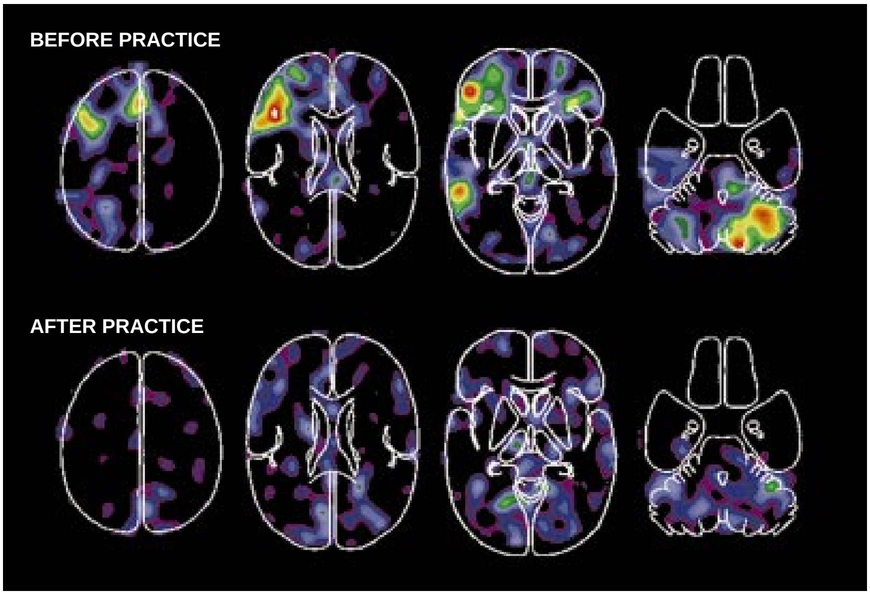
To identify the areas of the brain used in a particular operation, we composed four levels of information processing. Such a hierarchy has become standard among laboratories doing this type of research [*see bottom illustration on opposite page*]. In the first level, subjects were asked to fix their gaze on a pair of small crosshairs—the arrangement looks like a small plus sign—in the middle of a television monitor. At the same time, a PET scan measured blood flow in the brain, providing a snapshot of mental activity.

In the second level, subjects continued to maintain their gaze on the crosshairs as blood flow was measured, but during this scan they were exposed to common English nouns. The nouns either appeared below the crosshairs on the television monitor or were spoken through earphones (separate scans were performed for visual and auditory presentations). In the third level, subjects were asked to recite the word they viewed or heard. Finally, in the fourth level, the subjects said out loud a verb appropriate for the noun.

Subtracting the first level from the second isolated those brain areas concerned with visual and auditory word perception. Deducting the second level from the third pinpointed those parts of the brain concerned with speech production. Subtracting level three from level four located those regions concerned with selecting the appropriate verb to a presented noun.

The final subtraction (speaking nouns minus generating verbs) was of particular interest, because it provided a portrait of pure mental activity (perception and speech—or input and output—having been subtracted away). This image permitted us to view what occurs in our brains as we interpret the meaning of words and, in turn, express meaning through their use. It renders visible conscious function because much of our thinking is carried out by concepts and ideas represented by words.

The results of this study clearly demonstrate how brain imaging can relate



LEARNING-INDUCED CHANGES in neural activity are revealed by PET imaging. The top row shows the brain of a subject who must quickly generate verbs appropriate to visually presented nouns. The bottom row shows the result of 15 minutes of practice; the regions activated are similar to those used in simply reading out loud.

mental operations of a behavioral task to specific networks of brain areas orchestrated to perform each operation. As anticipated by cognitive scientists and neuroscientists, the apparently simple task of generating a verb for a presented noun is not accomplished by a single part of the brain but rather by many areas organized into networks. Perception of visually presented words occurs in a network of areas in the back of the brain, where many components of the brain's visual system reside. Perception of aurally presented words occurs in an entirely separate network of areas—in our temporal lobes.

Speech production (that is, simply repeating out loud the presented nouns) predictably involves motor areas of the brain. Regions thought to be Broca's and Wernicke's areas do not appear to be engaged routinely in this type of speech production, an activity that would be viewed by many as quite automatic for most fluent speakers in their native language. This finding suggests what we might have suspected: we occasionally speak without consciously thinking about the consequences.

Regions of the left frontal and temporal lobes (those corresponding in general to the respective locations of Broca's and Wernicke's areas) only become active when two tasks are added: consciously assessing word meaning and choosing an appropriate response. Moreover, two other areas come into play under these circumstances, forming a network of four brain regions. Interestingly, two areas used in the routine repetition of words were turned off. This shutdown suggests that the demands of generating a verb to a presented noun does not simply build on the task of just saying the noun. Rather the act of speaking a verb to a presented noun differs from speaking the noun, as far as the brain is concerned.

This finding caused us to pause and consider what would happen if we allowed subjects a few minutes of practice on their task of generating verbs. Although subjects initially discover that forming verbs rapidly is difficult (nouns are presented every 1.5 seconds), they become relaxed and proficient after 15 minutes of practice. An examination of the brain after training reveals that practice completely changes the neural circuits recruited [*see top illustration on this page*]. The circuits responsible for noun repetition now generate the verbs. Thus, practice not only makes perfect (something we have always known) but also changes the way our brain organizes itself (something we may not have fully appreciated).

As cognitive neuroscientists demonstrated the utility of PET technology, a newer method swiftly emerged that could compete with PET's abilities. Magnetic resonance imaging, or MRI, has now become a fairly common tool for diagnosing tissue damage. Recent developments have vastly increased the speed with which MRI can form images, thus making it suitable for research in cognitive neuroscience.

MRI derives from a potent laboratory technique known as nuclear magnetic resonance (NMR), which was designed to explore detailed chemical features of molecules. It garnered a Nobel Prize for its developers, Felix Bloch of Stanford University and Edward M. Purcell of

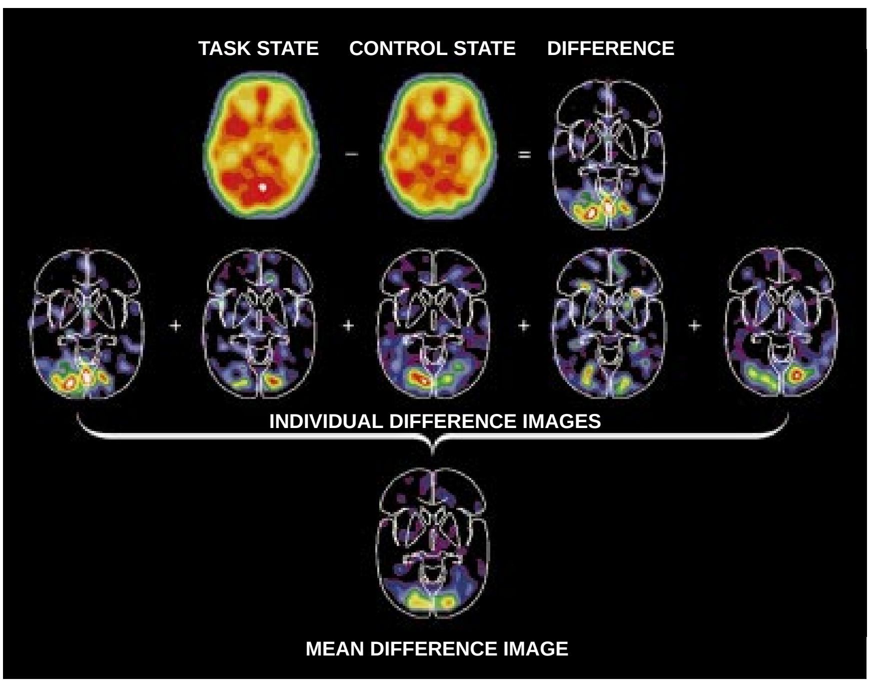


IMAGE SUBTRACTION AND AVERAGING serve as the foundation of functional brain imaging. Researchers subtract the PET blood-flow pattern of a control state from that of a task state to produce a difference image (*top row*). Data from different subjects are averaged (*bottom two rows*) to eliminate statistical fluctuations.

Harvard University, in 1952. The method depends on the fact that many atoms behave as little compass needles in the presence of a magnetic field. By skillfully manipulating the magnetic field, scientists can align the atoms. Applying radio-wave pulses to the sample under these conditions perturbs the atoms in a precise manner. As a result, they emit detectable radio signals unique to the number and state of the particular atoms in the sample. Careful adjustments of the magnetic field and the radio-wave pulses yield particular information about the sample under study.

NMR moved from the laboratory to the clinic when Paul C. Lauterbur of the University of Illinois found that NMR can form images by detecting protons. Protons are useful because they are abundant in the human body and, by acting as little compass needles, respond sensitively to magnetic fields. Their application resulted in excellent images of the anatomy of organs that far surpassed in detail those produced by x-ray CT [see "NMR Imaging in Medicine," by Ian L. Pykett; SCIENTIFIC AMERICAN, May 1982]. Because the term "nuclear" made the procedure sound dangerous, NMR soon became known as magnetic resonance imaging.

The current excitement over MRI for brain imaging stems from the technique's ability to detect a signal inaccessible to PET scans. Specifically, it can detect an increase in oxygen that occurs in an area of heightened neuronal activity. The basis for this capacity comes from the way neurons make use of oxygen. PET scans had revealed that functionally induced increases in blood flow accompanied alterations in the amount of glucose the brain consumed but not in the amount of oxygen it used. In effect, the normal human brain during spurts of neuronal activity resorts to anaerobic metabolism. Few had suspected that the brain might rely on tactics similar to those used by sprinter's muscles. In fact, this form of metabolism occurs despite the presence of abundant oxygen in the normal brain. Why the brain acts this way is a mystery worthy of intense scientific scrutiny.

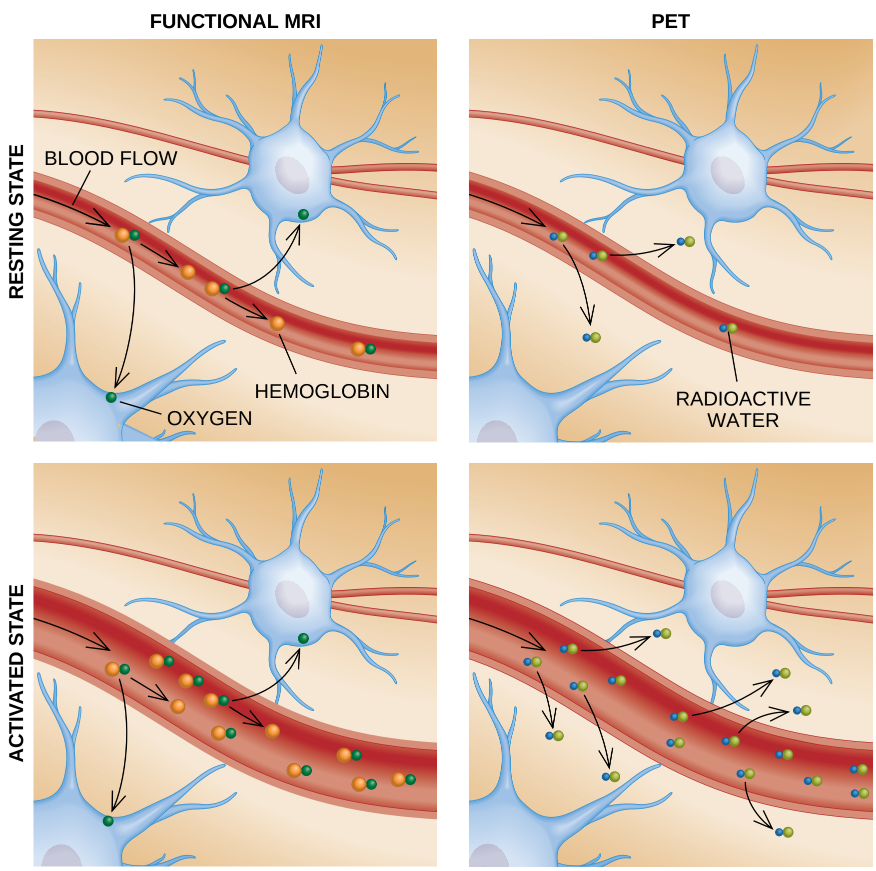
Additional blood to the brain without a concomitant increase in oxygen consumption leads to a heightened concentration of oxygen in the small veins draining the active neural centers. The reason is that supply has increased, but the demand has not. Therefore, the extra oxygen delivered to the active part of brain simply returns to the general circulation by way of the draining veins.

Why does oxygen play a crucial role in MRI studies of the brain? The answer lies in a discovery made by Nobel laureate Linus C. Pauling in 1935. He found that the amount of oxygen carried by hemoglobin (the molecule that transports oxygen and gives blood its red color) affects the magnetic properties of the hemoglobin. In 1990 Seiji Ogawa and his colleagues at AT&T Bell Laboratories demonstrated that MRI could detect these small magnetic fluctuations. Several research groups immediately realized the importance of this observation. By the middle of 1991 investigators showed that MRI can detect the functionally induced changes in blood oxygenation in the human brain. The ability of MRI machines to detect functionally induced changes in blood oxygenation leads many to refer to the technique as functional MRI, or fMRI.

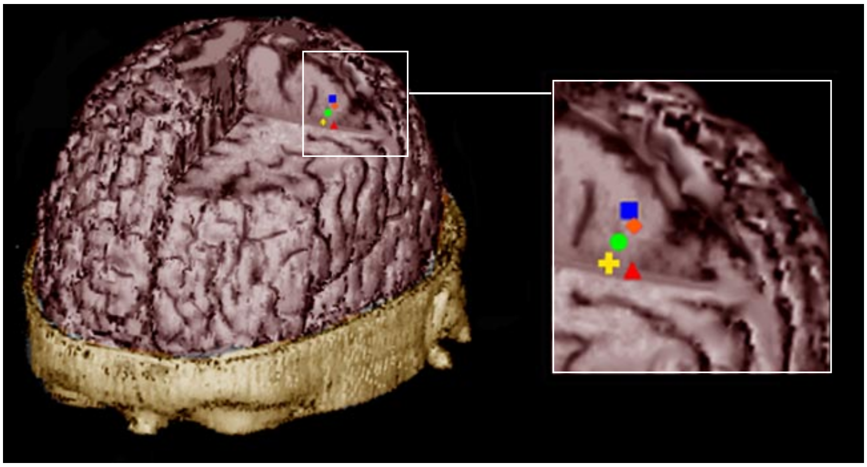
Functional MRI has several advantages over x-ray CT and other imaging techniques. First, the signal comes directly from functionally induced changes in the brain tissue (that is, the change in venous oxygen concentration). Nothing, radioactive or otherwise, needs to be injected to obtain a signal. Second, MRI provides both anatomical and functional information in each subject, hence permitting an accurate structural identification of the active regions. Third, the spatial resolution is quite good, distinguishing parts as small as one to two millimeters (better than PET's resolution). Fourth, when properly equipped (that is, given so-called echoplanar capability), MRI can monitor the rate of change in the blood-flow-induced oxygen signal in real time [*see illustration below*].

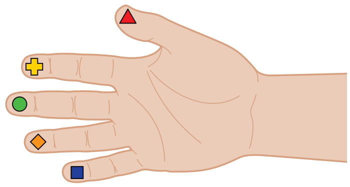
Finally, MRI has little, if any, known biological risk. Some workers have raised concerns about the intensity of the magnetic field to which the tissues are exposed. So far most studies have found the effects to be benign. The largest drawback is the claustrophobia some subjects may suffer. In most instrument designs the entire body must be inserted into a relatively narrow tube.

Several intriguing results with functional MRI were reported this past year. Robert G. Shulman and his colleagues at Yale University have confirmed PET findings about language organization in the brain. Using conventional, hospital-based MRI, Walter Schneider and Jonathan D. Cohen and their colleagues



BLOOD FLOW to the brain provides the signals detected by functional MRI and PET. When resting neurons (*top*) become active (*bottom*), blood flow to them increases. MRI (*left*) detects changes in oxygen levels, which rise in the nearby blood vessels because active neurons consume no more oxygen than when they are at rest. PET (*right*) relies on the increased delivery of injected radioactive water, which diffuses out of the vessels to reach all parts of the brain.





MAGNETOENCEPHALOGRAPHY, or MEG, captures neural activity too brief to be detected by PET or MRI. Above, MEG has located the areas in the normal adult somatosensory cortex associated with the digits of the right hand (*colored symbols*). The symbols on the MRI image of the brain correspond to those on the fingers.

at the University of Pittsburgh have corroborated work in monkeys that indicated the primate visual cortex is organized into topographic maps that reflect the spatial organization of the world as we see it. Other groups are actively trying to visualize other forms of mental activity, such as the way we create mental images and memories.

The ability of MRI systems to monitor the oxygen signal in real time has suggested to some the possibility of measuring the time it takes for different brain areas to exchange information. Conceptually, one might think of a network of brain areas as a group of individuals in the midst of a conference call. The temporal information sought would be equivalent to knowing who was speaking when and, possibly, who was in charge. Such information would be critical in understanding how specific brain areas coordinate as a network to produce behavior.

The stumbling block, however, is the speed of neuronal activity compared with the rate of change of oxygenation levels. Signals from one part of the brain can travel to another in 0.01 second or less. Unfortunately, changes in blood flow and blood oxygenation are much slower, occurring hundreds of milliseconds to several seconds later. MRI would not be able to keep up with the "conversations" between brain areas. The only methods that respond quickly enough are electrical recording techniques. Such approaches include electroencephalography (EEG), which detects brain electrical activity from the scalp, and magnetoencephalography (MEG), which measures the magnetic fields generated by electrical activity within the brain.

Why don't researchers just use EEG or MEG for the whole job of mapping brain function? The limitations are spatial resolution and sensitivity. Even though great strides in resolution have been made, especially with MEG [*see illustration above*], accurate localization of the source of brain activity remains difficult with electrical recording devices. Furthermore, the resolution becomes poorer the deeper into the brain we attempt to image.

Neither MRI nor PET suffers from this difficulty. They both can sample all parts of the brain with equal spatial resolution and sensitivity. As a result, a collaboration seems to be in the making between PET and MRI and electrical recording. PET and MRI, working in a combination yet to be determined, can define the anatomy of the circuits underlying a behavior of interest; electrical recording techniques can reveal the course of temporal events in these spatially defined circuits.

Regardless of the particular mix of technologies that will ultimately be used to image human brain function, the field demands extraordinary resources. Expensive equipment dominates this work. MRI, PET and MEG equipment costs from $2 million to $4 million and is expensive to maintain. Furthermore, success requires close collaboration within multidisciplinary teams of scientists and engineers working daily with these tools. Institutions fortunate enough to have the necessary technical and human resources need to make them available to scientists at institutions less fortunate. Although some radiology departments have such equipment, the devices are usually committed mostly for patient care.

In addition to the images of brain activity, the experiments provide a vast amount of information. Such an accumulation not only yields answers to the questions posed at the time of the experiment but also provides invaluable information for future research, as those of us in the field have repeatedly discovered to our amazement and delight. Recent efforts to create neuroscience databases could organize and quickly disseminate such a repository of information.

Wise use of these powerful new tools and the data they produce can aid our understanding and care of people who have problems ranging from developmental learning disorders to language disabilities arising from, say, stroke. Researchers have begun to use functional brain imaging to learn about the mood disturbances that afflict patients with such mental illnesses as depression. The technology could guide neurosurgeons in the excision of brain tumors, enabling them to judge how the removal of tissue will hamper the patient. Centers across the world are investigating such other mental activities as attention, memory, perception, motor control and emotion. Clearly, we are headed toward a much richer grasp of the relation between the human mind and the brain.

FURTHER READING

INTRINSIC SIGNAL CHANGES ACCOMPANYING SENSORY STIMULATION: FUNCTIONAL BRAIN MAPPING WITH MAGNETIC RESONANCE IMAGING. S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S.-G. Kim, H. Merkle and K. Ugurbil in *Proceedings of the National Academy of Sciences,* Vol. 89, No. 13, pages 5951–5955; July 1, 1992.

SOMATOSENSORY CORTICAL PLASTICITY IN ADULT HUMANS REVEALED BY MAGNETOENCEPHALOGRAPHY. A. Mogilner et al. in *Proceedings of the National Academy of Sciences,* Vol. 90, No. 8, pages 3593–3597; April 15, 1993.

IMAGES OF MIND. M. I. Posner and M. E. Raichle. W. H. Freeman and Company, 1994.

PRACTICE-RELATED CHANGES IN HUMAN FUNCTIONAL ANATOMY DURING NONMOTOR LEARNING. M. E. Raichle, J. A. Fiez, T. O. Videen, A.-M. K. MacLeod, J. V. Pardo, P. T. Fox and S. E. Petersen in *Cerebral Cortex,* Vol. 4, No. 1, pages 8–26; January/February 1994.

Chemistry and Physics in the Kitchen

Bon appétit! *Scientists are beginning to understand how chefs accomplish their culinary masterpieces—and are making modest recipe suggestions of their own*

by Nicholas Kurti and Hervé This-Benckhard

Interest in the application of science to the art of cookery is growing. Cooks once regarded the introduction of scientific reasoning, let alone laboratory techniques, into their kitchens with suspicion, even with hostility. That time seems to be past. Nevertheless, both in restaurants and in domestic kitchens, many cooks tend to remain faithful to the grand culinary traditions and practices they were taught, without knowing why (or really even whether) those practices guarantee the best results. Thus, cooks add pinches of flour when heating custards to prevent them from curdling; they rigidly follow certain protocols in making soufflés; they generally do not vary the proportions of ingredients in their recipes, and so on. Perhaps for that reason, culinary superstitions and old wives' tales continue to flourish.

The mistrust of scientific explanations for culinary mysteries is all the more surprising given that music, painting, sculpture and the performing arts have prospered with experimental scrutiny and discovery. Science has improved the technologies for preserving, reproducing and disseminating works of art, which has led to a greater appreciation of those works by a wider audience. There is no proof that science and technology have compromised creativity in any way; they may even have helped it.

We believe it is the duty of scientists to acquaint culinary artists with principles and techniques that may stimulate their imagination, just as they have previously done for painters, composers and musicians. The time seems ripe for such an approach. Physics is beginning to explore the state of emulsions, suspensions, solid dispersions and foams—"soft matter," as physics Nobelist Pierre-Gilles de Gennes has called them—that often occur in cooking. Advanced structural chemistry can now elucidate the behavior of large molecules such as complex carbohydrates and proteins. New chromatographic methods make it possible to isolate the components of foods that give rise to tastes and smells. Scientific explanations are already appearing for many old and seemingly obscure culinary tricks.

In effect, a new discipline is being born: molecular and physical gastronomy, the science of food and its enjoyment. We would like to offer a small feast of discoveries from this field—concerning appetizers, main courses, desserts and beverages—that may be of practical value and interest to cooks. Several of our examples are drawn from discussions at the First International Workshop on Molecular and Physical Gastronomy, which we organized in August 1992 in Erice, Sicily.

A popular first course, *oeuf dur mayonnaise* (hard-boiled egg with mayonnaise), gives us the opportunity to examine the molecular and physical properties of emulsions. Mayonnaise, cream, butter and béarnaise sauce are all emulsions, in which droplets of one liquid are suspended in another with which it is immiscible.

Mayonnaise consists of vegetable oil, vinegar or lemon juice, and egg yolk. Because half the yolk is water, mayonnaise is actually an emulsion of oil in water. Ordinarily, no matter how thoroughly you whisk a mixture of water and oil, the two components separate into distinct layers. Mayonnaise is stable because the egg yolk contains so-called surface-active molecules such as lecithins. The two ends of these rod-shaped molecules have different properties: one is hydrophilic (it has an affinity for water), and the other is hydrophobic (it shuns water). Each oil globule in the mayonnaise

NICHOLAS KURTI and HERVÉ THIS-BENCKHARD have collaborated in the new field of molecular and physical gastronomy since 1989. Kurti, emeritus professor of physics at the University of Oxford and a former vice president of the Royal Society, was born in 1908 and educated at the University of Paris and the University of Berlin. During his long career, he has achieved many honors, including election as an honorary member of the American Academy of Arts and Sciences. His many papers and books cover topics in cryophysics, magnetism, energy and, more recently, culinary physics. This-Benckhard is the co-editor of *Pour la Science,* the French edition of *Scientific American.* He has written extensively about chemistry, mathematics and culinary science. He holds a degree in physical chemistry from the École Supérieure de Physique et de Chimie Industrielles de Paris and has studied modern literature at the University of Paris.

INSTANT ICE CREAM, produced by the vaporization of liquid nitrogen, is one of the culinary delights being created by a hitherto unused technique. The joint efforts of scientists and cooks are explaining why certain cooking practices work and, in some cases, how they can be improved.

suspension is coated by at least one layer of lecithin, which allows it to mix freely with the aqueous medium because the hydrophilic part contacts the water while the hydrophobic part faces the oil. The droplets do not readily coalesce, because the hydrophilic ends protruding from their surfaces usually acquire small electric charges and consequently repel one another.

Cooks generally accept that one egg yolk can emulsify only 150 to 250 milliliters of oil. Yet that ratio is a gross underestimate, as Harold J. McGee, a writer on science and cooking, has revealed. Knowing that a yolk contains about two grams of lecithin and assuming that the oil droplets in a mayonnaise are one hundredth of a millimeter in diameter, he calculated that one yolk could stabilize 3.5 liters of

oil—about 20 times the recommended quantity. The trick is to keep adding enough water, wine, vinegar or other aqueous solutions to keep the oil droplets separate.

McGee's work shows how science can extend the limits of a culinary process. We do not mean to suggest that one should abandon the traditional rule in favor of "one yolk per 3.5 liters," but the knowledge that it can be disregarded might sometimes prove useful. Suppose that a dinner guest desires an *oeuf dur mayonnaise,* but you can find only a single egg and no mayonnaise in the larder. With a hypodermic syringe, withdraw one milliliter of yolk from the egg, use that to make a tablespoon of mayonnaise and hard-boil the rest of the egg.

Some concerns have been raised, particularly in the U.S. and the U.K., about eating eggs contaminated by salmonella bacteria. In November 1988, for example, a junior minister at the British Department of Health made the exaggerated announcement that most eggs produced in Great Britain were infested with salmonella. Within a couple of weeks, egg consumption had dropped by one half, and several food writers were depicting a gloomy future devoid of soft-boiled eggs or fluffy omelettes.

One of us (Kurti), wondering whether scientific reasoning could avert this calamity, looked more closely at the problem of soft-boiled eggs. (We should note that the British soft-boiled egg has a soft but coagulated white and a creamy yolk, unlike the French three-minute *oeuf coque,* which is more liquid throughout.) He first ascertained that egg yolk coagulates between about 62 and 65 degrees Celsius and that salmonella organisms cannot survive more than a few minutes' exposure to 59 degrees C. A safe cooking method should therefore ensure that the temperature of the yolk never exceeds 62 degrees C but still remains between 59 and 61 degrees C for at least six minutes.

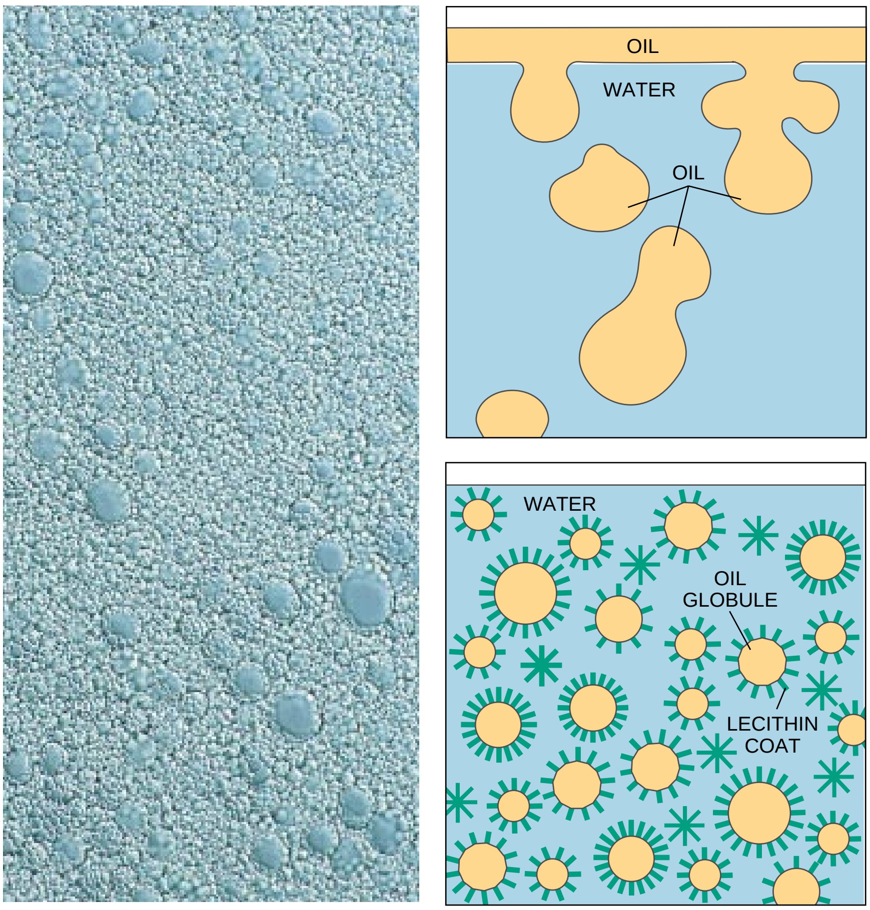
The temperature at the center of the yolk of an intact egg can be measured with a thermocouple, an electrical sensor whose output voltage varies with temperature. A fine thermocouple wire can be threaded through a hypodermic needle and anchored to its tip; the thermocouple can then be connected to a microvoltmeter calibrated in degrees C.

That instrument has made it possible to develop a recipe for turning a contaminated egg into a safe soft-boiled egg. First, place a 60-gram egg into boiling water for 3.5 minutes. As Richard Gardner and Rosa Beddington of the University of Oxford have shown, that immersion will cook the egg white to the proper consistency but will raise the yolk temperature to only about 30 degrees C. Second, immediately transfer the egg into a water bath at 60 degrees C. The yolk temperature will then gradually rise to 59 degrees C in another 7.5 minutes. Thomas J. Humphrey of the Exeter Public Health Laboratory has shown that eggs inoculated with one million *Salmonella enteritidis* organisms were rendered safe by this treatment after a total of 18 minutes of cooking.

Can any treatment make raw eggs safe for use in mayonnaise and other dishes? Yes, because the temperature that kills the salmonella is below that of yolk coagulation. Put the yolks in a bowl, then set the bowl in a water bath at 62 degrees C for about 15 minutes, giving the yolks an occasional stir. Alternatively, mayonnaise can also be made from hard-boiled egg yolks diluted with some vinegar (the coagulation of the egg does not destroy the lecithins). In fact, most French cookbooks recommend hard-boiled egg yolks for preparing *mayonnaise tartare,* which contains onion, capers and herbs in addition to the usual flavorings of vinegar, mustard, salt and pepper.

Despite such advances, some mysteries about cooking with eggs remain. For example, in the preparation of many foods—custard, zabaglione (egg punch) and various savory sauces thickened with egg yolks—the yolks must be heated in the presence of another liquid. Very commonly, however, these mixtures will curdle over the heat. As chefs have long known, a pinch of flour can prevent this problem. In Erice, researchers discussed the mechanism of the chefs' solution.

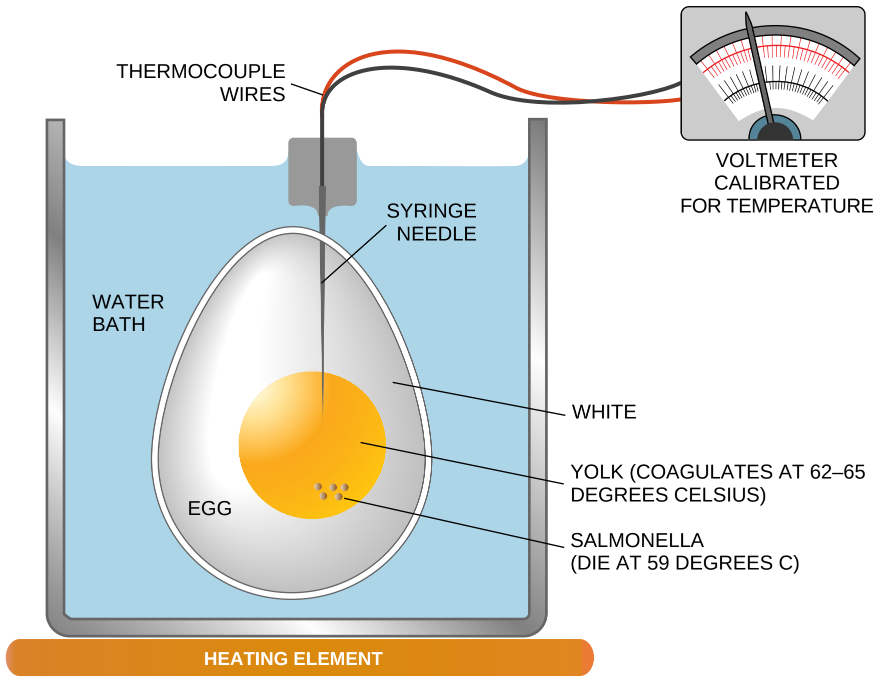
The curdling occurs because, in a watery solution, long protein molecules from the egg yolk are freed from some of the weak bonding forces (such as hydrogen bonds and sulfur bridges) that



MAYONNAISE (*left, magnified 100 times*) is fundamentally a suspension of oil in water. Ordinarily, mixtures of oil and water separate into distinct layers (*top right*). Mayonnaise is more stable, however, because of lecithin in the added egg yolk. Because lecithins have one hydrophilic end and one hydrophobic end, coatings of lecithin can surround the oil globules and allow them to disperse in the aqueous solution (*bottom right*). The average size and size distribution of the oil drops depend on the proportion of the oil and the amount of whipping energy: as the amount of oil and whipping increases, the size decreases, and the range of diameters narrows.

held them in specific coiled conformations. Sufficient heating makes the proteins denature, or uncoil, and then form new weak bonds with other molecules. When the proteins meet and aggregate into clumps, they form curds.

Flour can stop this clumping because its particles consist mainly of two types of starch: amylose, a linear polymer of the sugar glucose, and amylopectin, a highly branched polymer of the same sugar. At high temperatures, these starches fall away from the flour granule and dissolve. The long starch molecules seem to inhibit curdling by limiting the movements of the proteins and stopping them from aggregating. Experimental studies should confirm or refute this explanation. They should also help quantify how much flour is needed to prevent curdling and what kind of starch works best. The number of experiments that just this one phenomenon suggests underscores the huge amount of work needed to understand fully even simple cooking facts.



TEMPERATURE INSIDE an egg being cooked can be measured using an electrical sensor called a thermocouple and a properly calibrated microvoltmeter. Experiments using this equipment have made it possible to devise a recipe for soft-boiling an egg in such a way that any salmonella bacteria would be killed.

Let us now consider some more complicated cooking: the soufflé. This dish is a pleasure to eat, but its preparation can be like a walk on a tightrope. The ideal soufflé expands greatly during cooking and has a crisp outside encasing a fluffy, slightly creamy interior. The basic ingredients are beaten egg whites, a viscous preparation such as béchamel sauce (a cooked mixture of butter, flour and milk) and—often—egg yolks. Fish, cheese, chocolate and fruit purée are also sometimes added and become important parts of the soufflé structure; vanilla, liqueurs and other ingredients can be used purely as flavorings. Recipes vary greatly in the proportions of all these ingredients. But all cookbooks agree on the importance of mixing the viscous preparation and the beaten egg whites uniformly; they also emphasize that when cooks are dispersing the foam in the béchamel, they must take care not to break the myriad air bubbles.

Some chefs believe that after the soufflé mixture is ladled into a soufflé dish, it should go immediately into a preheated oven. Others say the mixture can be safely kept for an hour or so at kitchen temperature or in a water bath of 40 degrees C for up to 30 minutes. According to some reports, small, individual soufflés may be deep-frozen and defrosted before cooking. We recently experimented with all four of those techniques, using small, individual soufflés [see "The Kitchen as a Lab," "Amateur Scientist," page 120]. All of them gave acceptable results, but the soufflé placed in the oven right after the whisked egg whites were folded in came out best. We presume that the other methods allowed bubbles in the mixture to coalesce and to escape from the mixture.

Whatever its method of preparation, a soufflé must be served immediately because, if it has risen noticeably, it will begin to collapse within seconds or minutes after it leaves the oven. Although some details of a soufflé's rise and fall are shrouded in mystery, the general explanation is relatively simple. Some observers have suggested that when the viscous mass of flour and egg yolks is heated, the air bubbles expand and raise the soufflé. The coagulation of the eggs then makes the material between the bubbles rigid enough to prevent the soufflé's collapse—at least until the temperature drops. Yet the heat-driven expansion of the air could account for only about a 20 percent increase in volume, whereas a soufflé can rise to more than three times its original size. It is actually water vapor that inflates the soufflé, as can be readily demonstrated: cut open a cooked soufflé, and water vapor escapes.

The first measurements of the changing temperature inside a soufflé being cooked were made 25 years ago. A thermocouple enclosed in a hypodermic needle was anchored to the dish, its tip 20 millimeters below the surface of the soufflé at the beginning of the experiment. That work showed that during the first 10 minutes, the temperature inside a soufflé reaches about 45 degrees C. It then sometimes levels out (perhaps because of protein coagulation) and may even dip as the thermocouple touches cooler, uncooked parts of the mixture. After a further 25 minutes or so (the time depends on the size of the soufflé), the temperature again rises rapidly as the water quickly evaporates from the top. The rapid rise indicates that the soufflé is done. Amateur cooks who are not averse to the introduction of thermocouples into the kitchen may find the remote sensing of the soufflé's progress a convenient aid.

So far in our discussion we have mainly considered ways to understand various cooking processes rather than to improve their results. But there are cases in which even small modifications of traditional methods might produce changes in flavors. The concentration of a bouillon for the preparation of a *fond* is a good example.

A bouillon is made by boiling meat and vegetables in water. Aside from their value as foods themselves, bouillons also form the *fond,* or base, of many sauces. For that use, bouillons are boiled to reduce, or concentrate, them to one tenth or one twentieth of their original volume. But when you boil a mixture of chemical compounds, the composition of the vapor usually differs somewhat from that of the liquid. For instance, when you boil wine, a mixture of alcohol and water, more of the alcohol evaporates first. The difference between the vapor and the liquid depends on the heating temperature. Would a *fond* prepared by reducing a bouillon at 100 degrees C be different from one prepared at 80 or 60 degrees C?

At the Erice workshop, participants were shown a makeshift apparatus that demonstrated how a bouillon could be reduced at a lower temperature by lowering the air pressure. A glass jar containing bouillon was connected to a filter pump, and its internal pressure was lowered to half an atmosphere (or about seven pounds per square inch). At that pressure, the bouillon boiled away at a temperature of only 80 degrees C. A better apparatus is now being assembled to establish whether low-pressure cooking could have gastronomic advantages.

Before we proceed to our main course—a cooked meat—we must say a few words about heating. Traditional cooking embraces two basic methods of bringing heat to foods. One is to expose the material to a hot liquid (as in boiling, stewing, frying and sautéing) or to a hot gas (as in oven roasting and baking). When the molecules of the heating medium strike the surface of the food, they transfer their kinetic energy to it. In the second method, represented by grilling, electromagnetic radiation strikes the food and is converted into heat. In both cases, the heat reaches the inside of the foodstuff only through convection and conduction. Because different parts of the food being cooked vary in their exposure to the heat, the taste and texture are not uniform. Our enjoyment of a rare steak, a crisp bread roll or a fluffy omelette is greatly influenced by such temperature gradients and discontinuities in texture and composition.

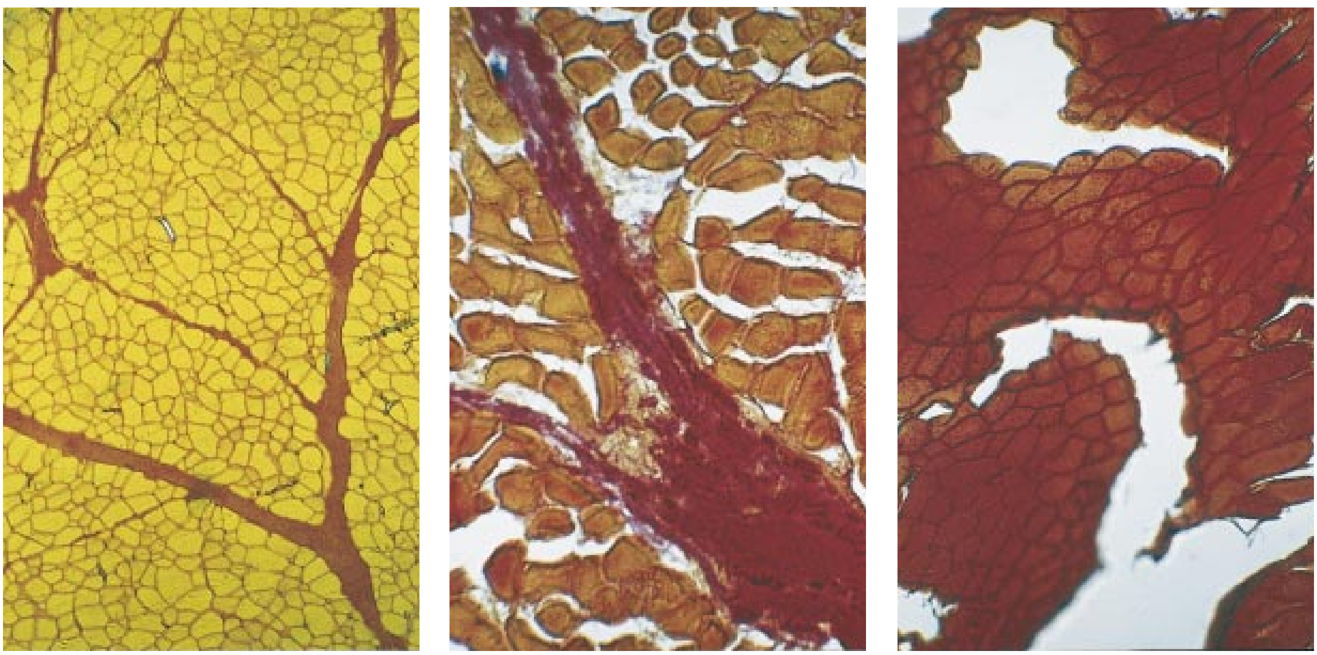
The possibility of heating the inside of a material without first heating the outside arose during World War II through a chance observation that microwaves of about a 10-centimeter wavelength could pass through considerable thicknesses of materials while giving up some energy as heat. The effect depends on the presence of polar molecules such as water, which are electrically neutral but carry asymmetrically arranged charges. Microwaves can make polar molecules rotate or oscillate; friction within the material converts that kinetic energy into heat. Ice does not absorb microwaves, because its water molecules are locked into crystals and unable to rotate. Thus, using microwaves, one can boil water inside an ice block or create an inverted Baked Alaska—the so-called Frozen Florida, which has a hot interior and a frozen shell.

When meat is prepared in a microwave oven, it uniformly warms to 100 degrees C and stays at that temperature for as long as it still contains water. This microwave method of cooking meat has two advantages: it is faster and more energy efficient. On the other hand, conventional boiling in a bouillon containing herbs and vegetables can lend additional flavors to meats.

Roasted meat is more flavorful than boiled meat because of browning reactions that intensify above 100 degrees C. Sugars and amino acids in the meat can then cross-link and create many kinds of compounds, some flavorful, some dark brown in color. These Maillard reactions, as they are known, produce the crackling on roasted meats. Some types of browning reactions—such as caramelizing sugar—are easy to induce during microwave cooking, but those are not Maillard reactions.

For the tastiest results, a chef may wish to combine traditional grilling and microwave cookery. A good example is roast duck Pravaz-Cointreau, named in part for the French physician Charles Gabriel Pravaz, one of the inventors of the hypodermic syringe. Pieces of a jointed duck are first grilled or fried to brown them, then injected with the orange liqueur Cointreau (a good absorber of microwaves because of its high water content). They are then placed in a microwave oven to cook their insides, which takes only a few minutes. By such a method, the meat is boiled from the inside in an orange medium; the dish looks like a modernized version of *canard à l'orange.*

No good dinner is complete without dessert. From the world of physics comes a recipe that not only eases the task of a chef but also produces a magnificent spectacle. This dessert, instant



COOKING OF MEAT produces a variety of physical and chemical changes in it. In its fresh, uncooked state (*left*), the meat fibers (*red*) are separated by collagen (*yellow*). After a short period of heating (*center*), the collagen begins to break down. Thoroughly cooked meat (*right*) is tender because almost all the collagen has dissociated into gelatin.

ice cream, was devised by Peter Barham of the University of Bristol. As a suitable finish for a public lecture on ice cream, Barham developed a way to make enough of it in about two minutes to feed an entire audience. The same recipe can be adapted to a domestic scale.

Good ice cream contains abundant air bubbles (to keep it light) and only very small ice crystals (so that the texture is smooth). Traditionally, ice cream makers have churned the mixture of milk, eggs, sugar and flavorings as it slowly chilled: the churning folded air into the material while also continuously breaking up large ice crystals. A simpler and more efficient way is to pour liquid nitrogen directly into the ingredients. At a temperature of –196 degrees C, liquid nitrogen can freeze the ice cream mixture so fast that only small ice crystals have time to grow. As it furiously boils, the liquid nitrogen also creates plenty of small gas bubbles. And as a further delight, the cold produces a cloud of dense fog, thus adding a crowd-pleasing, highly dramatic touch.

ROAST DUCK PRAVAZ-COINTREAU is prepared using both a conventional and a microwave oven. The conventional oven induces tasty browning reactions in the skin. Microwaves acting on liqueur injected into the duck then quickly cook the interior in an orange-flavored medium.

You will need about equal volumes of liquid nitrogen and a mixture for ice cream or sorbet. After preparing the mixture in the usual way, place it in a large metal bowl. (Do not use a glass or plastic bowl, which might break from thermal shock.) While observing the proper safety precautions (as set out below), pour in about half the liquid nitrogen, stirring gently with a wooden spoon. Continue to stir while adding more of the coolant until the ice cream is nice and stiff. Make sure the ice cream has stopped giving off fog—which signals that all the nitrogen has evaporated—before serving.

Two important safety points need to be made. First, always wear gloves and safety glasses when handling the liquid gas or any objects that have been exposed to its extreme cold. Second, if you are making the ice cream in front of guests, be sure they are out of range of any splashes. You should be able to obtain liquid nitrogen (or directions to a commercial source for it) from your local university's physics or chemistry department or from a hospital. The best way to transport liquid nitrogen is with a vacuum flask; inside a well-made one, it will last for up to a day.

Would you like a postprandial drink? Much research has been done in the past decade on the chemistry and biochemistry of wine and spirit production. Oenological research institutes have studied the composition of some of the great wines and counted in some cases more than 500 compounds that may contribute to a wine's character.

Vintners routinely age many spirits and some wines in oak barrels because chemical reactions with the wood improve the liquors' flavor. Wood contains many complex chemical compounds, among them cellulose, hemicellulose, lignin, tannins (chemicals that are often astringent) and resinous molecules. Oak is the preferred material for those barrels because it is strong and waterproof and contains none of the resins that give retsina wines their flavor.

In the late 1970s and 1980s Jean-Louis Puech of the National Agronomic Research Institute in Montpellier demonstrated how the ethyl alcohol in a spirit extracts tannins and lignin from its wooden container. Through the simple but lengthy experiment of putting alcohol in an oak barrel for more than 10 years, and also by macerating pieces of wood in alcohol, he studied how both the liquid and the wood changed over time. The concentration of tannins in the wood decreased by 75 percent. Moreover, the extracted tannins had been oxidized into a variety of flavorful compounds. The concentration of lignin on the inside of the barrel was 5 percent lower than on the outside. The cellulose content was almost unchanged, but the hemicellulose had been modified: it had dissociated into sugars, such as fructose, xylose, arabinose and glucose.

For the molecular gastronome, perhaps the most significant discovery was that vanillin, the major aromatic molecule of vanilla, was a final product of lignin degradation during aging. Indeed, the vanilla flavor can be detected in old cognacs, rums and whiskeys. The manufacturers of wines and spirits are typically forbidden by law to improve the taste of their products by adding sugar or other chemicals. Nevertheless, if the consumer wants to use the results of chemical research to enhance the qualities of inferior wines or spirits, should he or she not be encouraged to do so? A few drops of vanilla extract may wonderfully enrich the flavor of a bottle of cheap whiskey.

This kind of experiment can be extended to a large number of beverages and dishes. Perhaps in the cookbooks of the future, recipes will include such directions as "add to your bouillon two drops of a 0.001 percent solution of benzylmercaptan in pure alcohol."

Science can explain, analyze and help in the creation of new dishes. But even though we are convinced that science has an important role in gastronomy, we also firmly believe the scientist will never dethrone the chef. The great culinary creations will be, as they have always been, the result of artistic imagination seasoned with a blend of empiricism and tradition and only a soupçon of science. These sentiments lead us to hope that Comus, the patron spirit of the culinary artists, will join the Muses in accepting science as an ally in the practice of the arts.

FURTHER READING

THE PHYSICIST IN THE KITCHEN. Nicholas Kurti. *Proceedings of the Royal Institution of Great Britain,* Vol. 42, Part 6, No. 199, pages 451–467; 1969.

ON FOOD AND COOKING: THE SCIENCE AND LORE OF THE KITCHEN. Harold McGee. Macmillan, 1988.

BUT THE CRACKLING IS SUPERB. Edited by Nicholas Kurti and Giana Kurti. Institute of Physics Publications, 1988.

THE CURIOUS COOK: MORE KITCHEN SCIENCE AND LORE. Harold McGee. North Point Press, 1990.

LES SECRETS DE LA CASSEROLE. Hervé This. Belin Publishing, Paris, 1993.

The Dilemmas of Prostate Cancer

Do the risks of aggressive treatment for early prostate cancer outweigh the benefits? This question is one of several unresolved issues faced by those who treat, and those who have, prostate cancer

by Marc B. Garnick

This year cancer of the prostate gland is expected to be diagnosed in 200,000 men in the U.S. and to take the lives of 38,000 American males. Such numbers make prostate cancer the most frequently diagnosed malignancy (other than that of the skin) in American males and the second leading cause of cancer-related deaths (behind lung cancer) in that group. Recently musician Frank Zappa, theater producer Joseph Papp and actors Telly Savalas and Bill Bixby all lost their battles with the disease. It does not kill gently. In their final months, most people who succumb to it endure excruciating pain that is difficult to control.

Sadly, prostate cancer, which becomes increasingly common with advancing age and is typically diagnosed in men older than 65 years, has historically been the subject of relatively little research. Physicians therefore lack the information they need in order to decide on the best therapy for many patients. A patient who finds himself diagnosed with prostate cancer in 1994 is thus quite likely to discover that even leading experts disagree on the best course of action in his particular case. Indeed, doctors and policymakers are mired in controversy over how the disease should be managed, especially in its early stages, and over whether many seemingly curable cases should be treated at all. These and other issues relating to prostate cancer are in urgent need of resolution if deaths and suffering are to be reduced significantly.

MARC B. GARNICK is associate professor of medicine at the Dana-Farber Cancer Institute at Harvard Medical School, where he cares for patients with genitourinary cancer. He is also vice president of clinical development at Genetics Institute, Inc., a biotechnology firm in Cambridge, Mass. Garnick was a leader in the clinical testing of leuprolide for the treatment of metastatic prostate cancer and is now developing other drugs.

The ongoing arguments can be best understood if one has some knowledge of how prostate cancer is currently diagnosed and treated. Physicians usually detect cancers by finding a lump in the prostate gland, which is a walnut-shaped structure that helps to maintain the viability of sperm. Such lumps may be discovered during a routine checkup or an examination prompted by a patient's complaint of sudden urinary discomfort or, occasionally, impotence. To examine the gland, the physician inserts a finger into the rectum and, feeling the gland through the rectal wall, searches for abnormalities in its size, contour or consistency.

The symptoms that send men to their doctors often arise when a cancerous mass causes the prostate to press on nearby structures. The gland lies underneath the urinary bladder, and it surrounds the urethra—the tube through which urine passes from the bladder to the penis and outside the body. For instance, if the tumor pushes on the bladder or pinches the urethra, it can cause men to have to urinate at night, with unusual frequency or with great urgency; it can also lead to difficulty initiating or maintaining a urine stream.

In some instances, prostate cancer is detected in quite a different way—as a by-product of treatment for a disorder called benign prostatic hyperplasia. This condition, an aging-related enlargement of the prostate, affects more than half of all men older than 45 and may give rise (albeit more gradually) to the same urinary troubles caused by a tumor. If the symptoms become too distressing, a surgeon may attempt to relieve them by performing transurethral resection of the prostate, a procedure in which parts of the gland are scraped away. Whenever resection is done, the excised tissue is analyzed under a microscope for evidence of malignancy, which is occasionally found.

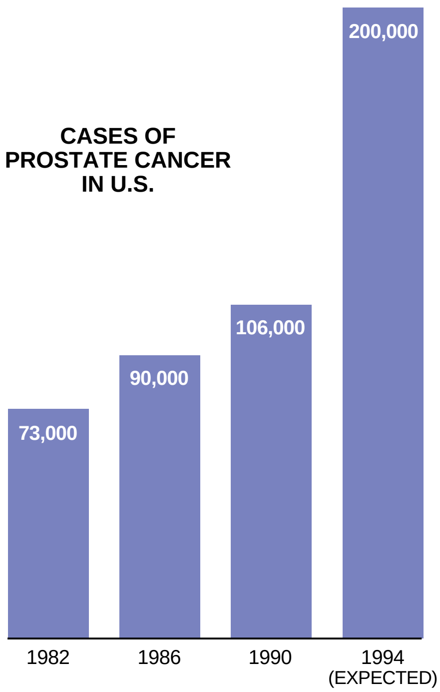
A simple blood test constitutes a third means of detecting prostate cancer, and it can signal the presence of cancer in individuals who display no symptoms of prostate abnormalities. This test measures the level of a glycoprotein called prostate-specific antigen (PSA), one of many molecules secreted by the prostate gland. In most versions of the test, which became widely available in 1986, levels that exceed four nanograms in a milliliter of blood suggest cancer might be present; levels greater than 10 are especially suggestive. Most of the tumors detected by the PSA test are still microscopic.

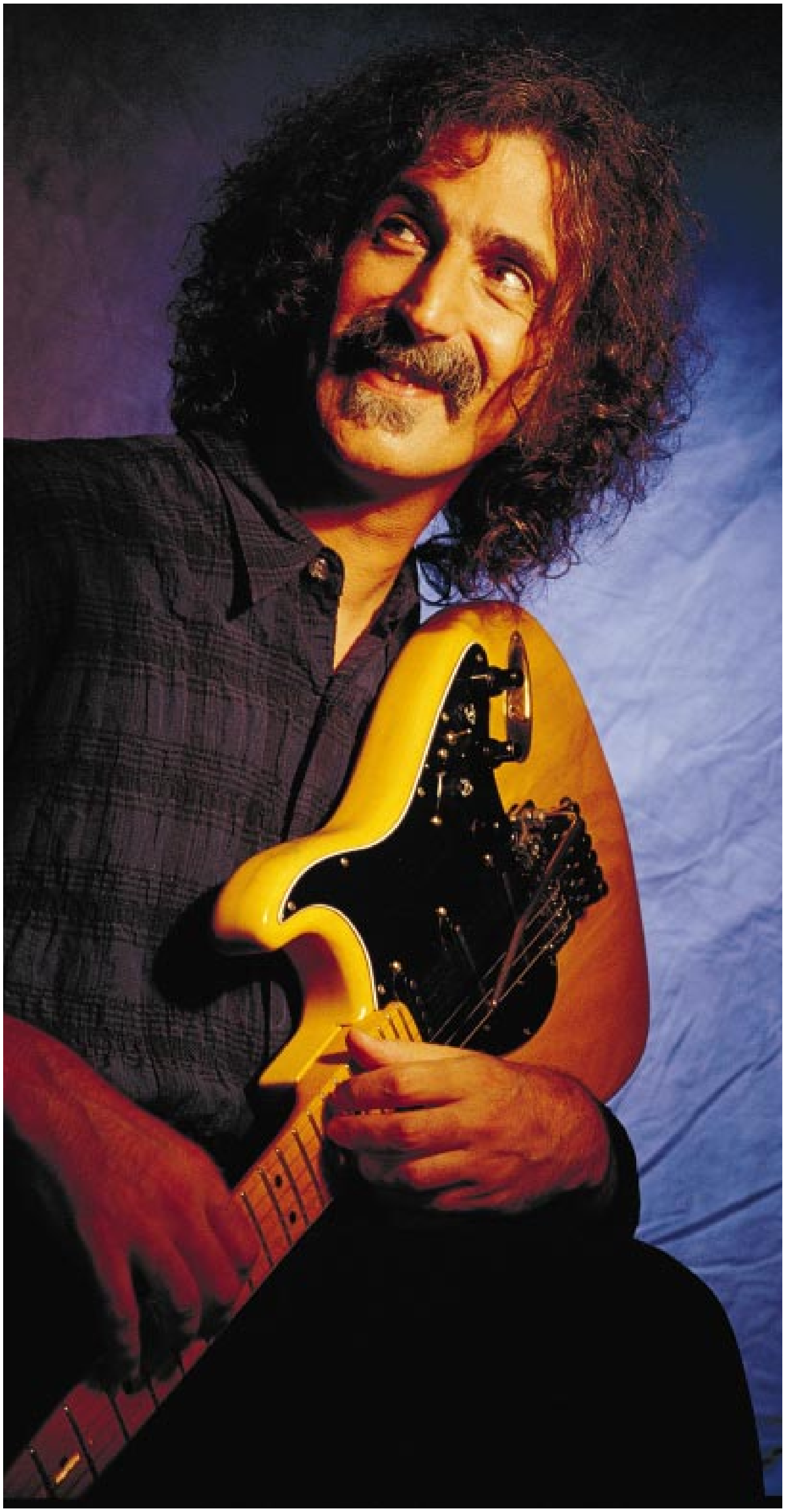
An elevated PSA reading is by no means proof that cancer is present, however. Factors other than cancer—such as development of benign prostatic hyperplasia, inflammation of the prostate (prostatitis) and mechanical pressure on the gland—can cause the level to rise. Conversely, in many men who have cancer, the PSA level is normal at the time of diagnosis. Investigators are currently exploring ways to overcome these drawbacks.

Regardless of the test's shortcomings, its ease and relatively low cost ($30 to $80) have made it a popular tool for detection of cancer in asymptomatic men. It is so popular, in fact, that it probably accounts for much of the recent, striking increase in the number of men who are diagnosed with prostate cancer every year. The number of cases identified in 1994 will be more than double the 90,000 cases identified in 1986.

As will be seen, the rise in cases has increased the intensity of the debate over whether patients with small cancers should be sought out and given aggressive treatment shortly after diagnosis in an attempt to effect a cure. Controversy has arisen in part because no one can yet distinguish conclusively among microscopic cancers that will remain latent (causing no symptoms in the patient's lifetime) and those that will become clinically significant (growing enough to cause symptoms or become life-threatening). Doctors can make educated guesses about virulence by examining such characteristics of tumors as their size and microscopic appearance. But they cannot determine with certainty which early tumors need treatment and which do not.

Improvement in the ability to identify small tumors in men has sparked arguments because it may lead to delivery of risky treatment to huge numbers of men who would have died with, not of, prostate cancer. Autopsy studies of males who died from other causes indicate that about a third of men older than 50 have at least some cancerous cells in their prostate and that the incidence increases steadily after age 50 (so that 90 percent of men older than 90 years are affected). Yet most men who acquire prostate cancer do not die from it. (About 3 percent of American men are expected to die eventually from the disease.) Some experts favor minimizing the number of men who re-





MUSICIAN FRANK ZAPPA died last year from prostate cancer at age 52. Many physicians in the U.S. advocate routine screening for this cancer in men older than 40 or 50 years, assuming that detection of early cancers will lead to more cures and help prevent thousands of deaths every year. Opponents worry that increased screening—thought to account for much of the sharp rise in diagnosis seen since the mid-1980s (*graph*)—will be counterproductive. A blood test now widely used for screening often leads to discovery of microscopic cancers, many of which will not become lethal. Hence, skeptics argue, increased screening may lead to risky therapy in multitudes of men who would have been better off without it.

ceive unnecessary treatment; others worry that lack of screening and treatment will cause thousands who might have been saved each year to suffer a cruel death.

Of course, no one is treated on the basis of a PSA test or discovery of a lump alone. Detection of a possible cancer is usually only the first step in diagnosis. When a rectal examination or an elevated PSA level reveals that a cancer might be present, physicians generally follow up with an ultrasound examination. This procedure can often pinpoint the location of a tumor and can aid in the next step of diagnosis: biopsy of the prostate and study of the suspicious tissue under a microscope.

If the microscopic analysis confirms the presence of a malignancy, physicians try to determine its stage of advancement. They do so because current guidelines for treatment are based on the extent of progression. Classification systems vary, but a common one divides tumors into four lettered stages. Stages A, B and C include cancers that have not metastasized; they have not spawned new tumor colonies in other tissues. Stage D consists of tumors that have already metastasized. Usually prostate cancer spreads first to the lymph nodes that are immediately downstream from the prostate gland, then it appears in the bones and other organs.

The first three stages are distinguished from one another by the size of the tumors. Stage A malignancies are microscopic; they are the type that have traditionally been revealed by transurethral resection. Such cancers can be divided into two subclasses. Stage A1 tumors are confined to one small area of the prostate and are composed of relatively well differentiated tissue—that is, despite some obvious abnormalities in the cancerous cells (such as enlarged nuclei), they, like healthy gland cells, are of uniform size and closely packed. Stage A2 cancers are more diffuse, consist of moderately to poorly differentiated tissue, or display both characteristics. Multiple tumor sites in the prostate gland or poor differentiation implies that the cancer is likely to behave aggressively.

Stage B malignancies are palpable (large enough to be felt as a nodule during a rectal exam) but rarely cause discomfort. Stage C tumors have spread through most or all of the gland, making it rock-hard, and have typically pushed past the borders of the prostate into surrounding structures. Patients bearing stage C malignancies are often diagnosed after urinary symptoms cause them to seek medical help.

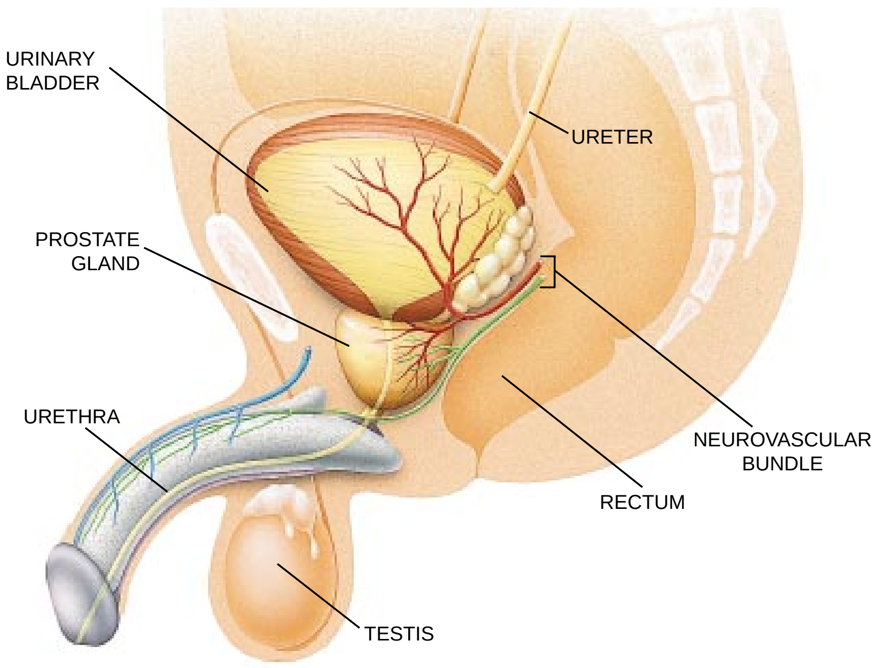
To "stage" tumors, physicians first combine the information gleaned from the rectal exam, the ultrasound and the biopsy with information provided by other, noninvasive tests. For instance, they may do a computed tomographic scan of the abdomen and the pelvis, searching for evidence of cancer in the lymph nodes. Experience teaches us that cancer in these nodes often signals the presence of cancer elsewhere in the body. Doctors may also perform a specialized kind of scan to look for metastatic deposits in the bones. With these clinical findings in hand, they assign the tumor to a tentative stage.

Regrettably, the diagnostic tests often cannot end there, because 25 to 50 percent of tumors that are assigned initially to stage A2, B or C turn out, on further testing, to be metastatic, stage D cancers. (The rate for stage A1 tumors is undoubtedly much lower.) Discovery of metastasis is important because metastatic malignancies call for a therapeutic approach different from those applied to less advanced tumors. To ascertain the stage more accurately, physicians may biopsy or remove lymph nodes from the pelvis, so that nodal tissue can be analyzed directly for evidence of cancer. Unfortunately, this surgical-staging technique cannot detect stray cancerous cells that have escaped into the blood and lodged in the bones, and so some patients who are treated as if they had early disease will actually have metastatic cancer.

Today standard protocols in the U.S. recommend that most patients with stage A or B disease (except perhaps for older individuals with stage A1 disease) be treated promptly with one of two potentially curative therapies. These treatments, which are thought to be about equally effective, involve removing the prostate gland (a radical prostatectomy) or irradiating the gland to kill the cancerous cells within it. Radiation is often preferred for men who are too frail to withstand surgery.

Physicians are further taught that radiation is the treatment of choice for stage C disease, because surgery cannot fully eradicate tumors that have pushed their way past the borders of the prostate gland. Textbooks also note that neither surgery nor radiation is likely to cure metastatic, stage D disease. Individuals with such advanced cancer are therefore better served by systemic therapy aimed at slowing the progression of metastatic deposits and at easing pain and other symptoms.

For the past 50 years, doctors have attempted to inhibit progression of advanced cancer by initiating hormonal therapies. This approach is based on the Nobel Prize–winning discovery, by Charles B. Huggins of the University of Chicago, that male hormones (androgens) can markedly accelerate the growth of prostate cancer and that



PROSTATE GLAND is a walnut-shaped structure (*center*) that lies below the urinary bladder and surrounds the urethra—the tube that carries urine from the bladder, through the penis and out of the body. Many men are unaware of the gland's existence until cancer or an aging-related enlargement of the gland leads to urinary abnormalities or, in some cases, to erectile difficulties.

Typical Therapies for Prostate Cancer

Treatment for prostate cancer is based on the stage to which the disease has advanced. The standard recommendations in the U.S. are listed in the rightmost column. Use of aggressive therapy—radiation or radical prostatectomy (surgery to remove the prostate gland)—for stage A and B disease has become increasingly controversial because some data suggest nontreatment may result in a comparable survival rate or in a better quality of life for certain patients. Those findings, however, are inconclusive and remain a subject of debate.

STAGE OF DISEASE		STANDARD THERAPY
STAGE A MICROSCOPIC CANCER WITHIN PROSTATE GLAND	**A1** Cancer is confined to one site and is well differentiated	Observation, radiation or radical prostatectomy
	A2 Cancer occurs in many sites or is moderately to poorly differentiated	Radiation or radical prostatectomy
STAGE B PALPABLE LUMP WITHIN PROSTATE GLAND	**B1** Cancer forms a small, discrete nodule in one lobe of gland **B2** Cancer forms a large nodule or multiple nodules, or it involves both lobes or is moderately to poorly differentiated	Radiation or radical prostatectomy
STAGE C LARGE MASS INVOLVING ALL OR MOST OF PROSTATE GLAND	**C1** Cancer occurs as a continuous mass that may have extended somewhat beyond the gland **C2** Larger cancer occurs as a continuous mass that has invaded structures surrounding the gland	Radiation; some physicians administer hormonal therapy with radiation
STAGE D METASTATIC TUMOR	**D1** Cancer appears in the lymph nodes of the pelvis **D2** Cancer involves tissues beyond the lymph nodes, usually including the bones	Hormonal therapy once symptoms arise (or possibly as soon as metastatic deposits are found) and palliative therapy for pain and other discomforts

withdrawal of such hormones can retard its growth. Androgen levels in the body can be reduced by removing the testes (a bilateral orchiectomy), where 95 percent of testosterone—the main male hormone—is made. They can also be lowered by various drugs, such as estrogen, that interfere with the actions or synthesis of androgens. Unfortunately, almost all metastatic tumors become resistant to hormonal therapy at some point and then progress rapidly. Usually patients die within two to five years after metastases are discovered.

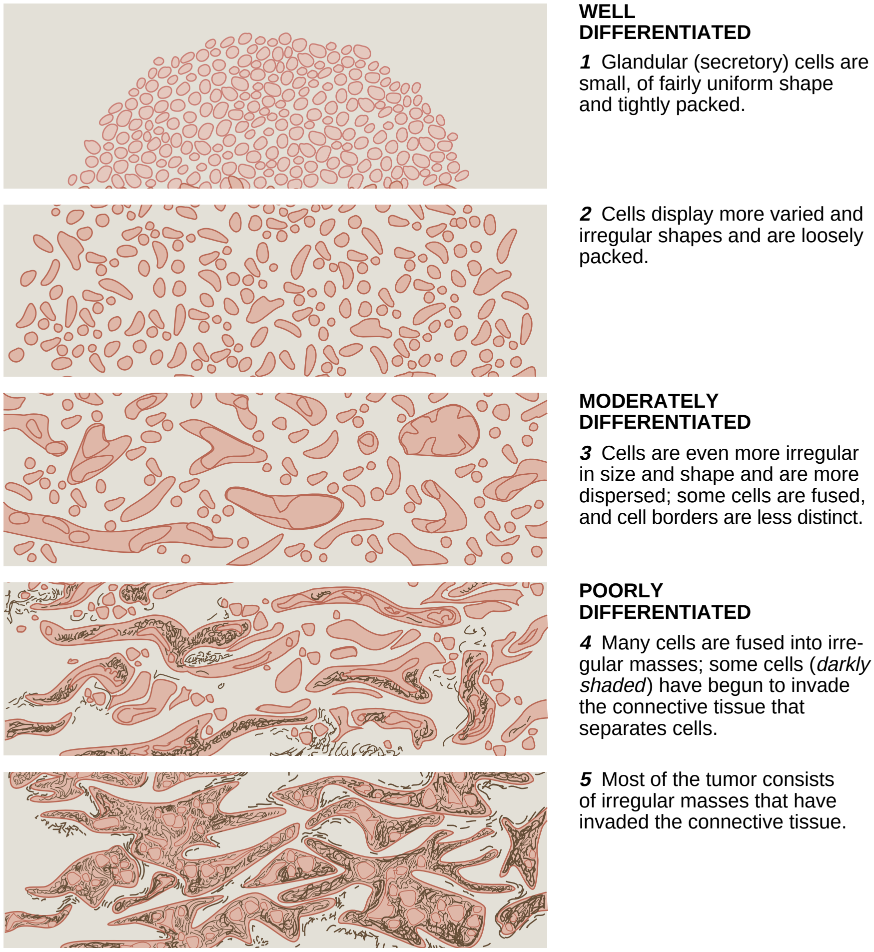
Most physicians are still in basic agreement on the merit of the standard strategies for coping with stage C and D disease. In contrast, experts increasingly disagree over the best approach to stage A and B disease and over the advisability of mass screening for early cancers. The conflict partly stems from the fact that the treatments are risky [*see box on page 78*]. Published rates of complications vary, but both radiation and surgery can lead to considerable impotence and to incontinence, bowel injury and, rarely, death. (Older patients are more prone to complications than are younger ones.)

Complication rates for surgery can go way down, however, if the operation is done by a team that has great expertise in performing the nerve-sparing technique that was introduced by Patrick C. Walsh of the Johns Hopkins University School of Medicine in the early 1980s. Surgeons who use this procedure avoid cutting into two bundles of nerves and blood vessels that are needed for a penile erection and that touch the surface of the prostate gland. The procedure, which is now applied in many medical centers, also reduces bleeding and can facilitate reconnection of the urethra to the bladder after its severance during surgery.

Proponents of screening and prompt treatment of small tumors understand that there are serious risks involved in undergoing aggressive treatment aimed at a cure. They, like opponents of such treatment, are also disturbed by the lack of methods for definitively distinguishing patients who have indolent tumors (and so need no intervention) from those whose tumors are likely to progress to a higher stage. But they believe that the patients who have the best chance of avoiding the horrors of metastatic disease should be given the opportunity to try to do so. Those patients are ones who are asymptomatic and who have small tumors that are confined to the prostate gland. To find these people, they reason, one must do screening tests.

Because of the risks and costs, even advocates of aggressive therapy for early cancer are divided over which age groups to include for screening and treatment. Some would rule out patients older than 70, on the theory that they are likely to die of other causes before their tumors progress enough to cause serious trouble. Others are more flexible and would actively seek and treat patients who were well into their 70s, if the men were in good health.

Most advocates of invasive therapy for early-stage disease (and some who are skeptical of its benefits) would, however, favor treating patients who are in their 40s and 50s, including those with well-differentiated, microscopic malignancies. First, some data suggest that prostate cancer is more virulent in younger men. Second, even if a slow-growing tumor developed in these people, chances are quite good that they would live long enough to see the tumor progress and metastasize.



SHAPE, SIZE AND ARRANGEMENT of cancerous cells in the prostate gland are often examined for clues to a tumor's virulence. Well-differentiated tumors (*top two panels*)—those that most resemble normal glandular tissue—are thought to behave less aggressively than do less organized, moderately (*middle panel*) or poorly (*bottom two panels*) differentiated tumors. In the grading scheme depicted above (the Gleason system), the degree of differentiation is indicated by a numerical value ranging from 1 to 5. The score for the predominant pattern is often added to that of the next most prevalent pattern to yield a final Gleason score ranging from 2 to 10.

One of my patients at the Dana-Farber Cancer Institute at Harvard Medical School would agree with treatment enthusiasts. He was diagnosed with prostate cancer at the age of 72 after his PSA score was slightly elevated in two successive tests (performed a year apart). A biopsy of the prostate indicated that a single cluster of cancerous cells was present. After many additional tests, including biopsy of his lymph nodes, he was deemed to be free of metastasis. Wanting to be rid of the cancer, he underwent a radical prostatectomy, from which he had a basically uneventful recovery. Although he was moderately sexually active before surgery and his potency had not yet returned when I last saw him, he remained delighted with his decision.

Yet one could argue that this patient should never have been treated. Opponents of early therapy are worried by an almost sixfold increase in the number of prostatectomies that were performed between 1984 and 1990, many in men older than 70 years. They point out that there still is no proof that aggressive therapy for early disease prevents the development of metastatic tumors and saves lives. That being the case, they fear that such treatment—which is aimed at avoiding a possibility that may never materialize (metastasis to the bones)—is condemning too many men to years of impotence, incontinence and other disabilities.

They contend as well that even if a cure were achieved, the side effects of treatment would often outweigh the benefits of any extra survival time gained. If treatment is undesirable, it follows that screening asymptomatic men for evidence of prostate cancer is unnecessary and wasteful.

Included in the concern of skeptics are men such as another of my patients. He was a 51-year-old, sexually active businessman who underwent a nerve-sparing radical prostatectomy after noninvasive and surgical tests showed he harbored microscopic cancer throughout his prostate gland (stage A2 disease). After the operation, he eagerly awaited the return of erectile function, knowing that about 70 percent of initially potent men in his age group who underwent the same procedure regained such function within a year.

Unfortunately, 18 months later, his ability to have an erection had not returned. Greatly concerned, he tried penile injections of substances that dilate the blood vessels, but they did not work well for him. In spite of counseling, he has had bouts of severe depression and has lost enthusiasm for maintaining his vigorous business interests. He is now considering insertion of a penile prosthesis, and he wonders if having surgery was the right decision.

Three recent papers are often cited as evidence against invasive treatment for localized disease. One was published in 1992 by Jan-Erik Johansson of the Örebro Medical Center Hospital in Sweden and his colleagues. The team followed 223 patients (mean age of 72 at diagnosis) who had localized prostate cancer (basically stage A or B) and were managed with "watchful waiting." That is, they received therapy (such as transurethral resection) for urinary symptoms or other discomforts but received no other treatment until their tumors metastasized. Then they were given hormonal therapy. At the end of 10 years, 124 subjects had died, but only in 19 of them (8.5 percent of the 223) was cancer listed as the cause. The survival rate was comparable to that typically reported in patients who receive aggressive treatment.

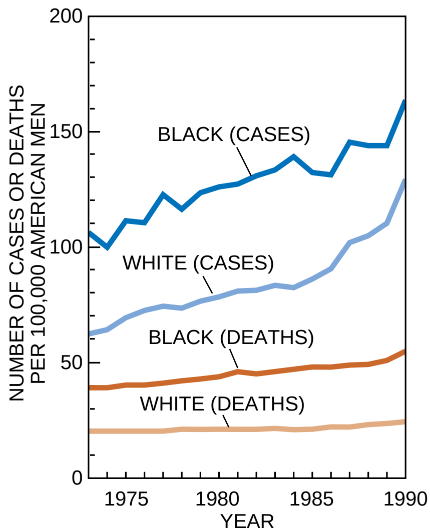
Another study favorable to watchful waiting was published in 1993 by Craig Fleming of the Oregon Health Sciences University and his colleagues in the Prostate Patient Outcomes Research Team (PPORT). The investigators in this study, from several medical centers, used data on treatment outcomes in the existing medical literature to try to calculate risks and benefits of treatment for men who are older than 60. They asked whether any extra survival time resulting from treatment of stage A and B disease would be outweighed by reductions in quality of life caused by treatment.

The PPORT group members concluded that for patients with well-differentiated tumors, treatment offered little benefit over watchful waiting. They also determined that for patients with moderately or poorly differentiated tumors, treatment might be worthwhile—providing up to 3.5 good years of life in some scenarios but no benefit in others. They also asserted that “men aged 75 years and older are not likely to benefit from either radiation therapy or radical prostatectomy when compared with watchful waiting.”

The third study, reported early this year, was carried out by Gerald W. Chodak of the University of Chicago and many co-workers. The group analyzed published data from six trials in which men with localized cancer (again, basically stage A or B) received no immediate therapy. The investigators found that aggressive treatment in such men may result in a lower death rate from cancer after 10 years; however, the difference appears to be small. In addition, the team noted, aggressive treatment “may have a substantial adverse effect on the quality of life.”

These studies are provocative. Yet they are seriously flawed and thus are far from definitive. For example, in the study by the Swedish group, not all patients were selected randomly. Many were chosen specifically because they had well-differentiated tumors. Further, a disproportionate number of subjects were elderly or had other life-threatening disorders. Therefore, the conclusions drawn cannot safely be applied to men who have moderately or poorly differentiated cancers, who are young or who are healthy except for the presence of prostate cancer. Similarly, the Chicago study has been criticized for containing mainly elderly patients, which puts into question its applicability to younger individuals.

The conclusions of Fleming’s group are problematic as well. At least some of the patients who were identified as having been followed with watchful waiting did not, in fact, go without treatment. They were given hormonal therapy early in the course of their disease. Such therapy is not a standard treatment for stage A or B disease, but it is not equivalent to a lack of treatment. Indeed, there is reason to think that hormonal therapy given early in the course of prostate cancer may benefit patients. Moreover, only sexually active men were selected; for them, impotence would reduce quality of life more than it would for other men. Hence, the decision to focus on sexually active men probably biased the findings against treatment. Thus, the benefits attributed to watchful waiting by the PPORT group may be smaller than has been claimed. In fact, uncritical interpretation of the findings reported in this and related studies may be leading many men who need treatment to opt for watchful waiting and to thereby lose their chance for a cure.



BLACK AMERICANS (*dark lines*) are more likely than their white counterparts (*light lines*) to be diagnosed as having prostate cancer and to die from it, according to the National Cancer Institute. More research is needed to explain these worrisome disparities.

Financial arguments are also invoked against widespread screening for, and treatment of, localized prostate cancer. More than 10 million Americans are thought to have microscopic prostate cancer right now. If all those men were found and treated in one year (at a cost of $8,000 to $11,000 for radiation alone and $10,000 to $18,000 for surgery alone), the bill would be astronomical.

On the other hand, it is improbable that all 10 million men with microscopic cells would be identified. For instance, detection efforts would probably be concentrated in selected age groups. It is also possible that the money saved by preventing metastatic disease would ultimately be offset by the costs of screening for, and treatment of, early prostate cancer.

Perhaps most important, some recent analyses imply that the PSA test does not readily detect indolent tumors and would not result in the diagnosis of cancer in millions of men who harbor quiescent microscopic cancers. Notably, data collected by William J. Catalona of the Washington University School of Medicine suggest that about 90 percent of the cancers revealed by the PSA test are diffuse or are moderately to poorly differentiated—features that have been associated with some degree of virulence. This finding is worth exploring

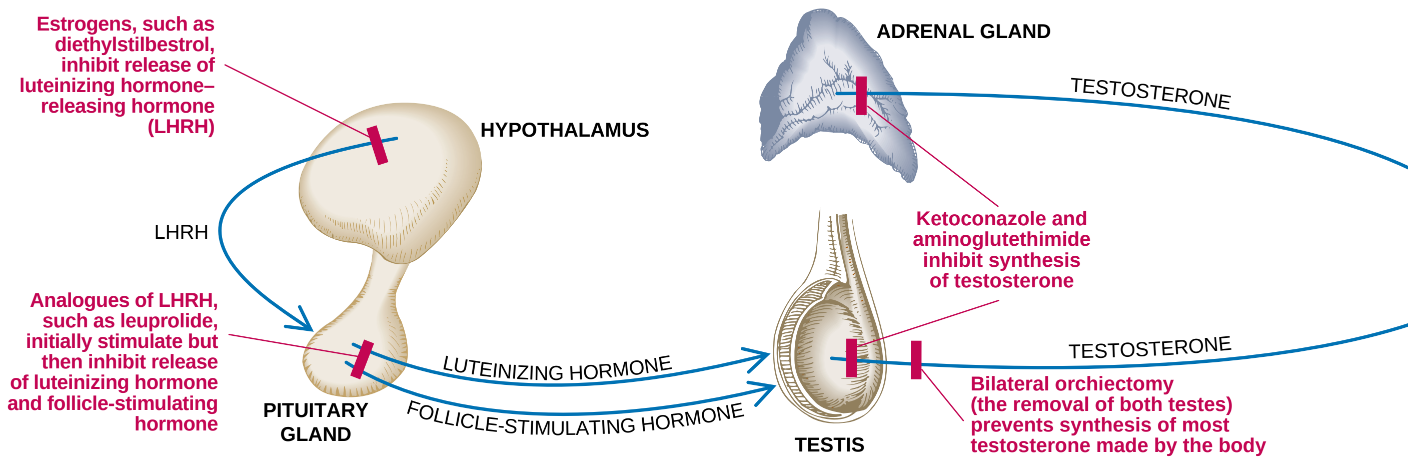
Risks of Aggressive Treatment

Complications are suffered by a fair percentage of patients who undergo aggressive treatment—radiation therapy or surgery—for localized prostate cancer, according to typical reports in the recent medical literature (*top two rows*). The reported complication rates for surgery can go even higher when patients are surveyed (*in parentheses*), as was done by Floyd J. Fowler, Jr., of the University of Massachusetts at Boston and his co-workers. The surgical risks often decline (*third row*), however, if patients are relatively young and have minimal tumors and if the operation is performed by a team having extensive experience with nerve-sparing procedures pioneered by Patrick C. Walsh and his colleagues at the Johns Hopkins University School of Medicine.

	PERCENT WHO DIE FROM THERAPY	PERCENT WHO BECOME IMPOTENT	PERCENT WHO SUFFER MILD TO SEVERE INCONTINENCE
RADIATION	<0.1	30–60	0–2
SURGERY	0.6	30–50 (*70 or more*)	2–15 (*30*)
SURGERY AT TOP FACILITIES	0.1	20–30	Severe: 0 Mild: 4–7

Options for Treating Advanced Prostate Cancer

Various therapies (*red type*) for metastatic (stage D) prostate cancer (and for selected cases of stage C cancer) share a common aim: interruption (*red bars*) of the biological pathways (*blue arrows*) that lead to the synthesis of testosterone and to its action in prostate cells (*far right*). Such a blockade prevents testosterone from fueling the growth of tumors derived from cancerous prostate cells. It can also shrink tumors (at least for a time). Recent studies suggest that administration of two therapies together (such as leuprolide and flutamide) is often more effective than using one by itself. The drug finasteride, taken by many men who have benign prostatic hyperplasia, interrupts this same



more thoroughly because it implies that the PSA test might consistently reveal cancers that deserve medical attention.

In the absence of conclusive data that watchful waiting is a safe alternative to treatment, I could not in good conscience tell a patient who had a potentially curable cancer that active intervention was inappropriate. I might consider monitoring an elderly patient with a stage A1 tumor and not intervening until I had some evidence that the cancer was becoming more active. But I see no proof that withholding treatment from other patients with localized disease is the ideal course of action. I share with my patients a dread that they will develop metastatic disease in their bones.

Because of this view, I am among the physicians who think the screening recommendations of the American Cancer Society are reasonable for now. The society proposes that men older than 40 who have no symptoms of prostate cancer undergo an annual rectal exam, that men older than 50 also obtain a PSA test and that males at high risk for prostate cancer (such as those who have a family history of the disease) be screened at younger ages.

I should note, however, that the National Cancer Institute (NCI) recommends only the use of the rectal exam in men older than 40 and is currently assessing the value of using the PSA test for screening. Many physicians in Europe have little enthusiasm for screening. Although I favor screening in principle, I also think it is important that patients be apprised of the benefits and risks of treatment before they are given a PSA test. Patients who say they would forgo aggressive therapy for localized disease could also forgo the test.

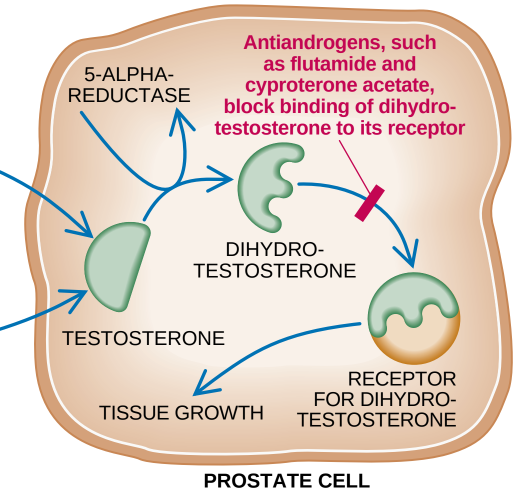
Stage A and B treatments are not the only focus of dispute. Another, gentler debate revolves around the use of experimental therapies for stage D disease. The disagreement is best illustrated by the experience of a third patient of mine. This man was 65 years old when he was found to have prostate cancer that had spread extensively. At the time, 1987, the standard treatment consisted of bilateral orchiectomy, delivery of the estrogen diethylstilbestrol (more commonly known as DES) or administration of analogues of the hypothalamic product LHRH (luteinizing hormone–releasing hormone). My collaborators at several institutions and I had recently shown that LHRH analogues were as effective as DES but caused milder side effects.

I was also able to offer him a promising experimental therapy: an LHRH analogue combined with an investigational antiandrogen known as flutamide. Although prevailing opinion held that combining hormonal therapies added no benefit, preliminary evidence to which I was privy indicated that the experimental combination might well extend survival time. I also knew how to obtain the investigational drug and permission to use it.

For the sake of his self-image, my patient did not want to have his testicles removed. He pinned his hopes on the combination drug therapy. Between April 1987 and the middle of 1991, his metastatic disease almost disappeared. As would be true with any treatment he chose, he was rendered impotent. He also lost interest in sex and had less physical intimacy with his wife. But he seemed to accept these losses and remained in all other ways fully functional and productive. In fact, his emotional intimacy with his wife increased.

In 1991 his cancer became resistant to hormonal treatment. In the last four months of his life, he suffered many complications of prostate cancer: agonizing bone pain, weight loss, muscle weakness, an inability to walk and constipation brought on by the narcotics he needed to control his pain. He died 54 months after his diagnosis, having survived much longer than the 18 to 24 months that would have been expected (given the extent of his disease) had he been given a standard therapy.

Chances are excellent that the experimental treatment was responsible for his prolonged survival. This possibility raises the question of whether all patients with terminal illnesses should have access to new, promising treatments, much as patients who suffer from AIDS often have hastened access today. This question continues to be pathway. In prostate cells, it blocks the enzyme 5-alpha-reductase from converting testosterone to dihydrotestosterone. It has not shown value for treating prostate cancer but is under study as a preventive.



hotly debated by physicians, pharmaceutical companies, regulatory agencies and patient-advocacy groups.

In my patient's case, the preliminary encouraging data on the combination therapy were later supported. Today it has governmental approval and is in broad use. Yet for every instance in which preliminary data are eventually confirmed, there is another in which an initially exciting treatment proves useless or harmful. There is also the risk that giving patients early access to drugs will prevent clinicians from carrying out large trials in which patients are randomly chosen to receive either an experimental treatment or a standard therapy. Failure to perform prospective, randomized trials may hinder efforts to evaluate safety and efficacy.

There is no easy answer. Nevertheless, when people are dying and existing treatments are inadequate, waiting for the perfect studies to be conducted is far from ideal. Some intermediate ground must be discovered, in which regulatory agencies maintain flexibility, and physicians and patients are kept up-to-date about preliminary findings. At the same time, pharmaceutical companies must be encouraged to make drugs available to patients who have life-threatening illnesses and must themselves be assured of continuous feedback about side effects and efficacy.

My patient's rejection of orchiectomy, together with the 51-year-old's despair over his impotence, highlights another shortcoming of current treatment. As women who have breast cancer know too well, the medical system does not deal particularly well with the psychological impact of threats to one's self-image and sexual activity. Men who must consider sacrificing their potency in order to gain extra years of life should be given as much emotional support as possible.

The three patients I have mentioned so far are all white. For black men, further treatment-related questions arise. Black Americans may have the world's highest incidence of prostate cancer, and they are twice as likely as white Americans to die from the disease. Many explanations have been proposed but none proved. Some evidence suggests that college-age black men have higher levels of androgens circulating in their blood; could the elevated levels lead to more cancer? Do black American males die more often (and more quickly) because, as a group, they have less access to care or because they seek medical treatment later, when their disease has progressed to an advanced stage? Or is prostate cancer simply more virulent in these individuals or less responsive to therapy? Might these men also be given fewer opportunities to try investigational therapies? Such questions need answers, and soon.

Clearly, prostate cancer poses complex medical management issues. Unfortunately, virtually all those issues will remain contentious until the data needed to resolve them are collected. Both governmental and private sources of funding for health care should make research into prostate cancer a greater priority. For fiscal year 1994, the NCI is estimated to have budgeted some $40 million for such study. This amount is about $26 million more than was spent in 1991, but it is still a far cry from the funds that are required. (By comparison, more than $250 million is expected to be spent for breast cancer, which will kill some 46,000 women in 1994.)

Much additional research is needed on the causes of prostate cancer, so that ways to prevent it can be developed. It is also crucial to identify the factors that spur dormant cancers to grow rapidly only in some individuals. Knowledge of such factors, and identification of the molecular abnormalities that foreshadow and accompany the transition to aggressiveness, could help physicians distinguish patients who can benefit from vigorous treatment from those who would not. Physicians additionally need ways to identify microscopic metastases so that they can best match treatment to the stage of disease. And research into the causes of hormonal resistance is critical to extending the health and life of patients with metastatic disease.

Clinical research must be expanded as well. The only way to determine whether mass screening for prostate cancer makes sense, whether the benefits of aggressive treatment outweigh the risks and whether prostatectomy and radiation are equally effective is to carry out large, long-term trials in which comparable patients are randomly assigned to different experimental conditions. The NCI is beginning trials to address some of these questions, but more work is essential. Research should also pursue suggestions that the administration of hormonal therapy can prolong life if given to patients with metastatic disease when the metastases are still minimal and are not yet causing symptoms. In parallel with such studies, there must be broader investigation into new approaches to treatment.

Only with adequate funding will the ongoing debates give way to well-grounded consensus. And only then will we achieve the shared goal of everyone involved in the battles over screening and treatment: a marked increase in survival time and in the quality of life enjoyed by the men in whom prostate cancer develops.

FURTHER READING

NATIONAL CANCER INSTITUTE ROUNDTABLE ON PROSTATE CANCER: FUTURE RESEARCH DIRECTIONS. Andrew Chiarodo et al. in *Cancer Research,* Vol. 51, No. 9, pages 2498–2505; May 1, 1991.

HIGH 10-YEAR SURVIVAL RATE IN PATIENTS WITH EARLY, UNTREATED PROSTATIC CANCER. Jan-Erik Johansson et al. in *Journal of the American Medical Association,* Vol. 267, No. 16, pages 2191–2196; April 22/29, 1992.

PROSTATE CANCER: SCREENING, DIAGNOSIS, AND MANAGEMENT. Marc B. Garnick in *Annals of Internal Medicine,* Vol. 118, No. 10, pages 804–818; May 15, 1993.

A DECISION ANALYSIS OF ALTERNATIVE TREATMENT STRATEGIES FOR CLINICALLY LOCALIZED PROSTATE CANCER. C. Fleming, J. H. Wasson, P. C. Albertsen, M. J. Barry and J. E. Wennberg in *Journal of the American Medical Association,* Vol. 269, No. 20, pages 2650–2658; May 26, 1993.

THE PROSTATE-CANCER DILEMMA. Charles C. Mann in *Atlantic Monthly,* Vol. 272, No. 5, pages 102–118; November 1993.

RESULTS OF CONSERVATIVE MANAGEMENT OF CLINICALLY LOCALIZED PROSTATE CANCER. Gerald W. Chodak et al. in *New England Journal of Medicine,* Vol. 330, No. 4, pages 242–248; January 27, 1994.

Precious Metal Objects of the Middle Sicán

A Peruvian culture older than the Incas made unprecedented use of gold and other metals. Studies of Sicán metalworking techniques offer hints about this mysterious society

by Izumi Shimada and Jo Ann Griffin

Gold ceremonial masks and knives are popular symbols of pre-Hispanic Peruvian culture. Examples adorn the covers of books on Peru and serve as emblems for some Peruvian institutions. These precious metal artifacts are often attributed, even by knowledgeable persons, to the Incas or to their coastal rivals, the Chimú. Yet many of them are not Incan or Chimú at all: they were created much earlier by the Sicán culture, which was centered in Batán Grande in northern Peru and flourished from the eighth to the 14th centuries.

The Middle Sicán era, between A.D. 900 and 1100, produced enormous quantities of precious metal artifacts, many showing extraordinarily high craftsmanship. Recently we and our colleagues from several disciplines scrutinized the metalwork from one Middle Sicán trove in an attempt to reconstruct the technology and organization of precious metal production and to define the meaning of those products within the culture. We determined that the scale and the range of metal use by the people of the Middle Sicán was unprecedented in the pre-Hispanic New World. That culture's extensive production of arsenical copper ushered the bronze age into northern Peru. Gold alloys were the most prestigious media for political, social and religious expression. In fact, we suspect that metallurgical production was the prime mover of Middle Sicán cultural developments.

IZUMI SHIMADA and JO ANN GRIFFIN have joined forces to investigate the precious metal artifacts of the Sicán. Shimada has conducted fieldwork on the north coast of Peru for the past 20 years and maintains interests in ancient technologies and the evolution of complex societies. A native of Japan, Shimada received his Ph.D. from the University of Arizona. He joined the department of anthropology at Southern Illinois University in 1994. Jo Ann Griffin has for almost 30 years been a goldsmith and conservator specializing in pre-Hispanic metallurgy, working with some of the largest public and private collections of pre-Hispanic gold in the world.

Ambiguity and ignorance have traditionally shrouded precious metal artifacts of the Middle Sicán. Almost all of those in private and public collections were looted from tombs within what is today the Poma National Archaeological and Ecological Reserve in the Batán Grande region, about 800 kilometers north of Lima. The modern period of grave robbery began in the 1930s. Treasure hunters sank vertical prospecting pits into likely spots, then dug horizontal tunnels outward. With the discoveries of more rich tombs, the extent of the looting continued to increase through the 1940s and 1950s. It culminated in the late 1960s, when a bulldozer was employed for a year to remove the surface soil so that outlines of the tomb pits could be seen more easily. Intense looting took place sporadically until the mid-1970s, effectively hindering any long-term scientific study of the regional prehistory. When one of us (Shimada) began fieldwork in 1978, he counted more than 100,000 looters' holes and hundreds of long bulldozer trenches on aerial photographs of the Poma National Reserve.

The lack of contextual information for those looted artifacts greatly limits understanding of their sociopolitical, religious and economic significance. Moreover, questionable and undocumented measures were often taken to "restore" them before they reached collectors or museums. Pigments, feathers and ancient tool marks on gold objects could have been removed by careless cleaning. "Missing" inlay pieces or bangles were often arbitrarily replaced. As a result, the appearance of objects cannot be taken at face value, which limits the information that can be drawn from them. Any attempt to understand the objects, their cultural significance and the techniques used in their manufacture is therefore best founded on those artifacts scientifically recovered from intact tombs.

The opportunity to gain just such an understanding came about with the first scientific excavation of the tomb of a member of the Middle Sicán elite at Huaca Loro, an adobe platform mound in the Poma reserve. The tomb was apparently one of a string left by the Middle Sicán, many of which had already been looted, along the east and north bases of Huaca Loro. Shimada recognized it during a survey of Batán Grande in 1978. He planned its excavation over the next 10 years as part of his broader sampling of Sicán tombs for the elucidation of that culture's social organization. Preparations included assembling a group of specialists and piecing together a Sicán cultural chronology, as well as the performance of other background research.

In particular, Shimada needed to make certain that the groundwater level was sufficiently low to allow safe excavation. He also gave a series of public talks on the scientific value of the planned tomb excavation to local residents, who were weary of tomb looting. The chamber was finally excavated, under Shimada's supervision, by the Sicán Archaeological Project between October 1991 and March 1992.

The central person buried in the tomb was a 40- to 50-year-old man who had been one of the elite. He was accompanied by the bodies of two young women and two children who apparently had been sacrificed. The six-month excavation yielded approximately 1.2 tons of diverse grave goods packed in a burial chamber roughly three meters on a side at the bottom of an 11-meter vertical shaft. By weight, metal objects and scrap account for nearly three quarters of the grave goods. Most of the objects appear to be gold of 14 to 18 karat. Some objects and nearly all the scrap are *tumbaga,* gold-copper or gold-silver-copper alloys roughly equivalent to the 10- to 14-karat gold commonly used today for commercial jewelry.

Arranged concentrically, the objects surrounded the body of the man, which was thoroughly painted with cinnabar (an intensely red paint of mercuric sulfide and a binder). The body was seated and placed in an inverted position. The head with its three sets of attached ear ornaments and its large gold mask was rotated 180 degrees and then bent back to point upward. A mantle (its cloth long since decayed) onto which nearly 2,000 gold foil squares had been sewn was laid atop the inverted body. Placed on, around and underneath the body were a staff with gold and *tumbaga* ornaments on top of a wooden shaft, a gold headdress with a sculptural representation of an animal head, a pair of gold shin covers, a pair of meter-long *tumbaga* gloves (one holding a gold cup with a silver rattle base), a silver ceremonial knife (or *tumi*) and a cluster of six magnificently made pairs of gold earspools. His chest was bedecked with a nearly 10-centimeter-thick layer of beads (made of sodalite, amethyst, quartz crystal, turquoise, fluorite, calcite, shell and other materials). Farther away, near the edges of the chamber, were about 500 kilograms of *tumbaga* scrap and more than 250 kilograms of arsenical copper implements.

By far the most impressive find within the tomb is Gold Cache 1, which we discovered at the northwest corner of the burial chamber. Inside a rectangular box lined with woven mats were at least 60 major objects, most of gold sheet, the balance of silver or *tumbaga.* This cache contains a peculiar mixture of ritual paraphernalia and personal ornaments: five crowns, four headbands, at least 12 *tumi*-shaped head ornaments, at least six sets of gold feather head ornaments, three *tumbaga* fans and 14 large disks that were either ornaments for staffs or the backs of headdresses. At the bottom were the largest objects: four sets of parabolic headdresses that would have been set on top and in front of the crown.

Two of the seven niches carved into the walls of the burial chamber also contained metal objects. A pit dug into the largest niche, in the east wall, was packed with an estimated 1,500 bundles of *naipes.* These arsenical copper objects of uniform shape and size may have served as currency. Each bundle consisted of 12 or 13 *naipes.* The pit also held two silver alloy *tumi* knives, thousands of small *tumbaga* foil squares and at least two dozen *tumbaga* masks identical in shape to the large gold mask on the buried man but smaller, technically inferior and less ornamented. A second niche—Gold Cache 2—contained another collection of gold ornaments and ritual objects. We also recovered from this grave more than 50 kilograms of diverse stone and shell beads, the carved wooden frame of a litter, about three kilograms of cinnabar and 21 ceramic vessels.

GOLD CUTOUT FIGURE is believed to represent a Sicán lord buried in a tomb at Huaca Loro in the Poma National Archaeological and Ecological Reserve in northern Peru. After decades of looting, this grave was the first tomb of one of the members of the ruling class to be excavated scientifically. The gold figure (*about 12 centimeters in height*) probably decorated an elaborate headdress.

Our study differed from earlier ones in two important respects: we had access to materials from an intact tomb, and we had the benefit of the insights of metalworking specialists. Most published studies have focused on single aspects of the manufacture or use of pre-Hispanic precious metal artifacts. Such studies relied primarily if not exclusively on laboratory analyses of looted objects. Also, the investigators who have made inferences about the manufacturing techniques have rarely had personal knowledge of metalworking as a craft. The reconstructions emerging from those narrow, academic studies are therefore tenuous and biased. What this area of archaeology needs is more comprehensive studies of scientifically excavated samples from a multitude of analytical and interpretive perspectives. Only then will we gain an in-depth appreciation of the organization of metal production and the meaning of metal products within a culture—a "holistic vision" of metallurgy.

One of us (Griffin) is a skilled metalsmith and longtime conservator who has experience with many pre-Hispanic collections. Because of her background, she was able to elucidate much about Sicán precious metal production by examining the Huaca Loro artifacts. Specialists in related fields helped us to interpret the evidence of other metal samples, feathers, beads and remains associated with the metal objects.

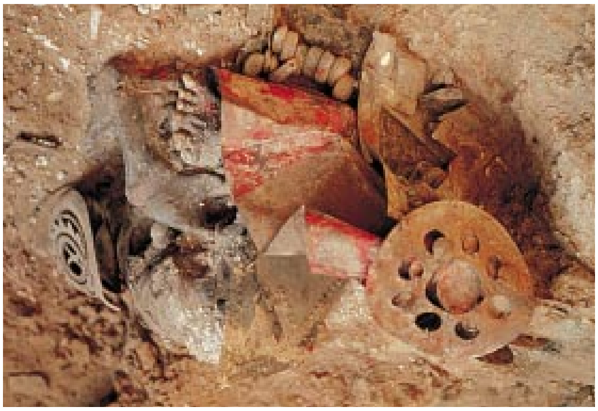
From its documented beginning, Andean metallurgy emphasized the use of

Excavating a Middle Sicán Tomb

The tomb at Huaca Loro held the remains of a member of the Sicán elite, a man 40 to 50 years of age. He was accompanied by two young women and two children, who had apparently been sacrificed. He was buried in a seated, inverted position (*1*), but the head had been rotated 180 degrees and bent back so that the top pointed upward. His face was covered by a large gold mask (*2*), which is shown shortly after removal from the tomb while it was still extensively deformed and spotted with soil (*3*). During the restoration process (*4*), patches of nylon fabric were placed over parts of the mask to prevent the cinnabar with which it had been painted from flaking off. The fully restored mask and headdress (*5; about one meter in height*) offer stunning evidence of a sophisticated group of master craftsmen and the society that supported them. In addition to the bodies and their elaborate coverings, the tomb contained 1.2 tons of diverse grave goods (*6*), most of precious metal.

1

2



3

4

sheet metal fashioned from ingots with stone anvils and hammers. Gold crowns and other ornaments found at various sites are all essentially gold sheets decorated with repoussé and cutout designs. In terms of the dimensions, smoothness, consistent thickness and overall quantity, the objects from the Huaca Loro tomb are prime examples of this sheet-metal tradition. The primary tools for making sheet and wire are handheld stone hammers and companion anvils, chisels and chasing tools made of arsenical copper and *tumbaga*. The hammers are commonly of magnetite, hematite or fine-grained basalt; they range from very tiny to the size of a man's fist. One hammer face is usually domed for stretching the metal, and the other face is flat for planishing—removing the shallow dimples left by the stretching blows. When properly executed, this technique yields a flat, smooth sheet. Although simple to describe, the task requires considerable time and skill. Because the sheet-metal maker must forge the gold while it is cold, the metal must be annealed regularly to prevent stress cracks.

5

6

Many gold objects in the Huaca Loro grave attest to remarkable expertise in sheet making. Consider, for example, the long borders on two parabolic head ornaments. Each is a two-meter-long continuous strip of forged metal with an even width (around 4.5 centimeters) and thickness (about 0.15 millimeter)—a virtuoso performance in sheet making. The mask (46 by 29 centimeters) that covers the face of the man buried in the tomb is another tour de force. It was fashioned from a sheet about 0.6 millimeter thick. The metal had to be thin to keep down the mask's weight (only 677 grams) yet thick enough to allow the large nose to be raised from the center.

More than a dozen *tumi*-shaped headdresses provide additional illustrations of expert sheet-metal making. The tang, or stem, of each headdress had to be narrow but sufficiently stout (about a millimeter thick) to stay upright when inserted into a turban or crown socket. At the top the sheet is only about 0.15 to 0.18 millimeter thick. The smith knew just how much to planish the sheet to give it the right amount of springiness so it would wave with each movement of the head but not crack or bend.

The same is true of six sets of gold feathers that are believed to have been part of an elaborate headdress. The sets we have studied consist of 11 or 12 feathers, each about 20 to 21 centimeters long and two centimeters wide. Each feather tapers in thickness from the stem (about 0.10 millimeter) to the upper tip (about 0.07 millimeter). The sets as a whole have a fan shape: the central feather is straight, and those to the sides are increasingly more curved to the right or left. The component feathers are mechanically joined by straps and slots near their stems. Below the straps, each feather has a slight ridge along the longitudinal axis to provide some rigidity. Preserved on the stems are the imprints of fine threads that apparently stitched them to some cloth backing. The overall design and structure, as with the *tumi*-shaped headdresses, allowed the feathers to sway gently with head movements while staying rigid and light enough to mount securely on the headdress.

The six pairs of gold earspools found near the southeast corner of the tomb demonstrate a level of technical mastery rarely seen in pre-Hispanic gold objects. They display a constellation of design features that may well represent a single school of goldsmiths. The varied and advanced metalworking effects, not usually found together, include forged wire, true filigree, excellent finish and polish, and protobrazing. Protobrazing is a superbly simple method of joining gold or silver alloys that utilizes either the copper in those alloys or verdigris (copper acetate) in an organic glue. The pieces to be joined are heated over charcoal in a reduction atmosphere; at the right temperature a new alloy forms where the metals touch.

To prevent the earspools from developing a torque while being inserted into the earlobes, the metalsmiths had made them from a relatively heavy gauge metal (about 0.35 to 0.55 millimeter thick). The front flange was domed by striking it from the back with a hammer against a shallow depression, probably in wood. The resultant domed flange greatly increased the twist strength while adding needed depth to the design.

In three of the pairs the metalsmiths implemented complex decorative designs by means of mechanical solutions that are simple and elegant. One such solution is the "tab and slot" joining of the sheet metal that forms the central rod connecting the front and back flanges. In several pairs the smiths also used wire structural supports to create the illusion that the central elements of the design "float" within the frame of the front flange.

One pair of the earspools, which might be called a "sketch" or trial piece,

Sicán Metalworking Techniques

Through its flaws, a gold beaker (*1*) helps to reveal the methods of the Sicán craftsmen. The beaker (*about 12 centimeters in height*) is decorated with the chased representation of a Sicán lord, shown in its entirety in a modern drawing (*2*). The tool used for chasing was too wide and its edges too sharp to execute the lord's round chin, resulting in a ragged line. The silver base of the beaker shows an "orange peel" texture, the result of overheating. There are also traces of silver on the lower part of the gold beaker, suggesting that a portion of the rim melted accidentally while the base was being joined to the bottom of the beaker. The small gold mask (*4; about seven centimeters in height*) shows the exquisite chasing-repoussé details more characteristic of these metalworkers. A close-up of the eyes and nose (*3*) illuminates the oval indentations left by the tools. Jo Ann Griffin, one of the authors, demonstrates the basic techniques used by the Sicán goldsmiths in photographs *5–9*.

1

5 **PRODUCTION of sheet-metal objects began with an ingot of metal. Using the domed face of the hammerstone, the ancient goldsmith struck the ingot on a stone anvil.**

6 **TO FLATTEN the sheet, the metalworker used the flat face of the hammer. This process also removed dimples caused by the domed face.**

7 **CUTTING the sheet to shape was done with a *tumbaga* chisel.**

shows how the Sicán goldsmiths gradually refined the mechanical solutions to problems posed by an intricate new design. In this pair, very fine gold wire was used to anchor a circular frame to the X-brace below it. Other pairs, which were presumably made later, show the use of permanent protobraze joins for the same purpose.

An additional indication of careful planning comes from the gold mask, which has its own pair of large earspools anchored directly on the metal earlobes by straps and slots. Not only do the size and shape of the earlobes match those of the back flanges of the earspools, but the slots for all three straps on each ear also match. The slots were punched through both pieces simultaneously by the same vertical strokes. It is quite likely that the mask and all the earspools were manufactured according to high-quality standards in the same workshop.

2

3

4

8 **REPOUSSÉ, employing another *tumbaga* tool, pushed up the gross form. The piece was worked against the resistant surface of a leather sandbag.**

9 **CHASING then applied details to the front of the object.**

Not all the objects found in the Huaca Loro tomb are flawlessly finished. One example is the small double-bottomed beaker (about 12 centimeters high and 10 centimeters in diameter) found in the hand of a ceremonial *tumbaga* glove. The base of the beaker is made of a raised silver sheet ornamented with cutout designs. It fit onto the base of the gold beaker and was intended to contain rattle stones. A portion of the rim of the silver base is melted, and the base shows an "orange peel" texture, the result of overheating. The lower part of the gold beaker displays a distinctly gray semicircular coating that extends up the side, where a portion of the silver base rim is missing. Those features suggest that part of the rim melted accidentally while the silver base was being joined to the bottom of the gold beaker. The cup was placed in the brazier upside down. It would not take more than a few seconds of overheating for the silver to melt and create the observed features. The base of the gold beaker was covered by a flash of melting silver as it flowed down toward the heat source.

Other technical details on the beaker are also informative. For example, it is decorated with three chased representations of a Sicán lord. To create that image, the goldsmith used a tool called a tracer about three to four millimeters in width. That tool was too wide and its edges too sharp for executing the round chin of the figure's face; as a result, the chin line is ragged. The metalworker made no attempt to correct the error when the chasing was finished on the outside of the beaker. This kind of error generally indicates haste or an apprentice worker.

The goldsmith would have stopped frequently to anneal the piece to prevent stress cracks from developing. How could he have known when to anneal? Immediately after annealing, the metal emits a dull sound when struck with a hammer. After repeated blows, the pitch of the sound becomes much higher, rising from a *thuk* sound to a *think*. With experience, one can tell from the pitch when it is time to anneal.

The impressive scale of sheet making during the Middle Sicán can best be seen in the 500 kilograms of scrap piled along the edges of the burial chamber at Huaca Loro. These piles are apparently not unique: local old-time looters recall finding similar quantities of scrap in other tombs nearby. In addition, we documented the extensive use of *tumbaga* sheets to line the interior of the gigantic Middle Sicán tomb at Huaca Las Ventanas. That tomb measured 15 by 15 meters at the mouth and three by three meters at the bottom, which was about 11 meters below the surface. Rectangular sheets of set dimensions were carefully placed side by side on the interior surface. They were then covered with cotton cloth showing elaborate polychrome religious images and scenes. The total surface area of the sheets lining this tomb may have exceeded 100 square meters.

The scrap is essentially small pieces left over from sheet-metalworking processes and rejects from manufacturing mishaps. It includes, for example, a partially used *tumbaga* ingot, square gold foils with poorly executed perforations, broken wires and bells, and sheet-metal trimmings that still retain the outline of the cutout pieces. Such scrap would have been carefully saved for recycling into new ingots.

The scrap clearly represents an enormous investment of manpower and materials. Its presence in the tomb testifies to the political power of the person buried there. In a recent experiment, using ancient stone hammers, Griffin needed about a day and a half to produce a uniformly thin sheet 10 by 15 centimeters in size from a 30-gram gold nugget. Moreover, the ancient Sicán metalworkers added another step: they treated *tumbaga* sheets with acid, which dissolved some of the base metal near the surface. As a result, the *tumbaga* sheets had an appearance that approximated that of 24-karat gold. This process is generally known as depletion gilding. The metalsmiths then burnished the sheets, which imparted an excellent finish and compacted the layer of gold left by the depletion gilding. In our opinion, this compacted layer is the peeling gold often seen on gilded *tumbaga* pieces. Some researchers have proposed that this gilt was deposited electrochemically, but none of the examinations of the sheets conducted by us and others using microscopes and electron microprobes can find any evidence to support that idea.

Sicán culture must have employed a sizable corps of master sheet makers who produced sheet goods for various applications. The remarkable degree of control over forging and planishing seen in these objects argues persuasively that those activities were in the hands of full-time specialists. These master sheet makers would have been assisted by perhaps dozens of apprentices who would have carried out the early stages of remelting scraps to prepare ingots for making sheets.

EARSPOOLS represent some of the most technologically sophisticated metalwork of the Middle Sicán. A back view (*bottom left*) shows construction details, including an X-shaped brace attached by protobrazing. A side view (*bottom right*) details another method of joining, the so-called slot-and-tab construction. Spools are roughly 10 centimeters in diameter.

This master-apprentice arrangement is clearly visible in the manufacturing stages of other objects. One crown in particular shows well-done chasing and perforations in the front but uneven hammer blows and perforations, as well as fine-scribed guidelines on the back. It is likely that the front was begun by a master who showed an apprentice how the remainder was to be done and then went on to another task.

This type of workshop would probably have required a series of multiroom shops, each with a fair number of apprentices and a sizable output. Sheet making, which entails long hours of rhythmic hammering periodically interrupted by annealing, most likely took place in a well-ventilated room. Polishing was probably done in a separate, well-protected room, because airborne sand and other contaminants would have wreaked havoc with the polishing efforts. Significantly, multiroom adobe structures atop the north platform of Huaca Loro and northeast of Huaca Las Ventanas have benches, split-level floors and many dispersed spots where one can find slag fragments, droplets of copper alloys and evidence that fire was used there. Those two areas were probably centers of metalworking.

Making metal sheet requires great finesse. The shaping and ornamentation of gold objects would have been in the hands of even more consummate master specialists. Because of their exceptional quality, innovative designs and technical distinctiveness, the mask and earspools in the Huaca Loro tomb are probably the products of only one or two masters. Other gold objects, we suspect, were manufactured in other workshops. Although those workshops may have performed different functions, it is unlikely that they were isolated: part of the apprenticeship training would have depended on frequent association with the masters. The apprentices were no doubt given repetitive tasks, such as making bangles, that were instructive but did not pose too many technical difficulties.

Some of the observed technical variation may reflect differences in the goldsmiths' personal styles. Many of the gold objects from Huaca Loro are near-

CUTOUT FIGURES (*about 12 centimeters in height*) once decorated the upper right sleeve of the dress worn by one of the two sacrificed women.

ly identical in size and shape but were clearly made in different ways. For example, on some rattles, bangles were attached to the "floating" circular bands by protobrazed wires, whereas others were attached by wire loops. Some of the sharp gold nails used on a dart thrower were cut in a sawtooth pattern from hand-forged wire with a chisel; others were cut from the end of wire that had been filed to a conical shape.

Such observations lead us to conclude that the production of metal objects was organized into task-specific work groups, which in turn were based on a nested hierarchy of masters, apprentices and other supportive personnel. Precious metalworking must not be viewed in isolation from other crafts. Considerable effort had to go into the procurement and preparation of feathers, cinnabar, hematite and other materials that covered the metal objects. Minerals, shells, bitumen and other substances were needed for inlays. Resins and pitch had to be prepared to make adhesives. Cloth had to be woven as a backing material. We know that arsenical copper was produced on a large scale at specialized settlements close to the mines. All these activities need to be considered to appreciate the impressive magnitude and complexity of the production of sumptuary goods during the Middle Sicán.

For the Sicán people to have invested so much effort in metalworking, metal objects must have held strong meaning for them. We have developed some working hypotheses about what that meaning was. Gold objects seem to have been the aesthetic locus of Middle Sicán art—they embodied the highest standards for artistic expression in the culture. And it is among the gold objects that we find the most explicit expressions of the important Middle Sicán icons and scenes. Ceramic decorations, on the other hand, present only partial or simplified versions of these portrayals.

Differential access to a range of metals seems to have marked the social strata. Approximately two dozen excavated burials can be grouped into those that contain no metal objects, those that contain only arsenical copper, those that have arsenical copper and *tumbaga* items and those that have gold in addition to those other materials. *Tumbaga,* along with gold and silver, seems to have been used to symbolize political power or social status and to convey religious messages. In terms of the scale of production and the range of use, *tumbaga* appears to top the list of precious metals. Yet it was secondary to gold in the perception of the Middle Sicán elite. The personal ornaments immediately surrounding the central body at Huaca Loro were all gold. The *tumbaga* objects were placed at the periphery of the burial chamber, and their use was probably auxiliary.

In other words, the gold objects were reserved for the personal use (including ornamentation and ritual paraphernalia) of the highest elite, whereas gilded *tumbaga* was used to decorate items associated with them as well as the objects used by lower echelon elites. *Tumbaga* allowed them to emulate their social superiors. Gilded *tumbaga* with relatively low gold content would have been a most practical substitute for meeting the broad demand for rich gold-colored sheet metal.

Many of the precious metal objects found in the tomb were probably used together in public settings for ostentatious displays to impress onlookers. The full ceremonial regalia of the important person buried in the Huaca Loro tomb offers a vivid example.

Depending on the ritual to be conducted, he would have worn various headdresses—sometimes a crown adorned with sets of gold feathers or *tumi*-shaped ornaments, sometimes a large parabolic headdress in addition to the crown. The upper perimeter and draping sides of that parabolic headdress would have been decorated with colorful bird feathers and bangles that reached almost to the shoulders. Over his face he would have worn a gold mask. He was probably carried on a wooden litter decorated with the carved heads of mythical animals. The litter was likely to have been flanked by people waving long *tumbaga* fans and preceded by someone holding a staff or standard almost two meters high, which was bright with gold and feathers. With each step, each breath of air, the bangles, gold feathers and other delicately articulated metal objects would have been set in motion to create a dazzling visual and auditory effect. It is not hard to be entranced by the thought of that luminous figure—or by thoughts of what future studies of Sicán artifacts may yet tell us about that lost culture.

FURTHER READING

PRE-COLUMBIAN SURFACE METALLURGY. Heather Lechtman in *Scientific American,* Vol. 250, No. 6, pages 56–63; June 1984.

COPPER-ALLOY METALLURGY IN ANCIENT PERU. Izumi Shimada and John F. Merkel in *Scientific American,* Vol. 265, No. 1, pages 80–86; July 1991.

A SICAN TOMB IN PERU. I. Shimada and J. Merkel in *Minerva,* Vol. 4, No. 1, pages 18–25; January/February 1993.

The Pioneer Mission to Venus

This multipart spacecraft spent 14 years scrutinizing the atmosphere, clouds and environs of the nearest planet. The results clarify the stunningly divergent evolutionary histories of Venus and the earth

by Janet G. Luhmann, James B. Pollack and Lawrence Colin

Venus is sometimes referred to as the earth's "twin" because it resembles the earth in size and in distance from the sun. Over its 14 years of operation, the National Aeronautics and Space Administration's Pioneer Venus mission revealed that the relation between the two worlds is more analogous to Dr. Jekyll and Mr. Hyde. The surface of Venus bakes under a dense carbon dioxide atmosphere, the overlying clouds consist of noxious sulfuric acid, and the planet's lack of a magnetic field exposes the upper atmosphere to the continuous hail of charged particles from the sun. Our opportunity to explore the hostile Venusian environment came to an abrupt close in October 1992, when the *Pioneer Venus Orbiter* burned up like a meteor in the thick Venusian atmosphere. The craft's demise marked the end of an era for the U.S. space program; in the present climate of fiscal austerity, there is no telling when humans will next get a good look at the earth's nearest planetary neighbor.

The information gleaned by *Pioneer Venus* complements the well-publicized radar images recently sent back by the *Magellan* spacecraft. *Magellan* concentrated on studies of Venus's surface geology and interior structure. *Pioneer Venus,* in comparison, gathered data on the composition and dynamics of the planet's atmosphere and interplanetary surroundings. These findings illustrate how seemingly small differences in physical conditions have sent Venus and the earth hurtling down very different evolutionary paths. Such knowledge will help scientists intelligently evaluate how human activity may be changing the environment on the earth.

Pioneer Venus consisted of two components, the *Orbiter* and the *Multiprobe.* The *Multiprobe* carried four craft (one large probe and three small identical ones) designed to plunge into the Venusian atmosphere, sending back data on the local conditions along the way. The *Orbiter* bristled with a dozen instruments with which to examine the composition and physical nature of Venus's upper atmosphere and ionosphere, the electrically charged layer between the atmosphere and outer space.

The *Multiprobe* was launched in August 1978 and reached Venus on December 9 of that year. Twenty-four days before its arrival, the *Multiprobe* carrier, or "bus," released the large probe; about five days later the bus freed the three small probes to begin their own, independent courses. The probes approached the planet from both high latitudes and low ones and from both the daylit and nighttime sides. In this way, information relayed by the probes during their descents enabled scientists to piece together a comprehensive picture of the atmospheric structure of Venus.

The *Orbiter* left the earth in May 1978, but it followed a longer trajectory than the *Multiprobe,* so it arrived only five days earlier, on December 4. At that time, the spacecraft entered a highly eccentric orbit that looped to within 150 to 200 kilometers from the planet's surface but carried it out to a distance of 66,900 kilometers. During its closest approaches to Venus, the *Orbiter's* instruments could directly sample the planet's ionosphere and upper atmosphere. Twelve hours later the *Orbiter* would have receded far enough from Venus so that the craft's remote-sensing equipment could obtain global images of the planet and could measure its near-space environment.

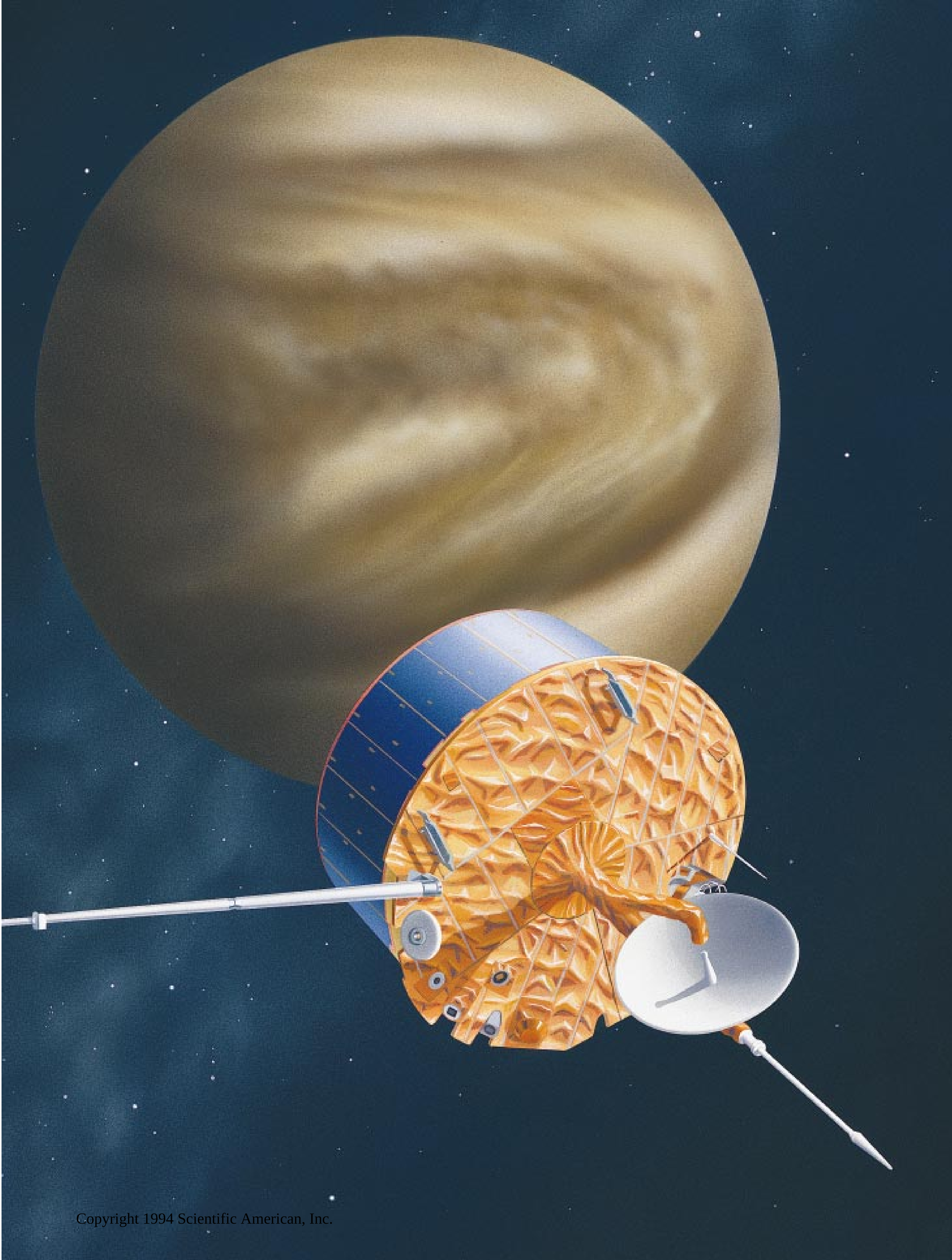
The gravitational pull of the sun acted to change the shape of the probe's orbit. Starting in 1986, solar gravity caused the *Orbiter* to pass ever closer to the planet. When the spacecraft's thrusters ran out of fuel, *Pioneer Venus* dove deeper into the Venusian atmosphere on each successive orbit until it met its fiery end.

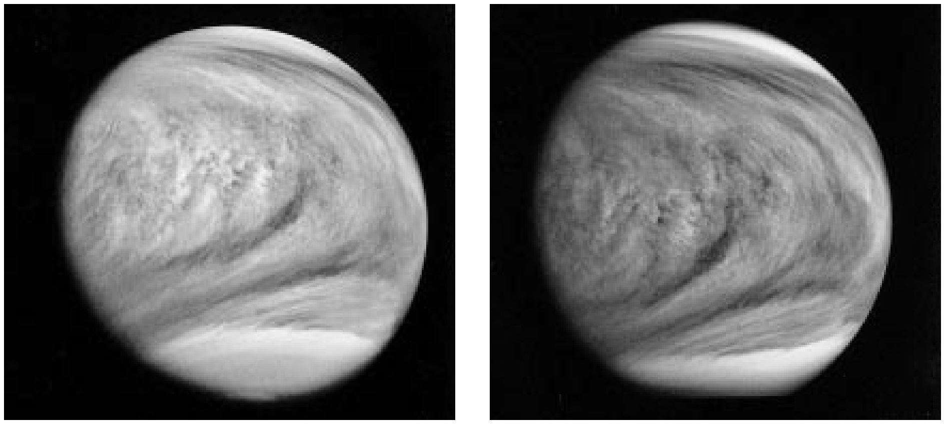
Well before the arrival of *Pioneer Venus,* astronomers had learned that Venus does not live up to its image as the earth's near-twin. Whereas the earth maintains conditions ideal for liquid water and life, Venus is the planetary equivalent of hell. Its surface temperature of 450 degrees Celsius is hotter than the melting point of lead. Atmospheric pressure at the ground is some 93 times that at sea level on the earth.

Even aside from the heat and the pressure, the air on Venus would be utterly unbreathable to humans. The earth's atmosphere is about 78 percent nitrogen and 21 percent oxygen. Ve-

JANET G. LUHMANN, JAMES B. POLLACK and LAWRENCE COLIN have studied different aspects of Venus as revealed by Pioneer Venus and by other missions. Luhmann received her Ph.D. in astronomy from the University of Maryland in 1974. Since 1980 she has been a research scientist at the Institute of Geophysics and Planetary Physics at the University of California, Los Angeles. Much of her work has concerned the interaction between planets and the solar wind. Pollack earned his Ph.D. in astronomy from Harvard University in 1965. He has been a research scientist at the National Aeronautics and Space Administration Ames Research Center since 1970. During those years, he has participated in numerous NASA planetary missions, including the current Galileo mission to Jupiter. Colin completed his Ph.D. in electrical engineering at Stanford University in 1964. Since then, he has worked as an aerospace technologist at the Ames Research Center, where he has specialized in atmospheric and space physics.

PIONEER VENUS ORBITER **regularly passed less than 200 kilometers above the planet's thick clouds of sulfuric acid, seen in exaggerated contrast in this artist's conception. During its 14-year lifetime, the probe circled Venus 5,055 times, gathering extensive information on the planet's atmosphere and outer surroundings.**





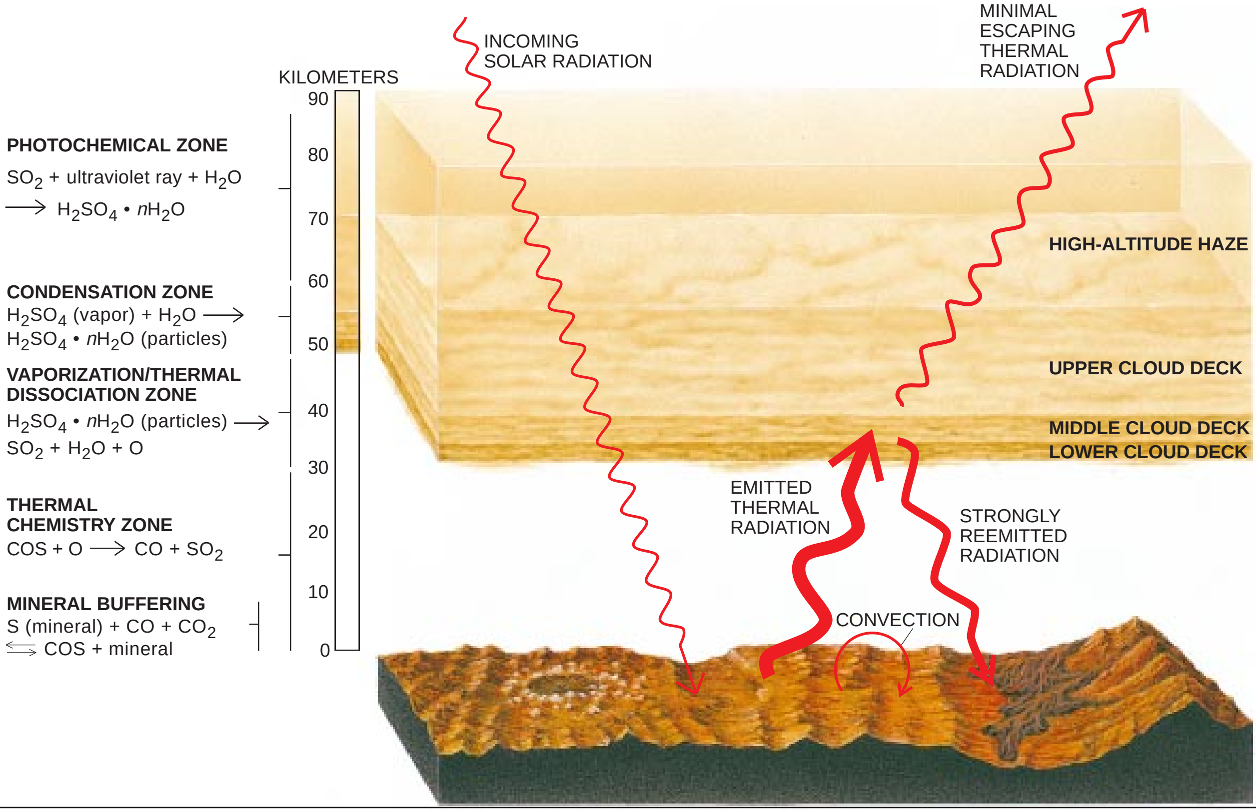
RAPID WINDS at Venus's cloud tops move 60 times as fast as the body of the planet. These false-color ultraviolet images from the *Pioneer Venus Orbiter* (*right*) show the quickly shifting cloud patterns. Atmospheric circulation is driven by solar radiation, which produces a north-south flow, known as a Hadley cell (*far right*). The rotation of the atmosphere transforms Hadley cells into predominantly westward zonal winds, which may be amplified by eddies.

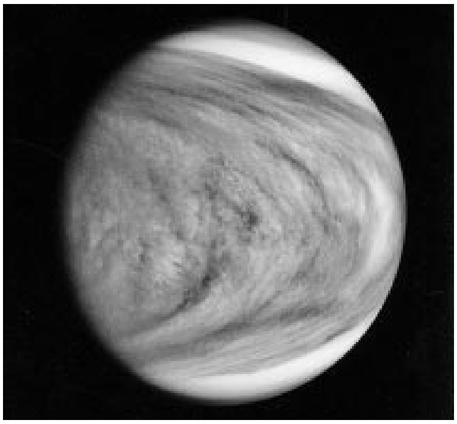
nus's much thicker atmosphere, in contrast, is composed almost entirely of carbon dioxide. Nitrogen, the next most abundant gas, makes up only about 3.5 percent of the gas molecules. Both planets possess about the same total amount of gaseous nitrogen, but Venus's atmosphere contains some 30,000 times as much carbon dioxide as does the earth's. In fact, the earth does hold a quantity of carbon dioxide comparable to that in the Venusian atmosphere. On the earth, however, the carbon dioxide is locked away in carbonate rocks, not in gaseous form in the air. This crucial distinction is responsible for many of the drastic environmental differences that exist between the two planets.

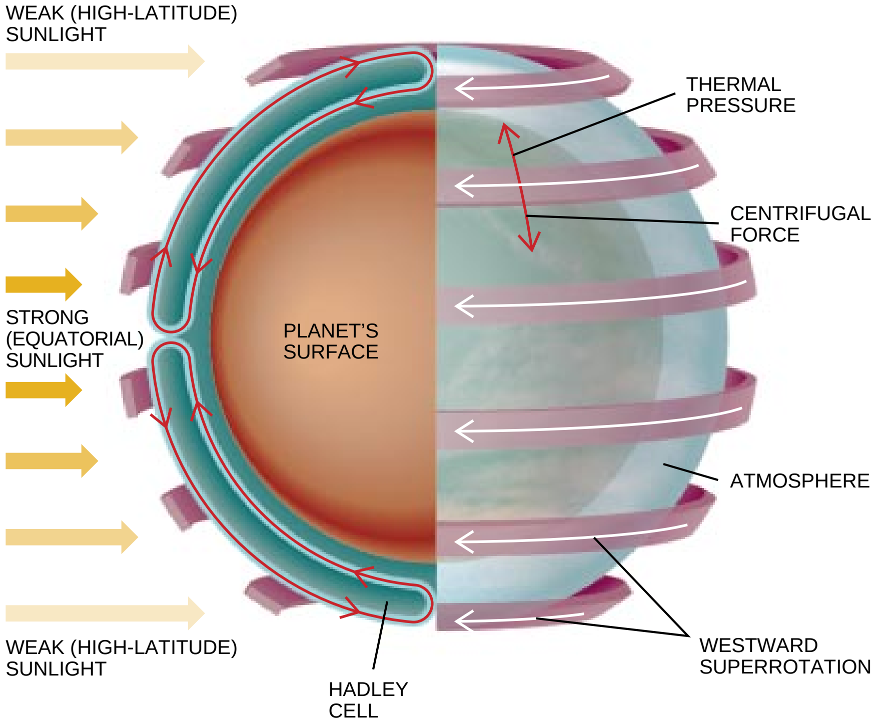
The large *Pioneer Venus* atmospheric probe carried a mass spectrometer and gas chromatograph, devices that measured the exact composition of the atmosphere of Venus. One of the most

The Hot, Poisonous Venusian Atmosphere

The opaque clouds and dense atmosphere of Venus effectively trap and reemit thermal radiation, resulting in a scorching surface temperature of 450 degrees Celsius. Convection acts inefficiently at removing heat from the surface. The clouds are the product of a chemical cycle involving the element sulfur (S). Through the mineral buffering process, sulfur from surface rocks reacts with carbon monoxide (CO) and carbon dioxide (CO_2) in the atmosphere to yield carbonyl sulfide (COS), which in turn interacts with oxygen-rich gases (O) to produce sulfur dioxide (SO_2). In the photochemical zone atop the clouds, sulfur dioxide interacts with water (H_2O) and solar ultraviolet rays, giving rise to sulfuric acid droplets (H_2SO_4). These droplets slowly sink through the condensation zone; along the way, they grow through mutual collisions and through drawing sulfuric acid vapor and water from the air. In the hot regions closer to the ground, the particles vaporize and dissociate into sulfur dioxide and water vapor.







stunning aspects of the Venusian atmosphere is that it is extremely dry. It possesses only a hundred thousandth as much water as the earth has in its oceans. If all of Venus's water could somehow be condensed onto the surface, it would make a global puddle only a couple of centimeters deep.

Unlike the earth, Venus harbors little if any molecular oxygen in its lower atmosphere. The abundant oxygen in the earth's atmosphere is a by-product of photosynthesis by plants; if not for the activity of living things, the earth's atmosphere also would be oxygen poor. The atmosphere of Venus is far richer than the earth's in sulfur-containing gases, primarily sulfur dioxide. On the earth, rain efficiently removes similar sulfur gases from the atmosphere.

Minor constituents of the Venusian atmosphere that were detected by *Pioneer Venus* offer clues about the internal history of the planet. The inert gas argon 40, for instance, is produced by the decay of radioactive potassium 40, which is present in nearly all rocks. As a planet's interior circulates, argon 40 that is trapped in deep rocks works its way to the surface and into the atmosphere, where it accumulates over the eons. *Pioneer Venus* found significantly less argon 40 in Venus's atmosphere than exists in the earth's. That disparity reflects a profound difference in how mass and heat are transported from each planet's interior to its surface. *Magellan* recently found evidence of earlier widespread volcanism on Venus but no signatures of the plate tectonics that keep the earth's surface geologically active and young.

Pioneer Venus revealed other ways in which Venus is more primeval than the earth. Venus's atmosphere contains higher concentrations of inert, or noble, gases—especially neon and other isotopes of argon—that have been present since the time the planets were born. This difference suggests that Venus has held on to a far greater fraction of its earliest atmosphere. Much of the earth's primitive atmosphere may have been stripped away and lost into space when our world was struck by a Mars-size body. Many planetary scientists now think the moon formed out of the cloud of debris that resulted from such a gigantic impact.

Venus's thick, carbon dioxide–dominated atmosphere is directly responsible for the inhospitable conditions on the planet's surface. On an airless body like the moon, the surface temperature depends simply on the balance between the amount of sunlight the ground absorbs and the amount of heat it emits back into space. The presence of an atmosphere complicates the situation. An atmosphere blocks some sunlight from reaching the surface and helps to carry heat upward. But more significantly, the atmospheric gases absorb infrared (thermal) radiation from the ground and reemit it back. The resultant warming of the surface is called the greenhouse effect because the atmosphere functions like a greenhouse: sunlight can get in, but infrared rays cannot get out, causing temperatures to rise.

The intensity of the greenhouse effect depends on how thoroughly the atmospheric gases capture infrared radiation. The principal greenhouse gases on the earth—carbon dioxide and water vapor—absorb complementary parts of the infrared spectrum. Adding more of these gases to the air would, in theory, increase the efficiency of the greenhouse effect, which is why people worry about the climatic impact of carbon dioxide released by human activities. The earth's atmosphere is largely transparent to infrared rays having wavelengths between eight and 13 microns, or millionths of a meter (although ozone, methane, freon and other gases do absorb rays in narrow portions of this band). This open "window" in the atmospheric greenhouse limits the amount of warming that the earth experiences.

Pioneer Venus showed that the greenhouse effect operates much more efficiently on Venus. Data from the four atmospheric probes enabled workers to construct a mathematical model that closely matches the observed temperatures at various altitudes. From that model, it was deduced that carbon dioxide is the most significant greenhouse gas on Venus but that its action is enhanced by the presence of water vapor, clouds, sulfur dioxide and carbon monoxide. The mixture of gases and particles in the Venusian atmosphere blocks thermal radiation at virtually all wavelengths, preventing heat from escaping into space and yielding torrid surface temperatures. These results emphasize the importance of learning more about how human-generated greenhouse gases might affect the terrestrial climate.

Astronomers have long wondered how Venus turned out so hot and dry compared with the earth, especially given that Venus and the earth probably started out with similar overall compositions. According to present theory, the two planets grew by colliding with and absorbing smaller bodies. In the process, each protoplanet would have scattered some smaller bodies into orbits that would have crossed the other protoplanet's path. Hence, the earth and Venus should have accumulated comparable quantities of water-rich bodies even if, at first, water was irregularly distributed through the infant solar system. The roughly equal quantities of carbon dioxide and nitrogen on the two planets support the notion that they once had comparable amounts of water as well.

The young earth and Venus quickly developed thick atmospheres consisting of gases expelled from their interiors and of the vaporized remains of icy, impacting bodies. Water in the earth's atmosphere condensed into lakes and oceans, which proved crucial to the planet's climatic development. Much of the airborne carbon dioxide was quickly sequestered into solid carbonate, a process that occurs through the chemical weathering of rocks in the presence of liquid water.

Venus, too, may have had broad oceans during its youth. The newborn sun was about 30 percent less luminous than it is at present, so temperatures on Venus could have been well below the boiling point of water. (Venus orbits at 0.72 times the earth's distance from the sun.) As the sun brightened, however, surface temperatures on Venus eventually rose above boiling. From then on, any carbon dioxide exhaled by volcanoes or delivered by impacts on Venus could no longer be removed from the atmosphere by chemical weathering. As carbon dioxide accumulated in the atmosphere, the greenhouse effect grew ever more intense. The ultimate result was the sizzling, carbon dioxide–dominated world of today.

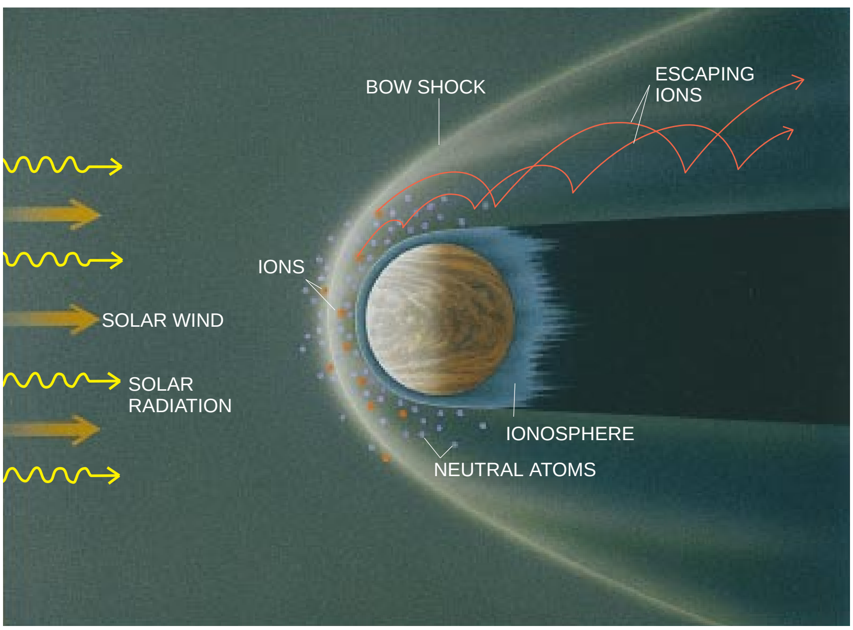
After the oceans boiled, the atmosphere of Venus should have been full of water vapor—in clear contrast to the data. Where has all the water gone?

Pioneer Venus has helped answer that question. The spacecraft documented that even now Venus continues to lose water. Water molecules that wander above the cloud tops react with solar radiation and other molecules. In the process the water molecules split into their oxygen and hydrogen components. The lightweight hydrogen atoms may escape into space by interacting with energetic atoms and ions in the upper atmosphere or with the solar wind, a flow of charged particles that issues from the sun. The leftover oxygen atoms may combine with minerals on the surface, or they, too, may escape by interacting with the solar wind.

A few billion years ago Venus's upper atmosphere contained much more water than it does now; the early sun also emitted far more energetic ultraviolet rays. Both factors greatly hastened the rate at which Venus's water was destroyed and carried off into space. Calculations indicate that over the 4.5-billion-year lifetime of the solar system, Venus could have lost as much water as resides in the earth's oceans.

The earth never experienced such large losses of water because of its moderate surface temperature. Water on the earth stays mostly on the ground or in the lower atmosphere; very little reaches the upper atmosphere, where it may disappear forever. Once the oceans of Venus boiled, in contrast, the planet's atmosphere grew ever hotter, driving more and more water vapor into the upper reaches of the atmosphere.

And yet some water remains. Observations of Venus's upper atmosphere made by the *Pioneer Venus Orbiter* imply that the planet now loses about 5×10^{25} hydrogen atoms and ions each second. At that rate, the entire amount of water in the atmosphere would be gone in about 200 million years. Venus is more than 20 times that old, so some mechanism must replenish the water that Venus is constantly losing. The water most likely derives from a mix of external sources (such as the impact of comets and icy asteroids) and internal ones (through volcanic eruptions or more widespread and steady outgassing to the surface). Because the understanding of Venus's water loss is still quite sketchy, it is possible that *Pioneer Venus* might actually have observed the last trickle from the planet's water-rich early atmosphere.



SOLAR WIND INTERACTS directly with the upper atmosphere of Venus because the planet has no substantial magnetic field. Where the solar wind skirts around the planet, a bow shock forms. Some neutral atoms (*purple*) at the top of the Venusian atmosphere become electrically charged ions (*orange*) that are then carried off in the solar wind. Radiation from the sun also gives rise to a permanent, charged layer, the ionosphere (*blue zone*), which forms at lower altitudes around the planet.

Despite its lack of water, Venus is cloaked in thick clouds that conceal its surface from conventional telescopes. The nature of those clouds has intrigued astronomers for centuries. By the time of the Pioneer Venus mission, planetary scientists had accumulated strong evidence that the clouds were largely composed of concentrated solutions of sulfuric acid and water. They could not determine, however, the source of the sulfur from which the cloud droplets arose.

Pioneer Venus finally settled the question. As the *Orbiter* circled the planet, it scrutinized the tops of the clouds using its ultraviolet spectrometer, which identifies the characteristic pattern of emission and absorption from various atoms and molecules. Also, the gas chromatograph on the large probe measured the composition of the region below the main cloud deck. The results of these studies show that the sulfuric acid in the clouds derives from sulfur dioxide in the atmosphere.

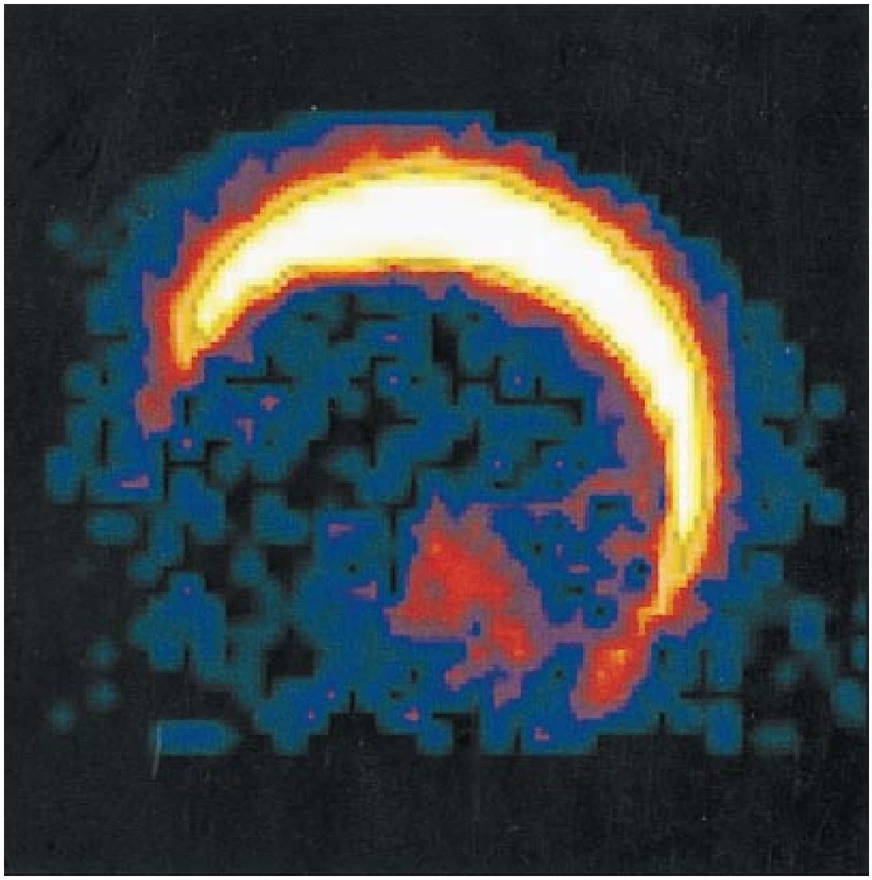
Near the top of the clouds, some 60 to 70 kilometers above the ground, ultraviolet rays from the sun split sulfur dioxide into molecular fragments known as radicals. These radicals undergo a series of chemical reactions with radicals derived from water, ultimately producing tiny droplets of sulfuric acid. Gravity and air currents cause the droplets to migrate downward. As the droplets fall, they grow by colliding with one another and by accumulating sulfuric acid vapor from the air. At and below the base of the clouds, sulfuric acid particles dissociate back into sulfur dioxide and water vapor.

Instruments on the *Pioneer Venus* probes detected tiny particles (less than a thousandth of a millimeter across) at altitudes between 48 and 30 kilometers, just below the base of the cloud deck. Atmospheric motions carry these particles, along with sulfuric acid vapor, to higher, colder altitudes. There the sulfuric acid rapidly condenses onto the particles, producing much larger cloud particles that are concentrated toward the clouds' base. The density of those particles varies from place to place in the lower cloud region, probably because of irregularities in the upwelling and downwelling motions.

A related *Pioneer Venus* observation has stirred great excitement and controversy. In the course of exploring Venus's sulfur chemistry, the *Orbiter* detected an apparent, steady decrease in the concentration of sulfur dioxide near the tops of the clouds. Some workers interpreted that measurement as evidence that a giant volcanic eruption spewed sulfur into the atmosphere just about the time that *Pioneer Venus* arrived—a tantalizing sign that Venus might have active, explosive volcanoes. Once the eruption ceased, the sulfur levels would have begun to drop, as observed. Other investigators have argued that the changes in composition could have resulted from normal variations in the atmospheric circulation. The issue remains frustratingly unsettled.

Although it could not resolve the sulfur dioxide puzzle, *Pioneer Venus* has provided many other intriguing details about the circulation of the Venusian atmosphere. Such information is a tremendous boon for scientists attempting to understand atmospheric dynamics because it shows how weather patterns operate on a planet that differs from the earth in several crucial aspects.

Venus rotates extremely slowly; the earth completes 243 daily rotations in the time it takes Venus to turn once with respect to the stars. Also, because of the dense atmosphere, the surface temperature on Venus is nearly constant, from equator to pole. One might naively assume, therefore, that winds on Venus would be very sluggish.



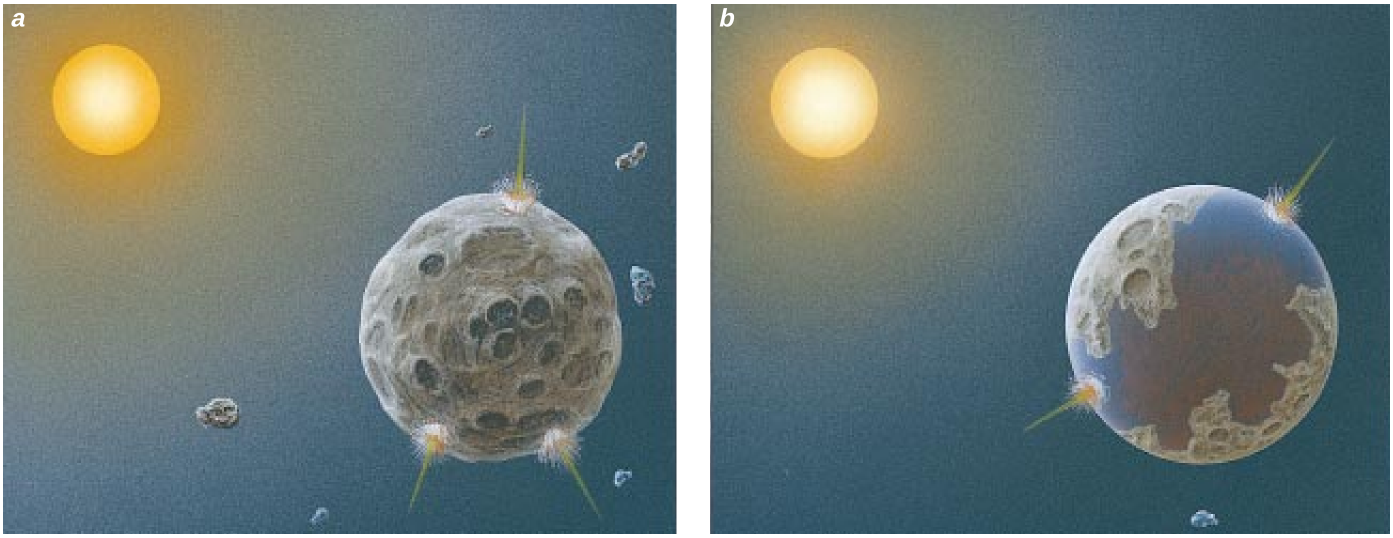
NIGHTSIDE AURORA shows up in this false-color ultraviolet image from the *Pioneer Venus Orbiter*. The bright crescent is the dayside of the planet, which reflects solar ultraviolet rays. The patchy auroras on the nightside occur when energetic particles, possibly from the solar wind, crash into the Venusian atmosphere.

Pioneer Venus proved that assumption false. On the earth, winds at low latitudes move more slowly than the rotation of the planet, whereas those at higher latitudes exceed the speed of the surface, a state known as superrotation. The atmosphere of Venus superrotates at all latitudes and at all heights from close to the surface to at least 90 kilometers above the surface. The winds attain their peak velocity near the cloud tops, where they blow at an unexpectedly rapid 100 meters a second, about 60 times as fast as the rotation of the underlying surface.

Winds in the atmospheres of the earth and the other terrestrial planets are driven by local imbalances between the amount of incoming solar energy and the amount of outgoing, radiated heat. In general, low latitudes, which receive the most sunlight, experience a net heating, whereas high latitudes, which receive the least incident solar energy, undergo a net cooling. As a result, the atmosphere develops a large-scale circulation pattern called a Hadley cell. In this pattern, hot air rises near the equator and travels toward the poles, where it sinks and returns toward the equator.

The spinning of a planet on its axis deflects the north-south (meridional) winds sideways, however, giving rise to east-west, or zonal, winds. Surprisingly, zonal winds almost always end up being much stronger than the north-south winds from which they derived. On the earth, Hadley circulation dominates atmospheric motions at low latitudes. Zonal winds close to the equator move slower than the earth's rotation (and hence are called easterlies); those closer to the poles form superrotating westerlies, culminating in the rapid flow of the jet stream. What is so odd about Venus's zonal winds is that they superrotate at almost all latitudes in the lower atmosphere.

Even now, planetary scientists do not fully understand why the entire lower atmosphere of Venus superrotates. The large fraction of solar energy that is absorbed high in the atmosphere, near the tops of the clouds, probably contributes to the brisk winds. The high-altitude heating of the atmosphere may set up a circulation system that is much less influenced by frictional interaction



EVOLUTIONARY HISTORY of Venus explains its present harsh surface conditions. During the first several hundred million years of its existence (*a*), Venus collected water and other frozen gases from collisions with icy bodies originating in the outer solar system; at the same time, water and other gases in the planet's interior emerged through volcanoes. The early sun may have been dim enough to permit the existence of hot oceans on Venus (*b*). As the sun grew brighter,

with the surface than is the case on the earth. The atmosphere of Venus might therefore be highly susceptible to the formation of eddies that can efficiently transport angular momentum. Such eddies could counteract the ability of the Hadley circulation to prevent superrotation at low latitudes. Cloud images taken by the *Pioneer Venus Orbiter* provide evidence of small-scale, eddylike variations in the winds.

High above the superrotating layers of the Venusian atmosphere lies the ionosphere, an extended zone of electrically charged atoms and molecules, or ions. The ions arise when high-energy ultraviolet rays from the sun knock electrons free from atmospheric gases. Every planet that has a substantial atmosphere possesses an ionosphere, but the one on Venus has a number of unusual traits.

The *Pioneer Venus Orbiter* monitored the passage of radio waves through the ionosphere and, during close approaches to the planet, measured its temperature, density and composition directly. As one might expect, Venus's ionosphere is densest in the center of the dayside hemisphere, near the equator, where the incoming sunlight is most direct. Because of the abundant chemical reactions occurring among the particles, Venus's ionosphere consists primarily of oxygen ions, even though carbon dioxide is the dominant gas at lower levels.

Unlike the earth and most other planets, Venus has no significant global magnetic field, for reasons not fully understood. The absence of a magnetic field significantly affects the structure of Venus's ionosphere. The *Orbiter* detected a weak ionosphere that extends beyond the day-night boundary. This finding was intriguing because, in the darkness, ions and free electrons should quickly recombine into neutral atoms. An instrument on the *Orbiter* found that on Venus ions from the dayside are able to migrate to the nightside. On the earth, the planetary magnetic field in the ionosphere inhibits such horizontal flow.

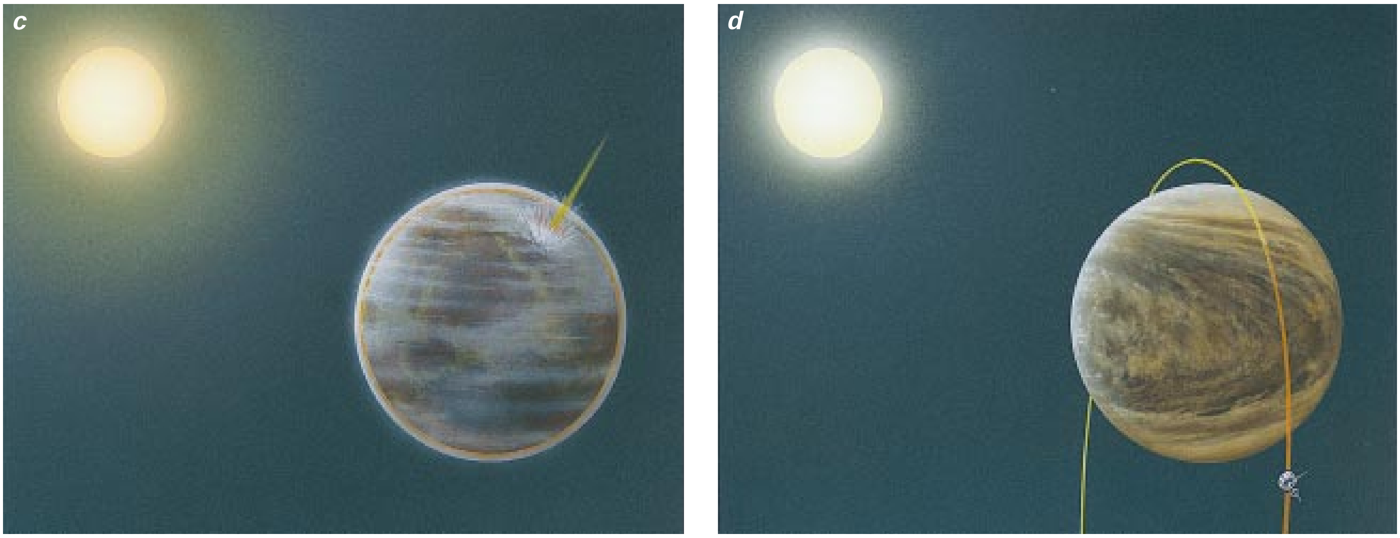
Images of the planet in ultraviolet radiation, obtained using the *Orbiter*'s ultraviolet spectrometer, detected a previously unknown, patchy aurora on Venus's shadowed hemisphere. Scientists attribute the aurora to energetic particles, probably fast-moving electrons, that crash into the atmosphere on the nightside. When these particles strike gas molecules in the atmosphere, they excite and ionize the molecules, further contributing to Venus's nighttime ionosphere. The excited molecules soon return to their normal, low-energy state by emitting radiation, which shows up as the aurora.

As is the case for terrestrial auroras, the particles that cause the auroras on Venus derive their energy from the solar wind. The solar wind is the sun's extended, rarefied outer atmosphere. It consists of plasma, or charged particles (primarily protons and electrons), racing from the sun at supersonic speeds. At the orbit of Venus the solar wind has a density of 15 protons and electrons per cubic centimeter and a velocity of 400 kilometers per second. As the solar wind blows past the planets, it carries part of the sun's magnetic field with it.

The intrinsic magnetic fields around the earth and other planets act as obstacles to the electrically charged solar wind. The wind flows around those fields along a surface (the magnetopause) where the pressure of the wind equals the opposing magnetic pressure. The extent of the deflection depends on the strength of the planetary magnetic field. Venus, which has virtually no field at all, creates an obstacle scarcely bigger than the planet itself.

Nevertheless, the spacecraft found that solar-wind plasma was clearly being diverted around Venus. That discovery confirmed theoretical predictions that a planet's ionosphere can effectively block the solar wind even in the absence of a substantial magnetic field. Like a magnetic field, the ionosphere exerts pressure against the wind, but in this case it is the thermal pressure of the charged gas that counters the force of the solar wind. On average, the balance point lies at a 300-kilometer altitude near Venus's noontime equator and at 800 to 1,000 kilometers above the day-night boundary.

The deflection of the solar-wind flow around large obstacles (such as planets) is preceded by a "bow shock," a sharp boundary closely analogous to the shock that forms in front of a supersonic aircraft. During most of its lifetime, the *Pioneer Venus Orbiter* crossed the bow shock twice each lap, enabling it to monitor continuous changes in the magnetic environment around Venus. The craft found that the bow shock expands and contracts in step with the 11-year cycle of solar activity. The ra-



the oceans boiled away, filling the atmosphere with water vapor and leading to significant greenhouse heating. Water in the upper atmosphere broke down into oxygen and hydrogen, which escaped into space (*c*). In the absence of oceans, carbon dioxide from volcanoes accumulated in the atmosphere, further intensifying the greenhouse effect. Sulfur derived from the surface formed sulfuric acid in the clouds. Thus, Venus acquired its present attributes (*d*).

dius of the shock in the plane of the day-night boundary ranges in size from about 14,500 kilometers at solar maximum to 12,500 kilometers at solar minimum. The expansion and contraction probably result from changes in Venus's upper atmosphere associated with the varying radiation flux from the sun.

Just downstream of the bow shock, the solar wind grows more dense, slows down and changes its direction of flow. Magnetic-field lines are frozen into the solar wind because it is completely ionized. After the solar wind passes through the shock, the frozen-in interplanetary magnetic field piles up.

Pioneer Venus mapped the large-scale magnetic-field geometry around Venus. These data give the impression that the magnetic-field lines eventually slip around the obstacle and into the wake that it creates in the solar wind. Researchers refer to this wake structure as an induced magnetotail because it derives from the interplanetary magnetic field rather than from the planet's own field, as is the case for the earth's own, much larger magnetic tail.

Because of its lack of a significant internal field, Venus interacts more directly with the solar wind than does the earth. Over the age of the solar system, that interaction has affected the atmosphere of Venus. The planet's upper atmosphere, where atomic oxygen predominates, extends well above the point at which the solar wind is diverted around the planet. This gas remains largely unaffected by the solar-wind plasma as long as it remains electrically neutral. If an oxygen atom is struck by an ultraviolet ray or if it collides with a particle in the solar wind, it can become ionized. The oxygen ion couples to the flowing plasma, which may carry it away from the planet and out of the solar system.

Instruments on board the *Pioneer Venus Orbiter* confirm that the solar wind truly does scavenge Venus's upper atmosphere. Measurements of the density of the Venusian ionosphere indicate that the uppermost layers—those above the deduced height of the solar-wind obstacle—appear to be missing. Evidently the ions created above the obstacle have been carried off in the manner just described. The *Orbiter* has also detected the oxygen ions escaping tailward in the solar wind. In essence, *Pioneer Venus* has captured a snapshot of one of the processes by which Venus evolved into a world so unlike the earth.

The extensive archive of data generated by *Pioneer Venus* has proved a wonderful resource for scientists studying the planet's atmosphere and near-space environs. Those data are all the more precious given that no nation has any plans for a follow-up venture. At present, major basic science projects face tight funding prospects in all developed countries. Nevertheless, the intriguing questions raised by *Pioneer Venus* have inspired studies for a possible return to the earth's cloud-enshrouded neighbor.

Even relatively inexpensive missions to Venus could deliver valuable results. A simple chemical composition probe, for example, could elucidate the nature of the atmospheric chemistry at various altitudes. A series of small craft deployed simultaneously all around Venus could yield a sharper picture of the planet's global weather patterns. A specialized orbiter could carry instruments to search for lightning storms and to measure in more detail the ions and atoms that escape from the planet. These scientific goals might be accomplished under the auspices of NASA's upcoming series of fast, low-cost "Discovery-class" planetary expeditions.

A better understanding of the global environment on Venus could be considered a worthwhile goal in itself. It also provides a perspective on the nature of the environment on the earth and on the delicate balance of physical processes that keep our world habitable. Planetary missions such as Pioneer Venus clarify the earth's unique place among the worlds of the solar system. For this reason, and to satisfy the fundamental human drive to explore, we hope such missions continue to receive support in the U.S. and abroad.

FURTHER READING

Special issue on *Pioneer Venus Orbiter* results. *Journal of Geophysical Research,* Vol. 85, No. A13; December 30, 1980.

VENUS. Edited by D. M. Hunten et al. University of Arizona Press, 1983.

THEORY OF PLANETARY ATMOSPHERES: AN INTRODUCTION TO THEIR PHYSICS AND CHEMISTRY. Joseph W. Chamberlain and Donald M. Hunten. Academic Press, 1987.

VENUS AERONOMY. Edited by C. T. Russell. Special issue of *Space Science Reviews,* Vol. 55, Nos. 1–4; January/February 1991.

Nurturing Nature

by Marguerite Holloway, *staff writer*

They used to stretch for hundreds of miles as a tawny sea of saw grass. Metallic-looking plankton added a golden patina to the shallow, slowly moving water that flowed between hammocks of tall grasses and stands of white-barked, high-kneed cypress trees. Even now, at half their original size, the Everglades appear to stretch forever—gilded, green, punctuated by the white of an ibis or a pink roseate spoonbill. Nothing could seem more natural.

Yet the most important aspect of this unique ecosystem is anything but natural. Four great gates at the northern end of Everglades National Park and 1,400 miles of canals and levees determine the quantity of water that can enter the area. Sugarcane plantations and vegetable farms to the north and east use fertilizers and pesticides that determine the quality of that same water. Demands for agriculture, urban living and flood control have made the Everglades too wet in the wet season, too dry in the dry season, too rich in nutrient phosphorus and therefore too close to extinction.

Because control has undone the Everglades, it seems appropriate that chaos be their salvation. Biologists and engineers will try to recreate some of the irregularity of nature by, among other things, delivering water on an erratic schedule, putting uncontrolled, meandering curves into a straight canal and fostering botanic biodiversity rather than biomonotony. Thus, the formerly watery wilderness will be the locus of the largest and most expensive attempt at ecological restoration yet undertaken.

The fact that restoration is being attempted on such a grand scale is testament to the growing status and popularity of ecological restoration, a young field that already carries a heavy burden. Environmentalists, government officials and business managers increasingly perceive restoration as a way to undo ecological damage and to compensate for development. The practice encompasses such diverse efforts as

Can we rebuild it? The field of ecological restoration is evaluating techniques to restore nature and is grappling with definitions of success

removing nonindigenous plants, reintroducing endangered fauna, transforming canals that replaced rivers back into rivers and donning scuba gear to plant sea grass on the ocean floor. Converts have swelled the ranks of the Society for Ecological Restoration to 2,200; when it was established in 1989 the group had a mere 300 members. A peer-reviewed journal, *Restoration Ecology,* was launched last year.

Despite its newfound prominence, restoration remains controversial because it has raised profound and unresolved questions. The idea of restoration seems disarmingly simple at first, but the goals are elusive. If, for instance, scientists want to return an environment to its "natural" state, they need a full understanding of what that is, how the particular ecology is constantly changing and how human beings fit into it. No one has, or is likely to have, such insight. Given, then, that an exact reconstitution is not possible, should researchers—and society—be content with achieving a semblance? Should a restored system be self-sustaining, or should it be managed? Given such uncertainties, how is one to judge success?

These fundamental complexities are further complicated by political strategy and public policy. Some biologists believe the promise of restoration fuels destruction. They argue that such pledges encourage thoughtless development and exploitation: if people believe nature can be rebuilt, there is no harm in losing more of it. Other researchers see restoration as the only possible way society can respond to an already irreversibly impaired environment.

The Everglades will serve as the testing ground

FLORIDA EVERGLADES are drying up because they receive only about one fifth of the water that they did at the turn of the century. This unique ecosystem, which encompasses a saw-grass wilderness (left) *as well as stands of cypress trees* (right), *will be the focus of perhaps the largest and most costly effort to restore an environment that has been attempted in the U.S.*

and battlefield where business leaders, government officials, biologists and the rest of the population address these questions on a vast scale. “The issue is so prominent, no one can afford not to have their name on it,” comments Thomas V. Armentano, acting director of research at Everglades National Park. “If it works, it will be unprecedented.”

Just Add Water

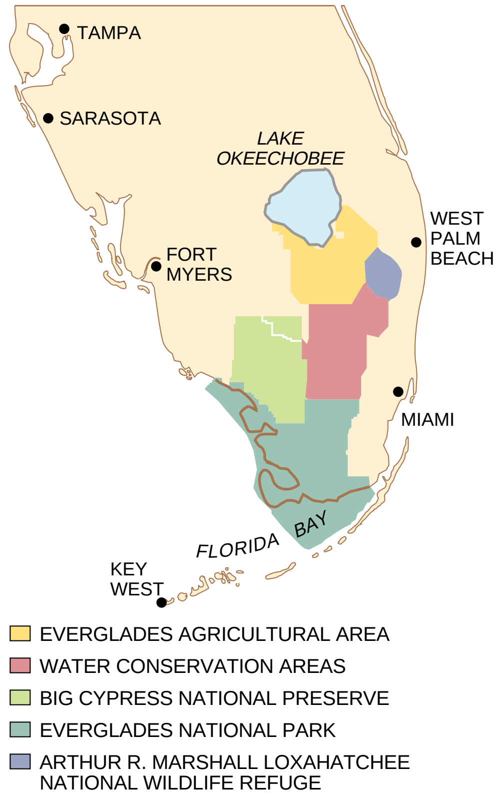
The starting point for restoration in southern Florida is water. As many conservationists working on the matter joke, the solution to the Everglades’ problem is perceived as a *Field of Dreams* theme: if you water it, they will come. This sea-monkey, just-add-water approach derives from the hydrologic history of the region. Efforts at restoration began in the 1980s, when it became evident that the Everglades were drying up. Only one fifth of the water that used to reach the ecosystem at the turn of the century was getting there, often at the wrong times. Only 5 percent of the wading birds that used to nest in the wetlands were still doing so.

Development, which began in the early 1900s as areas were drained for farming in the peat-rich soil, has been, and continues to be, rapid and intense. Southern Florida has one of the fastest-growing populations in the country; domestic water consumption is also high at an average of 123 gallons per person per day. (The national average is 108 gallons per person per day.) The remaining two million acres of the Everglades are contained in the park, Big Cypress National Preserve, the Arthur R. Marshall Loxahatchee National Wildlife Refuge and several water conservation areas. Each site abuts residences or farms.

Most of the chaos that has been introduced into the ecosystem so far has been political, not scientific. In the years since the ecological threats became widely known, nearly every agency and special-interest group has been on one side or another of at least one lawsuit to promote or stall restoration. The largest case was brought in 1988 by the federal government against the state for failing to protect the water supply. Although the parties reached an agreement in 1991, sugar growers opposed it. A new compromise fell through in December 1993. (Ironically, because sugar is federally subsidized at between $1.4 billion and $3 billion a year, it appears that the federal government has paid many of the court costs against itself.)

Despite the haggling about who will pay and the unresolved, baroque questions about which agency will control what, several restoration projects are under way. The South Florida Water Management District, which regulates water use in the watershed and oversees the water conservation areas, has constructed a 3,700-acre wetland to remove nutrients from agricultural water.

This runoff currently drains southeast into the wildlife refuge, where erect stands of cattails along the northern border attest to the influx of phosphorus. Although scientists debate the source of the mineral at some sites, the nutrient clearly fosters the growth of



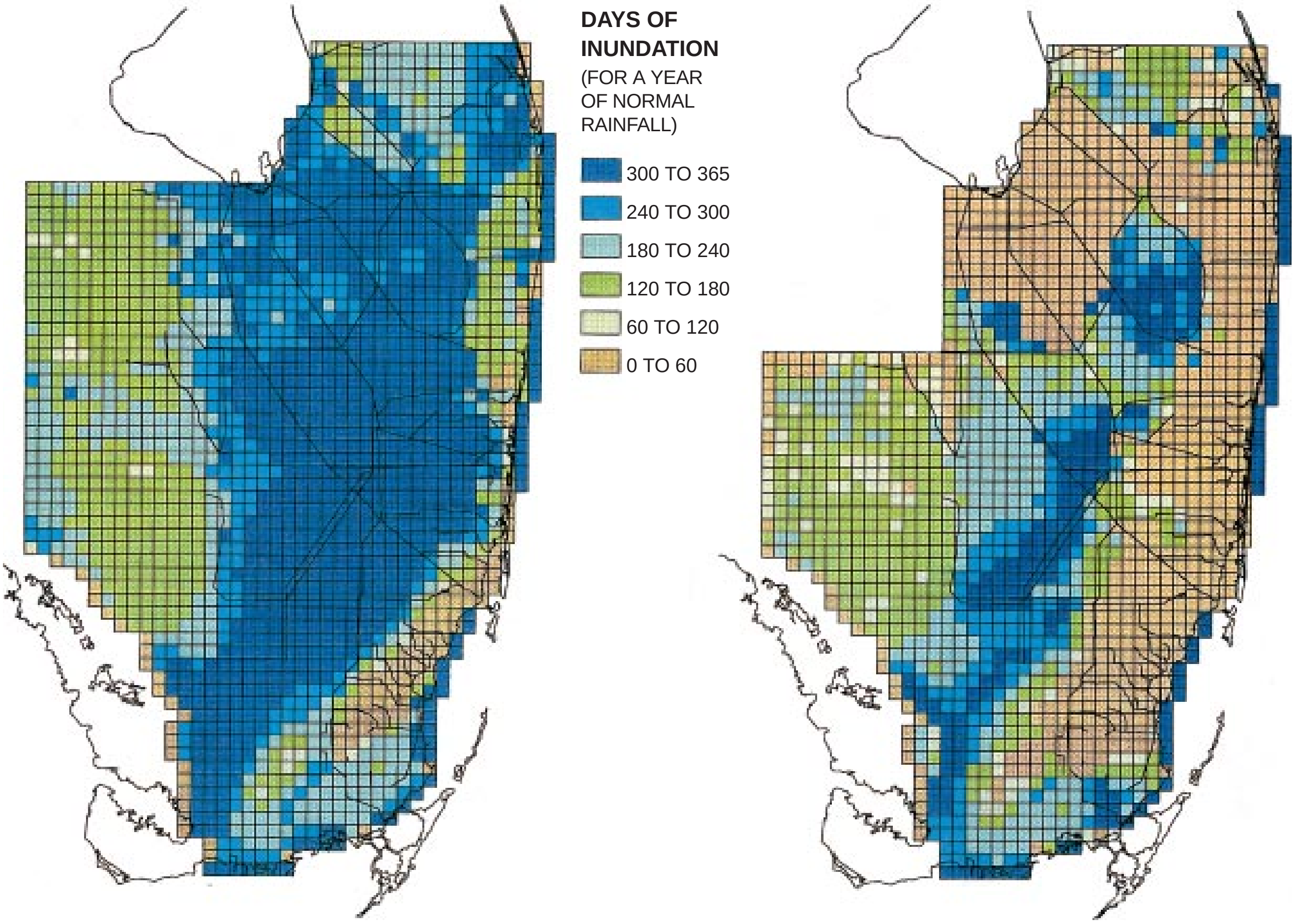
SOURCE: South Florida Water Management District

cattails. Cattails, in turn, replace saw grass and other indigenous plants, reducing floral diversity and habitat for waterfowl. The constructed marsh will turn cattails to advantage, using them to remove 75 percent of the phosphorus from incoming water.

The Everglades Nutrient Removal Project, as the marsh is called, constitutes the largest cleansing wetlands created in the U.S. Cattails and other plants used in the project are commonly employed in smaller-scale managed wetlands to purify water. Although conflicts in the approval process have held up the release of water, it is intended to serve as a demonstration. Similar facilities, covering a total of 40,000 acres, are also being planned to cleanse storm water.

Cleanup is not cheap. The cost of these efforts could reach $465 million or more, and it is this figure that makes many of the large sugar companies and farmers balk: under the most recent proposal, they would have been responsible for a total of between $232 million and $322 million over the next two decades. In January, however, Flo-Sun, Inc., reached an independent settlement and agreed to finance part of the cleanup. Given that many farms may not be around in several decades, the position of the holdouts makes some sense. Because growers keep the soil drier than it would normally be, the peat is being

FOUR WATER-CONTROL GATES, including this one, punctuate the Tamiami Trail, which runs across southern Florida from east to west. The gates regulate the amount of water that flows south into Everglades National Park.



COMPUTER MODELS designed by the South Florida Water Management District will be used to plan water use in an effort to save the Everglades and simultaneously satisfy urban and agricultural needs. The natural system model (*left*) will illustrate how water moved in the region around 1900. The managed system model (*right*) shows how water flows now that the area is crossed by 1,400 miles of canals. The natural system model has not been fine-tuned, but the comparison of the two models in this simulation conveys the magnitude of development in the region and its effects on hydrology.

oxidized and blown away at the rate of anywhere between one and three centimeters a year. Within a lifetime, many farmers may reach the limestone bedrock of southern Florida—at which point soil restoration may become the new focus of the region.

Scientists disagree on the relative importance of removing nutrients from the water or just getting the water back into the Everglades. Water formerly traveled over the area in a sheet that was sometimes as many as 60 miles wide and as much as a foot deep. Now it is ushered through a labyrinthine network of canals, pumping stations, locks and gates. Because the region's hydrology is so disturbed and there are so many demands for water, efforts to fix flow seem even more daunting than those designed to remove pollutants. Despite all the attention directed at the Everglades, the park only just recently received increased amounts of water. Armentano says the strategy is not working: the added water, which is coming into a slough from a canal to the east, is seeping right back out again.

Regional agencies have devised several long-term plans to address the hydrologic problems. Modelers at the South Florida Water Management District are writing computer programs to reconstruct how water traveled before the region was crosshatched by canals and levees. Once finished, this so-called natural system model will be superimposed on an existing model of how water currently flows. Researchers are also working to predict how hydrologic changes will affect vegetation.

Using these tools, experts hope ultimately to devise a politically acceptable and environmentally sound way to divide water in southern Florida, explains Jayantha T. B. Obeysekera, an engineer for the district. Such answers will not be available for a while. Models inevitably reveal the many gaps in data that have to be filled in by fieldwork. Furthermore, the graphic results must be viewed for what they are and with the same careful scrutiny that all the conflicting climatic change models garner. "It is like Disney World: it looks natural, but you go outside, and you realize it is not," Obeysekera cautions.

While modelers are trying to anticipate how the system could be manipulated, field engineers and biologists have been examining the feasibility of restoring the Kissimmee River. The Kissimmee was formerly a 103-mile-long rambling river that flowed south into Lake Okeechobee, which, in turn, sloshed into the Everglades. Together the river and the lake were responsible for the unique hydrology of the region. Like many rivers, the Kissimmee was unruly, inundating the wetlands adjacent to it, overflowing the lake after heavy rains, flooding farms, damaging property. So, between 1962 and 1972, the Army Corps of Engineers straightened it out. The Kissimmee is now a subdued, 56-mile canal called C-38.

Public opinion, however, proved less amenable to control. The army had only just completed its work when several groups, including the Fish and Wildlife Service and the Audubon Society—which had opposed the alignment in the first place—called for restoration. The conservationists, public and private, won their case, and the South Florida Water Management District was required to determine the feasibility of putting the curves back into the Kissimmee and water back onto its floodplains. The resulting study, as well as others, revealed that the loss of 30,000 acres of wetlands on either side of the river had diminished bird and fish habitat and had degraded water quality.

Between 1984 and 1989 the South Florida Water Management District conducted a demonstration project on 12 miles of the river to determine if restoration was indeed possible. “It would be the purest restoration project that I am aware of because it would eliminate any of man’s interference in this area,” yells Louis A. Toth, who directed the project. He and several colleagues are hovering above the project site in a helicopter, recording changes in vegetation that have taken place since they finished monitoring the project several years ago. A line of oaks a mile or so inland sketches the upper limit of the floodplain. The researchers see the succession they expected: in some sites, switchgrass is giving way to wax myrtle; willows are replacing woody species.

Toth is careful to emphasize that the project has not restored the floodplain—rather it shows that there is ecological benefit in trying to restore. As a pair of sandhill cranes dash through a marsh trying to escape the downdraft of the helicopter, Toth remarks that there was a 1,000 percent increase in the number of wading birds after reflooding: “It indicates that the species will return if there is further restoration.” He also explains that it took nearly 20 years to develop a definitive goal for the project—a span that exceeds the lifetime of most restoration efforts and most political administrations.

The full-scale project, slated to begin next year, will bring back 26,500 acres of wetlands. To obviate demands for flood control, the district will also have to purchase farms. In total, restoration of the Kissimmee will cost at least $372 million. “It is the new environmental pork barrel,” comments Daniel E. Willard of Indiana University. “It will cost 100 times as much to put the curves in as it did to take them out.”

Exorcists of Exotics

The last big piece of the restoration work in southern Florida has not received as much press as have pollutants and hydrology. Nevertheless, that element—the removal of nonindigenous plants, sometimes referred to as exotics—has become a pressing concern. According to a recent report by the Office of Technology Assessment, more than 2,000 plant species that came from somewhere else thrive in the U.S. today. Fifteen of them have caused over a half a billion dollars’ worth of damage since 1906. But even committed restorationists are divided on the necessity of destroying them. Some feel any growth in a disturbed area is better than nothing. “In some places, we have phragmites, which are hated by everyone,” Willard says, describing a marsh plant that can grow rife in areas outside its natural habitat. “But black night herons nest in phragmites. Is it a failure if we have them there?”

The exorcists of the exotics counter that native ecosystems and biodiversity

Several Nonindigenous Pest Plants in Southern Florida

MELALEUCA (*Melaleuca quinquenervia*)—This Australian tree in the myrtle family was brought to southern Florida to drain the Everglades in 1906. It is now considered the region’s most serious problem. The rapidly growing tree, which crowds out native vegetation and reduces habitat for animal life, can flower five times a year. It currently covers some 500,000 acres. Annual efforts to control its spread will cost about $1 million by 1995.

WATER HYACINTH (*Eichhornia crassipes*)—A fast-growing aquatic weed, water hyacinth originated in South America. It was introduced into Florida in the 1880s because of its beautiful purple flowers, and by the 1960s it covered 125,000 acres of public lakes and rivers. The plant blocks waterways, interferes with flood control and reduces biodiversity by preventing other species from growing. Possession of water hyacinth can lead to a $500 fine and 60 days in jail. Controlling this exotic weed costs about $3 million a year.

BRAZILIAN PEPPER (*Schinus terebinthifolius*)—This ornamental tree was brought to Florida in 1892 and is now rampant. After Hurricane Andrew, it spread into mangrove stands in Everglades National Park, where it has been difficult to remove. The pepper’s seeds are mildly hallucinogenic, and some biologists report seeing birds so intoxicated on these seeds that they cannot fly.

HYDRILLA (*Hydrilla verticillata*)—This plant from Sri Lanka was introduced in the 1950s as vegetation for aquariums. More than 40 percent of Florida’s public lakes and rivers are infested with hydrilla: a total of some 66,000 acres are covered with the exotic species. Hydrilla crowds out other plants and clogs waterways. It travels easily on boats and can produce millions of underground tubers that lie dormant for years. The state spent $48 million between 1980 and 1989 attempting to control the weed.

AUSTRALIAN PINE (*Casuarina*)—This tree, sometimes referred to as an oak, was brought to Florida in the late 1800s. It is replacing mangroves and other native vegetation, rendering habitats sterile because nothing thrives in its vicinity. It shades plants growing on dunes, thereby killing them and opening the dunes to erosion (the root system of *Casuarina* does not serve to stabilize the dunes).

SOURCES: Dan Thayer, South Florida Water Management District; Greg Jubinsky and Jeff Schardt, Florida Department of Environmental Protection.

cannot truly be restored as long as foreign species are present. "I am very anti–exotic species, even when they are not causing a problem," says Peter White, director of the North Carolina Botanical Garden. "They are the most irreversible of all human effects: we can clean the air, we can clean the water, we can restore wetlands, but exotic species are difficult to get rid of."

In the Everglades the most unwanted aliens are a Brazilian pepper plant and melaleuca, an Australian tree imported to help drain marshes. Both tend to grow in very dense stands, as do cattails, driving out other plants and reducing wildlife habitat. Keeping melaleuca controlled in the Everglades requires vigilance. Thick forests of the trees can be seen running along the canal bordering the eastern boundary; inside the saw-grass fields, hundreds of white corpses of the poisoned trees reveal the scope of the seek-and-destroy mission. In addition, two insect pests are being imported from Australia to control the pines.

At a site called the hole-in-the-doughnut, an even more dramatic campaign is being waged against the Brazilian pepper. Situated in the middle of the park, the hole-in-the-doughnut was farmed until the 1970s. It is now the location of the park's hurricane-decimated research center. The soil the farmers had broken up and tilled for so many years proved to be ideal for the pepper plant. So park biologist Robert F. Doren, who started the Exotic Pest Plant Council 10 years ago, decided to remove the topsoil the farmers had worked. Doren found that the pepper plants did not return to areas from which the soil had been stripped. But what will take their place remains unclear—the results of the succession experiment will become evident only in the next few decades. In that time, $44 million will be invested in clearing 100 acres of topsoil each year for 15 or 20 years, at a cost of $16,500 an acre, while the hole-in-the-doughnut is rehabilitated. "One of the questions is, What do you actually get?" Armentano says.

Whether it is used by scientists to play gardener and weed out exotics or to play God and part the waters, information about the effects of restoration is in great demand in southern Florida—and elsewhere. Knowledge is inadequate in all areas of science, including the interaction between hurricanes, fire and nutrient recycling in the Everglades, the characteristics of the soil and its microbial communities, and the effects of sea-level rise on the ecosystem. In addition, basic components of the hydrology are not understood. What is the role of changed water patterns and drainage on bird nesting and feeding habits? Has reducing the flow of fresh water through the Everglades caused the demise of Florida Bay?

But because there is little precedent for restoration of this magnitude, many scientists are advocating a one-day-at-a-time approach. Such a strategy would allow experts to experiment and revise plans if a particular line of attack did not seem to be working. "We have to take an adaptive approach. We are not going to have the degree of predictability that we want," explains Steven M. Davis of the South Florida Water Management District, an editor of the recent book *Everglades: The Ecosystem and its Restoration.* Davis thinks this concept is difficult for some researchers to accept. Many want to set targets for certain species: restoration is working if, say, the anhinga population increases by *x* percent. "We may have the illusion of control," he notes. "But no matter what we do, we are not going to put the Everglades back to the way they were. There are going to be surprises."

JOY B. ZEDLER monitors a San Diego marsh that was built 10 years ago to compensate for the construction of a nearby highway. The new wetland still does not resemble a natural one.

Indeed, many biologists argue that the flexible approach Davis champions should be the key element of restoration. Its absence is one of the biggest flaws in efforts to compensate for economic development, says Charles A. Simenstad of the University of Washington. Simenstad has done extensive work on the restoration of rivers and fish habitats in the Pacific Northwest. He found that when the failure of some aspect of a plan becomes apparent there is often little chance to correct the mistake. An inappropriate goal, dear to the agency granting the permit, is often the source of constraint.

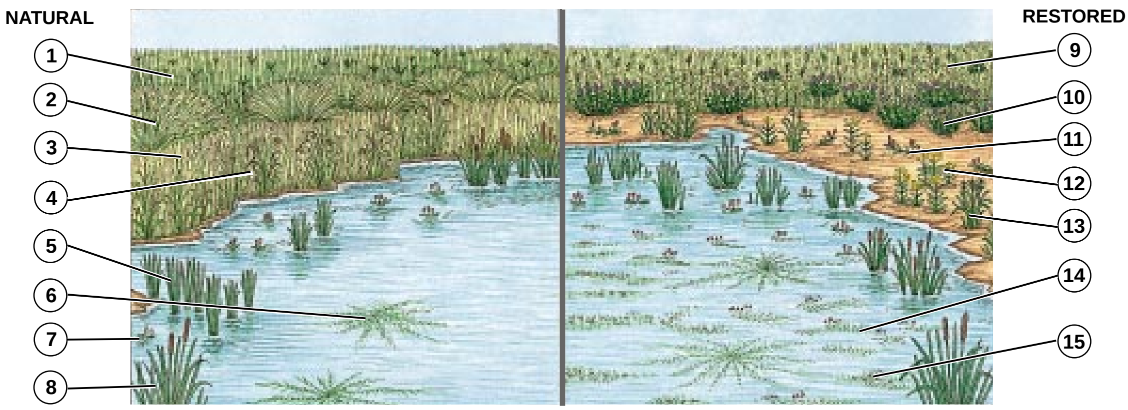
To make matters worse, many of these compensatory efforts do not yield information, because they are not framed in the context of experimentation and are not monitored. In an attempt to make projects more sound and more adaptive, Simenstad is encouraging multistage plans: try a few approaches, wait a few years, see which works best and then follow it.

Even though Simenstad, Davis and many other restorationists are trying to escape what may be a restricting emphasis on precise end points, this focus has arisen precisely because so many restoration projects have not had clearly defined goals. The adaptive strategy may make very good sense in places where engineers or scientists have a reasonable expectation that they can improve the functioning of an ecosystem. In those cases, perhaps rehabilitation is a better word: there is still a clear sense of what the original environment was, and getting back to it seems a somewhat reasonable feat. But a look at many past restoration projects suggests that such rehabilitation is not always possible, particularly where an entire ecosystem is being created. Furthermore, when the incentive for restoration is to compensate for development rather than to redeem an ecosystem, the lack of clearly defined goals for a project can conceal technical failure—or let a developer off the hook.

Joy B. Zedler of the Pacific Estuarine Research Laboratory at San Diego State University has conducted one of the most thorough dissections of one such restoration attempt. Since 1989 she has monitored a salt marsh that was built to mitigate the construction of a high-

Restoration of Prairie Potholes

There is a widely held belief that prairie potholes, many of which were drained for agricultural purposes, are easy to restore: just return the water. But the ecosystems of these ponds, originally created by the surging of glaciers, may not be so amenable to quick rehabilitation, says Susan M. Galatowitsch of the University of Minnesota. The botanist examined records of more than 1,000 restorations conducted by the Fish and Wildlife Service and closely scrutinized 62 of them. She and others found that the longer the pothole had been dry, the less viable the seeds lying dormant in the soil proved to be. Certain types of plants also did not return, and as a result, few birds came back. "There is a whole suite of species that rely on sedge meadows," Galatowitsch notes. "Wrens, yellow-throats, bitterns and rails need dense vegetation. So they have not come in." She says it may be necessary to reseed areas to get the original vegetation to come in again. "The important thing is for people to be constantly improving restoration."



1 BIG BLUESTEM GRASS *(Andropogon gerardi)*
2 CORD GRASS *(Spartina pectinata)*
3 BLUEJOINT GRASS *(Calamagrostis canadensis)*
4 SEDGE *(Carex languinosa)*
5 RIVER BULRUSH *(Scirpus fluviatilis)*
6 GREATER BLADDERWORT *(Utricularia vulgaris)*
7 AMPHIBIOUS SMARTWEED *(Polygonum amphibium)*
8 SILVER CATTAIL *(Typha glauca)*
9 SMOOTH BROME *(Bromus inermis)*
10 ALFALFA *(Medicago sativa)*
11 SMARTWEED *(Polygonum pensylvanicum)*
12 BUR MARIGOLD *(Bidens cernua)*
13 BARNYARD GRASS *(Echinochloa crusgalli)*
14 SAGO PONDWEED *(Potamogeton pectinatus)*
15 LEAFY PONDWEED *(Potamogeton foliosus)*

SOURCES: Susan M. Galatowitsch and Diana Lobien, University ot Minnesota.

way. The project was designed as a habitat for the endangered light-footed clapper rail. Ten years of salt-marsh building later, the rails still have not arrived. "I don't think we are ever going to get functional equivalency for the marsh," Zedler says.

Zedler also determined that for want of a bee, a marsh can be jeopardized. Some plants at the site were not setting enough seed, because pollinators were rare. The insects were not crossing the highway or making their way through urban areas to reach the wetland. "There are millions of pieces in an ecosystem, and we have looked only at a tiny fragment of them," Zedler cautions. "It is not as easy to restore these systems as developers would have us think. When you are trying to improve conditions, I think you can do a lot, but you can never get back what we lost."

This conclusion fits into a national pattern of restoration failures. Mary E. Kentula, director of the Environmental Protection Agency's Wetlands Research Program, and Jon A. Kusler, director of the Association of State Wetland Managers, edited a 1990 report on the status of the science of wetlands restoration. As Kentula notes, they determined that "the efficacy of restoration and creation methods remains uncertain. The technology is unproved for many types of wetlands, and the quality of completed projects is inconsistent."

Kentula also conducted a study of 150 young restoration projects in Oregon. She found that most of the new wetlands were very wet indeed. They were in fact about 90 percent open water. The natural sites they were meant to substitute were only 20 to 22 percent water; the rest was vegetation and wildlife habitat. The reason is that creating ponds is easier and cheaper than ensuring all the species that should be there are there, Kentula explains. But "wetlands are where we have the most experience in restoration, and we see the same mistakes being made over and over again." According to Kentula and Kusler, one of the hardest features of a project to get right is its hydrology.

Correct hydrology is precisely what Zedler ultimately found to be the most crucial missing element in the San Diego marsh. Zedler's observations have led her to be very outspoken about the dangers of mitigation—a position that has earned her the epithet "Joyless Zedler" among some of her colleagues. Zedler confines most of her concern to southern California. This region has lost more than 75 percent of its coastal wetlands—leaving only 31,700 acres of estuarine habitat, of which 18,600 is open water. Ninety-four animal and 187 plant species are endangered or threatened in the state. "California shows that if losses continue, you eventually get to a point where they cannot rebound if there is a catastrophic event," Zedler says. "Species have to come back from somewhere, and there is not enough habitat left so that they can recover."

Zedler and others believe science may not yet be up to the task of ecosystem duplication. The field is so young that it is lacking the quality control that it needs, states John Cairns, Jr., distinguished professor at the Virginia Polytechnic Institute and State University. Cairns chaired the committee that produced a 1992 National Research Council report on restoring aquatic ecosystems. Restoration needs "the kind of control in which ludicrous statements and publications are immediately pounced on and eliminated," he says. "The reason the new journal was founded was because of all the aggravation people in the field go through dealing with reviewers who do not know the existing literature, small as it is."

Other ecologists think these conclusions are unwarranted. They believe the shortcomings of many restoration projects are often perfectly explainable for several reasons. Regulations are one of the culprits. "For those projects that clearly failed, I would take the perspective that it is an agency's failure," argues Dennis M. King of the Center for Environmental and Estuarine Studies at the University of Maryland.

King makes a strong case. The surge in contemporary restoration activity was set in motion by changes in federal wetlands legislation in the 1980s. Before then, restoration in the U.S. had been largely nostalgic, such as the creation of a prairie because this symbol of the American landscape had disappeared, or practical, such as the prevention of erosion in a pit mine. Ten years or so ago mitigation was introduced. It is the process of last resort: compensating for development by creating, restoring or rehabilitating an ecosystem if destruction was unavoidable. If, for instance, a company wanted to build a mall on a marsh, it was unlikely to get a permit to do so from the Army Corps of Engineers unless it made a marsh elsewhere.

Restoration of Mined Land

Mined lands are some of the major sites of restoration efforts, often conducted in an attempt to prevent topsoil erosion. Nine years ago Andre F. Clewell was hired to restore a 3.8-acre forest that had been wiped out by phosphorus mining southeast of Tampa, Fla. The American Cyanamid Company had been required to mitigate—that is, to compensate for the destruction wrought by mining the area (*top*). Clewell, who runs A. F. Clewell, Inc., in Quincy, Fla., says the Hall Branch Restoration Project has turned out to be one of his most successful. Because adjacent forests had not been disturbed, the botanist had good data for how the site was originally—something he needed because "the regulatory agencies had those presettlement forests as stars in their eyes." The most crucial aspect of the work was getting the elevation of the land right, which, in turn, ensured proper hydrology. "A matter of a few centimeters in elevation could cause drastic changes in the project," Clewell explains. After re-creating the original topography, he was able to foster the growth of reforested wetland, including the characteristic trees, shrubs and herbs (*bottom*). "We do not have to do any more to the system," Clewell notes. "We just have to make sure that the land is protected."

A Mall for a Mud Puddle

In 1981 President Ronald Reagan's Task Force on Regulatory Relief pushed the Army Corps of Engineers to speed up the approval process. According to a 1990 study by William L. Kruczynski, then at the Environmental Protection Agency, the new rules limited the power of the EPA, the Fish and Wildlife Service and the National Marine Fisheries Service to review permits. The three agencies, charged with commenting on the environmental impact of development, increasingly recommended mitigation because it was clear that they could not influence the army's judgments.

But there were no mechanisms in place to monitor compliance. "The regulations had not been enforcing quality control," King explains. "And the market has been for low-cost permits, not high-quality restoration." One of King's favorite examples is a developer who chose a site on which to construct a wetland after receiving his permit to build a shopping center. He hit granite one foot below the surface at the site of the proposed marsh. Rather than go back through the permit process, the contractor blasted the granite and built a mud puddle that King says has no ecological value. The cost of the "restoration" was $1.5 million an acre.

With the right amount of money and follow-up, King says, anything is possible. "You see all those statistics on failure rates, and you talk to all the scientists and scratch below the surface and ask them why they failed, and they always know," he argues. "It is not a failure of science. The institutions are not holding the scientists' feet to the fire."

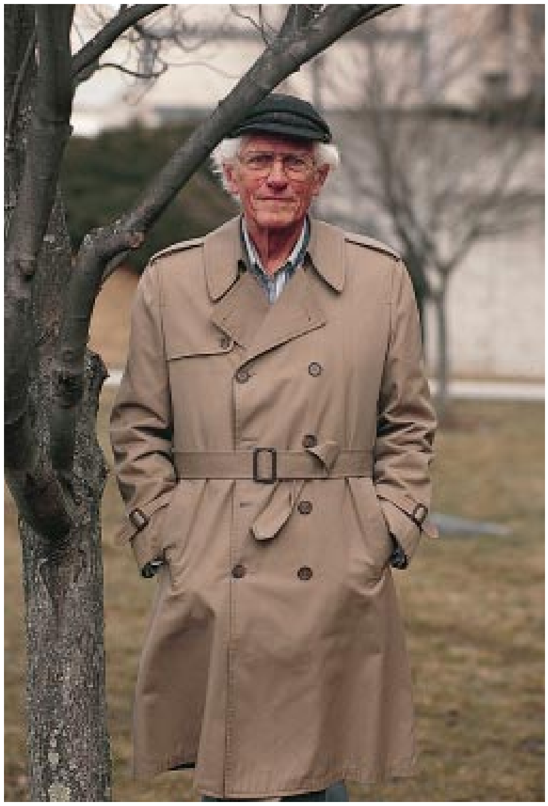
Mark S. Fonseca, a research ecologist at the National Marine Fisheries Service, agrees. He has been working on the restoration of sea-grass beds on the eastern seaboard for more than a decade. "The technical ability is there," Fonseca says. "It is more a problem of the scale of the losses." Fonseca explains that to be successful, restoration has to take place within the context of the preservation or rehabilitation of an entire ecosystem. He notes that he can restore an entire bay, even replant the entire seafloor with sea grass, but if the water column is not cleaned up and pollution persists, the grass will die all over again.

The importance of the Everglades plan derives from the fact that this vast wetlands provides such a context. The region is one focus of the attention of the National Biological Survey (NBS), a new agency established within the Department of the Interior to inventory all animal and plants as well as to study and identify areas at risk. The NBS will attempt to consider entire ecosystems rather than individual species, such as the northern spotted owl.

Because every gallon of water that flows through the Everglades is tracked by the South Florida Water Management District or the Army Corps of Engineers, the bureaucratic structure needed to monitor the region as a whole appears to be already in place. "Kusler and Kentula were looking at much smaller systems," Davis says. "You can set up end points when you have smaller areas. You have captive ecosystems: they are so highly managed they are like circuses. But the Everglades are huge. They will be less predictable in their responses."

The NBS approach, which embodies the prevailing ethic of sustainable development, resembles the perspective from which landscape architects and some western Europeans have traditionally viewed restoration. "They are used to looking at much larger systems than most ecologists are," Cairns says. "Most ecologists in the American and British mold try to find places where humans are not part of the system. The Europeans assumed that humans are always in the landscape. They are not running off to the Galápagos to find some untouched ecosystem."

Achieving that kind of perspective at the level of the evolving science is also crucial. At the moment, the field appears to be missing the very quality that ecologists are demanding from restoration projects: integration. The discipline is by its nature broadly based: good restoration includes information on soil, microbes, botany, hydrology and population ecology, to name a few aspects of any ecosystem. But specialists rarely come together because of what Cairns describes as tribal language and an intellectual electric fence. At meetings, some scientists "want to talk about the number of bugs they had after using a certain fertilizer. It is almost like gardeners talking to each other," King laments. "The perspective has got to widen."



JOHN CAIRNS, JR., chair of a National Research Council report on the restoration of aquatic ecosystems, is outspoken about the need for quality control and synthesis in the young field.

It is also clear that a degree of snobbishness exists on the part of people dealing with landscape-based projects toward those working on fragments of ecosystems. But doing one without taking the other into consideration may be useless, as the death of the pines in southern Florida attests. Barren pines, some with dead needles swinging from nearly naked branches, protrude from plush undergrowth on a 3.5-acre plot in Miami. Until Hurricane Andrew ripped through in 1992, this fragment was one of the last remnants of a coastal forest that covered southern Florida decades ago. Every tree on the site is now dead, as are most of the small pine forests in the region—only an 11,000-acre stand in the park remains.

No one knows what happened to the pines. Was it the force of the hurricane? Had the size of the forests made them vulnerable? Had municipal changes in water flow diminished the amount of sap the trees usually produced to push insect pests out of boreholes? "Everyone is watching the park, waiting to see what happens there," says George D. Gann-Matzen, president of EcoHorizons, a restoration firm, as he gazes at the dead trees. He was hired to remove an exotic Asian vine that had been smothering the pines, camouflaging them as great green druids, and to manage the tiny woods as if it were still wild, when the hurricane struck. Absent trees, Gann-Matzen is unsure whether to burn or plant seedlings.

Gann-Matzen's restoration and management problems appear isolated. But without this forest fragment in downtown Miami, the pines in Everglades National Park, just to the south of the city, could be in jeopardy. Ten of some 26 species of birds that inhabited the park's forest have not been seen in the past several years, or longer. And biologists speculate that the disappearance of the last few woods destroyed an arboreal corridor that these birds traveled southward along, down into the park.

So few large tracts of untouched land are left that restoration must work in tandem with preservation. Areas of conservation are needed to ensure that species can survive, particularly as the human population explodes and the pressures that regions such as southern Florida experience increase. There seems to be tacit agreement that most of nature is hardly natural anymore—the orchestration of water in the Everglades is but one example. Ultimately, the state of the science does not matter to the majority of restorationists. Their attitude is just do it, do something.

FURTHER READING

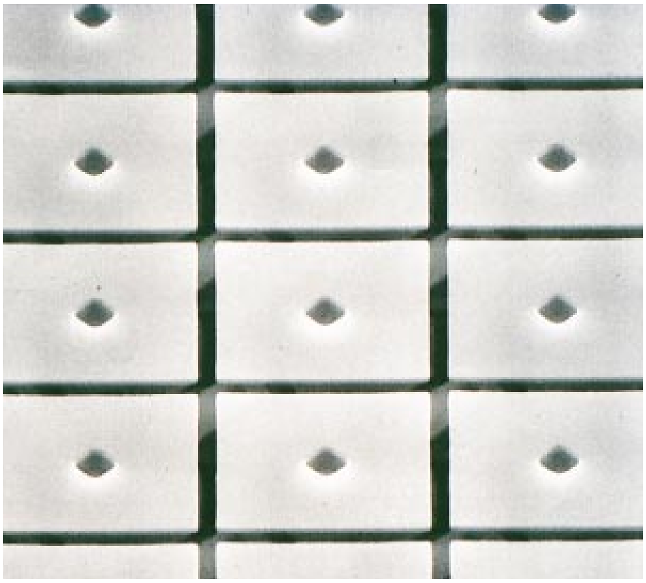
THE EVERGLADES: RIVER OF GRASS. Marjory Stoneman Douglas. Pineapple Press, Sarasota, 1988.

THE CONTROL OF NATURE. John McPhee. Farrar Straus Giroux, 1989.

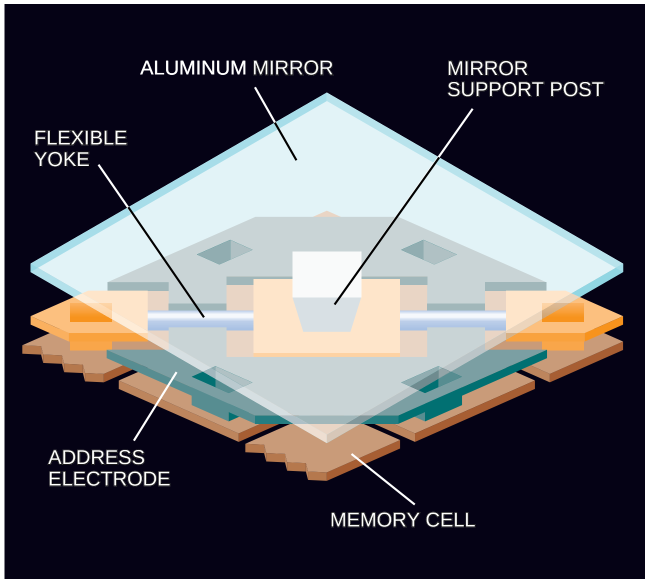
ENVIRONMENTAL RESTORATION: SCIENCE AND STRATEGIES FOR RESTORING THE EARTH. Edited by John J. Berger. Island Press, 1990.

WETLAND CREATION AND RESTORATION: THE STATUS OF THE SCIENCE. Edited by Jon A. Kusler and Mary E. Kentula. Island Press, 1990.

RESTORATION OF AQUATIC ECOSYSTEMS: SCIENCE, TECHNOLOGY, AND PUBLIC POLICY. National Research Council. National Academy Press, 1992.



TEXAS INSTRUMENTS



JARED SCHNEIDMAN *JSD*

MICROSCOPIC MIRRORS developed by Texas Instruments can project high-definition television images when assembled by the millions on top of a microchip* (left). *Each 16-micron-wide mirror* (right) *is attached to a flexible yoke that allows it to tilt 10 degrees to each side. Integrated circuits drive the mirrors to reflect pulses of colored light onto or away from the screen.

Mirror, Mirror

Micromechanical chips challenge tubes for large-screen television

When the Advanced Research Projects Agency (ARPA) offered five years ago to help fund industrial research on promising techniques for making larger, sharper and more affordable digital televisions, it was surprised by a grant proposal submitted by Texas Instruments. Other electronics companies had suggested using liquid crystals, plasma channels or miniature cathode-ray tubes to paint moving images on a screen. The Dallas-based semiconductor manufacturer offered to do it all with mirrors—2.3 million of them hung from hinges atop a microchip roughly the size of a commemorative postage stamp.

"That proposal was originally rejected by ARPA's review team as a totally ridiculous idea," recalls Gary A. Feather, marketing manager of TI's digital imaging division. But after a day's reflection on its mission, Feather says, ARPA relented and awarded Texas Instruments $10.8 million to build a prototype by 1994.

To the astonishment of some, TI claims it is now ready to deliver. In February the company demonstrated a high-definition television (HDTV) system based on three digital micromirror devices (DMDs), one for each additive primary color: red, green and blue. The company has already built less expensive systems that use just a single DMD to project standard-resolution video images onto screens measuring up to 16 feet on the diagonal.

Larry J. Hornbeck, a scientist at Texas Instruments, built the first simple micromirror device in 1987, using the same photolithographic equipment with which TI produces computer memory chips. The DMD is in fact itself just a memory chip with a few extra layers [*see diagram*]. Rising on a hinged post above the electrodes of each memory cell is a thin aluminum mirror, 16 microns on a side. (More than 400 such micromirrors could fit comfortably inside the period at the end of this sentence.) When a positive voltage is applied to one electrode (representing a binary "1"), the negatively charged mirror tilts 10 degrees to make contact, reflecting an incident light beam onto the screen. When the memory cell contains a "0," the mirror tilts 10 degrees in the opposite direction, scattering light away from the screen. Only when the DMD is turned off do the mirrors lie flat. Each mirror can move independently to switch a single pixel on or off up to 100,000 times per second.

Animated by a digital television signal, the 2.3 million mirrors can fill the entire 1,152 columns and 2,048 rows of the largest HDTV screen with pulses of light that blend into a vivid, moving image. The trick relies on several illusions, all of which work because the hinge is quicker than the eye. The illusion of motion is accomplished by changing the picture 60 times a second, just as conventional tube televisions do. DMDs redraw the whole image; standard TVs only every other line.

Producing millions of colors from a single white light requires two other illusions. A spinning red, green and blue filter in front of the lamp isolates each color once per frame. As colored light hits the DMD, each mirror flashes on for a period proportional to the brightness required for that color. To the relatively sluggish human eye, the sequential flashes of primary colors meld into an apparently full spectrum.

Because the DMD projects an image rather than drawing it directly on the screen, Feather believes it would be well suited only to large-screen displays of 40 inches or more. That market, which is currently dominated by rear-projection cathode-ray terminals (CRTs), generated $1.4 billion in U.S. sales last year, according to David E. Mentley, who tracks the display industry for Stanford Resources in San Jose, Calif. Now that an HDTV standard has been approved and high-resolution programming is emerging, he expects that demand for larger sets will accelerate beyond its current 10 percent annual growth rate.

Which technology will seize the day? Although liquid-crystal and plasma display panels have been built, they are still too expensive and power-hungry to produce in large sizes, Mentley argues. In the near term, large HDTVs will be lit by projectors using CRTs, small liquid-crystal displays—and perhaps micromirrors.

Of those three, micromirrors are the dark horse. Feather emphasizes that DMD production can begin soon and with little investment of capital. TI has already produced tens of thousands of the devices for research purposes, he adds. “We are now trying to complete the design of an initial production device that we can use for reliability and life-cycle testing.”

That will be a critical stage for a system that has millions of moving parts and is radically different from anything used in televisions before. While Feather asserts that demonstration DMDs have been run through “about 500 billion cycles” without a single mirror failure, he concedes the need for more tests to answer television manufacturers’ many questions.

Given the exquisite price sensitivity of the consumer electronics market, production cost will be high on their list. Mentley estimates that TI’s micromirrors “could be as cheap as $20 per chip. Even with the cost of the associated electronics and optics,” he figures, “that compares favorably to a CRT.” At this stage, of course, no technology is a safe bet, and TI is hedging with investments in plasma-driver technology and field-emission devices that use tiny cathode-ray emitters. Still, on reflection, the micromirrors may be the fairest of them all. *—W. Wayt Gibbs*

Shaking Quakes

Antiquake building methods begin to emerge from the lab

The January 17 earthquake that struck near Los Angeles took 61 lives and caused an estimated $20 billion in damage. And that was not even the much-feared “Big One.” Each time a major quake hits, it again raises the question, what can be done to protect structures—and hence human lives—from such future disasters?

Over the past decade, a series of innovations in materials science, engineering and electronics has opened the possibility of erecting buildings that shrug off the effects of a quake. One of the most obvious ways to keep a building stable is to make it respond as a human would, shifting its weight to maintain its balance. This “adaptive structure” concept forms the basis for a technique known as active mass damping. In this approach, a heavy mass mounted to the top of the building is attached to computer-controlled hydraulic actuators. Motion sensors in the building send signals to the computer, which in turn directs the actuators to move the rooftop weight so as to counteract the building’s motion.

In 1989 the giant Japanese construction company Kajima built the first building to use active mass damping. Since then, about a dozen buildings, all of them in Japan, have incorporated similar systems, and many more are on the way. The mass-damping system enables tall buildings to isolate their occupants from nausea-inducing motions caused by high winds. None of these buildings rely on active-control devices as a front-line defense against strong earthquakes, although the mass-damping system could in principle be modified to deal with them.

An alternative active technology for keeping buildings stable draws on another aspect of biological systems. It consists of actuators that pull on tendonlike bracings, thereby neutralizing wind- or earthquake-driven motions. Five years ago Tsu T. Soong and Andrei M. Reinhorn of the National Center for Earthquake Engineering Research (located at the State University of New York at Buffalo) began a collaboration with Takenaka Corporation in Japan to build a test structure. Tendon systems are particularly appropriate for being retrofitted onto weak or damaged structures, Soong reports.

Sami F. Masri and his colleagues at the University of Southern California are pursuing another approach that enlists chaos to fight tremors. When an earthquake begins to rattle a building, jet thrusters mounted on the roof would fire blasts of compressed air to trigger brief, chaotic motions. The chaotic forces would smear out the most damaging, harmonic oscillations of the buildings—the ones that cause catastrophic collapse. Tests in Masri’s laboratory convince him that “it’s a very effective way of actually putting chaos to work.”

Researchers are also calling on so-called smart materials—for example, substances whose mechanical properties are quickly and easily adjustable—to create more robust structures. A number of workers, including Robert

LES STONE *Sygma*

SUBSTANTIAL DAMAGE resulted from the recent, moderate-strength earthquake in Los Angeles. Systems that actively respond to ground motions, which are slowly moving from the lab to the real world, could ameliorate the toll from future tremors.

Hanson's group at the University of Michigan, are exploring the engineering possibilities of electrorheological fluids, which transform from liquid to jelly when exposed to electric current. Such fluids could be incorporated into the base of a building. When an earthquake strikes, the fluids would filter out the most damaging, high-frequency shocks, Hanson explains. A computer would then automatically electrify the fluid and stiffen it so that the structure could pass the energy from the low-frequency motions back into the ground. Such a system would have modest electricity demands, and so it could be powered by inexpensive storage batteries.

Smart materials may find their way into a variety of active designs. "There are many ways to control the motion of a building," observes Craig A. Rogers, who directs the Center for Intelligent Material Systems and Structures at Virginia Polytechnic Institute and State University. He is bullish on the electrostatic actuator, which he describes as "just a piece of aluminum foil and some Mylar." Through the selective application of an electric current, he notes, such a simple device can become a powerful muscle that could help hold an edifice steady during a quake.

Rogers and his colleagues are also exploring the potential of piezoelectric materials, which contract when exposed to electricity, and shape-memory alloys, which return to their original shape when heated, to help counteract the motions and stresses that buildings, bridges and other structures undergo during a seismic event. He views smart materials as "a big toolbox that will be available to us."

Despite the many ideas being explored about how to design better, safer and more reliable structures, concepts such as active controls and smart materials are not exactly flying out of the laboratory and into real buildings—at least not in the U.S. "It is embarrassing that in the U.S. there is not a single actively controlled building," Masri says. Japanese companies have shown that such buildings can be constructed within commercial economic constraints, so low-level motion compensation "is no longer even a research issue."

Indeed, both in government and in academia, nearly everyone associated with earthquake-resistant design laments the unwillingness of the American construction industry to adopt, much less instigate, innovation. "Civil engineers in this country are very conservative," says Shih-Chi Liu, who directs the National Science Foundation's five-year research initiative to explore the possibilities of active-control technologies. He notes that in Japan, unlike in the U.S., construction companies run their own active research and development programs.

Liu also decries the "lengthy and frustrating process" of changing building codes to recognize innovative technologies. He points to the example of base-isolation techniques—in essence, large rubber mats that sit under buildings and protect them from the full brunt of the shaking. The NSF supported base-isolation technology 20 years ago, he notes, but it has become an accepted part of American building codes only within the past two years or so.

Soong applauds the money and research coordination coming from the NSF. "The interest has increased tremendously since the program started. I would hope the L.A. quake will spur more activity," he says. Masri and George W. Housner of the California Institute of Technology are coordinating a pioneering, joint U.S.-Japanese conference on active structural control next summer, intended to expand awareness of the new engineering approaches among American industry.

But low construction R&D spending, sluggish building-code changes and the inevitable specter of lawsuits seem destined to keep the U.S. behind Japan and perhaps even other countries, such as China, that are jumping on the active-control bandwagon. "Unfortunately," Masri observes sadly, "it takes disasters like the L.A. quake to remind people of what technology can do to mitigate damage." *—Corey S. Powell*

The Threat of Buried Thrust Faults

For most of this century, Californians have kept a fearful eye on the San Andreas fault, which runs like a surgical scar down the length of the state. The San Andreas is a slip-strike fault, a vertical fissure between two tectonic plates moving in opposite directions. Geologists predict that someday this seam dividing the North American and Pacific plates will rupture and trigger a quake measuring eight or more on the Richter scale.

That "Big One" may yet occur. But the 6.6-magnitude temblor that jolted Los Angeles this past January has shown that "the greatest threat" to that region of southern California may be not the San Andreas but so-called hidden thrust faults, according to Ross S. Stein of the U.S. Geological Survey in Menlo Park. He points out that unlike the San Andreas, which passes 50 kilometers northeast of Los Angeles, these thrust faults lie directly underneath the city and may therefore do more damage.

Hidden thrust faults occur when a highly compressed plate shears deep underground so that one section heaves on top of the other; the faults are called hidden because unlike most slip-strike faults they do not extend to the earth's surface, where they can easily be detected and monitored.

Few geologists realized the danger posed by such faults before 1987, when a 6.1-magnitude quake struck eastern Los Angeles. Workers led by Thomas L. Davis, a geologic consultant, eventually traced the quake to a previously undetected thrust fault below the suburb of Whittier. Davis argued that the wavelike folds characterizing the region formed not gradually—as most geologists had thought—but through sudden, earthquake-triggering slips along thrust faults. That theory is now "completely accepted," says Lucile M. Jones of the USGS's Pasadena office.

Davis has now pinpointed the thrust fault system, buried 20 kilometers below the San Fernando Valley, responsible for the more recent disaster. He estimates that the fault slipped two meters during the recent quake and that it has slipped a total of 14.4 kilometers over the past two to three million years, with quakes occurring every 400 years or so on the average. Davis draws his conclusions largely from data gathered by oil companies, including seismic soundings and cores from oil wells. Conventional fault-detection methods, he says, which include digging trenches across surface faults, are inadequate for studying hidden thrust faults.

Workers at the Southern California Earthquake Center—established three years ago by the National Science Foundation in part to identify and assess the potential risk from hidden thrust faults—are desperately trying to obtain more information from oil companies. Thomas L. Henyey, executive director of the center, notes that even though most oil companies no longer prospect in the Los Angeles area, they have still shown reluctance to share their geologic information. After the recent devastating quake, Henyey says, "I think we will see a new effort to get these data." *—John Horgan*

Basic Strategies

Japanese companies cultivate research labs sown in the U.S.

Back in 1989, when Japan's industrial giants were flush with cash from an overheated domestic economy and a burgeoning trade surplus with the West, many began taking their research operations global. By 1991 corporations such as Mitsubishi, Ricoh and NEC had opened nearly a dozen basic research laboratories in the U.S. Along with similar facilities strategically placed near leading universities in Europe, the satellite labs clustered around Princeton, N.J., and Silicon Valley enabled their Japanese parent companies to mine the rich lode of scientific talent on the two Atlantic continents.

Small, independent and focused on technologies a decade or more from commercial use, the laboratories in the U.S. were bold departures from traditional Japanese research institutions. Three years later, as Japan lingers deep in recession and their benefactors look for expenses to cut, the young organizations are counting on innovative management strategies to help them meet their owners' expectations.

Those goals range from visionary to simply pragmatic. "Before I took over as president [of the NEC Research Institute]," recalls C. William Gear, formerly head of the computer science department at the University of Illinois, "I asked the president of NEC what his expectations were for the next five years. He said we would be successful when we were doing Nobel-quality research."

Such prestige may be desirable for its own sake, but it has short-term business value as well. "We also recognized the criticism that Japanese companies were simply picking up ideas conceived elsewhere and commercializing them," admits Tohei Nitta, president of Mitsubishi Electric Research Laboratories (MERL) in Cambridge, Mass. Given the trade friction between Japan and the U.S., says David T. Methé, an assistant professor of corporate strategy at the University of Michigan, "the Japanese fear that they could be closed out if they are not seen as also contributing."

The U.S.-based labs seem to have helped soothe concerns over Japanese predation of research. "After all," Methé points out, "Americans do this, too. AT&T has technology listening posts in Japan, and IBM, Intel, Texas Instruments, Du Pont and Dow have all set up research facilities there."

The difference is that Japanese firms hope to acquire basic research skills as well as marketable results. "For a long time, there was no need for Japanese companies to do fundamental science," Methé concludes. "Economically, it made better sense to import it." But as U.S. software, networking and biotechnology companies have widened their technological lead, Japanese corporations can no longer afford to be the second to learn of important scientific advances in these areas.

Balanced against that vision, a desire for more tangible returns from scientific investment is reflected in the choice of managers such as Laszlo A. Belady, chairman and technical director of MERL. As head of software engineering first at the IBM Thomas J. Watson Research Center and then at Microelectronics and Computer Technology Corporation (MCC), Belady has seen how short-term thinking and inattention to technology transfer can run a basic research lab afoul of its sponsors.

"We should not try to mimic an academic department," Belady admonishes. "I saw many efforts at IBM and MCC go down the drain because they tried to compete with universities." Yet MERL must compete with academia in one vital area: recruiting. Virtually all of MERL's 21 investigators were wooed from universities. That goes as well for the 40 Ph.D.'s at the NEC Research Institute and the staff of 22 at Canon's scientific center in Palo Alto, Calif. "Researchers are seriously concerned about the environment in which they work," Nitta says. "Openness, for example, is very important to them."

And so, the heads of each of these organizations emphasize, their labs are open. "We publish practically everything we do," boasts Peter E. Hart, who runs the Ricoh California Research Center in Menlo Park. "Last year we had about 700 technical visitors—half of them from universities," Nitta says. Preventing such a free flow of information from washing out the commercial value of these labs to their parent companies requires delicate management.

First, Belady says, the disciplinary boundaries must be torn down. More important, he adds, "is establishing contact with engineers back at Mitsubishi even before the research begins. Because if we wait until we have a prototype, it is too late"—the competitive edge will have been lost. NEC more bluntly reserves the right to sit on the results of its American scientists' work for five months before allowing them to be published.

The need to balance researchers' desire to publish their work with their employers' desire to use that work to competitive advantage may explain why so many of the Japanese laboratories in the U.S. concentrate on software. "It is easy to be open about software," Nitta remarks. Publishing the results and even ideas behind a program does not automatically allow a competitor to recreate it. A more compelling reason, suggests Ronald B. Richard, who oversees Matsushita's American labs, is that "computer software is a field in which the U.S. is so far ahead. You can get better scientists here."

At NEC, the most academic of the Japanese-American labs, one team is trying to discover fundamental limitations of parallel computing. Another is adapting linguist Noam Chomsky's ideas of a shared biological basis for all human languages to automatic translation software. Karvel K. Thornber took a break from physics to devise "analogic," a new system of logic that finally extends to many-valued logics (such as fuzzy logic) powerful rules of inference, some of which have no counterpart in classical logic.

MERL aims for a narrower mission. "In the future the entire workforce has to be continuously learning new tricks," Belady explains. "We are building a computing infrastructure on which we can investigate how people can exploit multimedia to learn and work together collaboratively." To this end, some MERL researchers are designing networks that can transmit real-time video and sound without slowing normal data communications. Others are writing software to enable computers connected to a network to divide a problem and conquer it together as a "virtual supercomputer."

Matsushita, which operates seven different labs in the U.S. through its Panasonic Technologies subsidiary, conducts most of its basic research at just two of them. Scientists at the Matsushita Information Technology Laboratory in Princeton are writing software for computers that might one day be embedded in everyday objects. At the Speech Technology Laboratory in Santa Barbara, 22 programmers and linguists work on speech synthesis and recognition. Last year they developed a video game for New York City's Lexington School for the Deaf that analyzes each student's speech and then demonstrates how to position the tongue and lips to improve it.

At Ricoh, Hart has been constructing a knowbot: "an intelligent agent that monitors the user's activity and figures out what kind of additional information might be helpful. The agent then constructs a query against almost any database, gently taps the user on the shoulder and says, 'If you are interested, I have something that might be relevant to exactly this point in your work.' "

Researchers at all these labs seem confident that the legendary Japanese job security extends to them as well. But until the recession lifts in Japan, their ranks are unlikely to swell. NEC, Ricoh and Matsushita have put plans for expanding their American labs on hold, according to Tamio Saito, who follows Japanese R&D for Big Bang Technology in San Jose, Calif. It is too early to tell whether the Japanese giants will pull back from their experiment in American science. Strategist Methé advises them not to cut bait. "For those companies with the patience," he appeals, "the payback from basic research is phenomenal." —*W. Wayt Gibbs*

Hot Property

Solar cells make a leap in cost-effectiveness

Photovoltaic solar cells, which convert free sunshine directly into valuable electricity, are so obviously a smart idea that one wonders why the people who sell them are not rich yet. The reason is that power generated by burning coal or oil is still usually cheaper. The dismal science trumps physical science: if power lines are nearby, solar cells are too expensive and too low powered to pay for themselves over their 20 years of life.

An increase in the power per unit cost could tip the economic balance, making solar energy a good investment in many more places. That is exactly what United Solar Systems in Troy, Mich., says it has now achieved. United Solar, a joint venture of Canon and Energy Conversion Devices, has developed a panel that uses low-cost, thin-film amorphous silicon to capture sunlight with an efficiency of 10.2 percent. That is high enough—and the panels are potentially cheap enough—to be competitive as an adjunct power source in homes, where the panels might be used as roofing shingle. Two years ago 6 percent was the best efficiency that could be obtained with this form of silicon.

Existing photovoltaic cells that can beat the canonical 10 percent figure are made either from single crystals of silicon or from the polycrystalline form of the material. But single crystals have to be made to high precision at high temperatures and command correspondingly high prices. Even polycrystalline silicon is much more expensive than the amorphous form. The new United Solar design, the work of Subhendu Guha, is a triple sandwich that consists of a layer of silicon combined with layers of two different silicon-germanium alloys. The layers, deposited at 300 degrees Celsius from a mixture of the gases germane, disilane and hydrogen, vary in composition across their thicknesses. That technological trick goes a long way toward explaining how the panels maximize light absorption, each capturing photons of a different energy. Silver electrodes carry the charge away. The sandwich is deposited on a ribbon of 12-inch-wide stainless steel, which suggests another potential advantage over existing panels: it can be bent.

Energy Conversion Devices, which owns half of United Solar, has long experience with solar power: the company is the creation of Stanford R. Ovshinsky, one of the pioneers of photovoltaics. The new solar panel was developed in a three-year, $6.26-million project carried out on a cost-shared basis with the Department of Energy.

Guha says the successful outcome is an argument in favor of strong government support of technology development: "This is an example of how by spending very little government can help build new industries and create high-quality jobs," he declares. United Solar is now building a plant in Newport News, Va., where starting in 1995 it plans to produce about a million square feet of solar panel per year. That is enough to produce 10 megawatts of power, or one sixth of current world demand for photovoltaics. "The Japanese are investing enormous amounts in this sort of technology," Guha warns.

"It's a nice result," says Frank J. Kampas, chief of research at Advanced Photovoltaic Systems in Princeton, N.J., which is also gearing up to mass-produce amorphous silicon cells. "But I don't think they're going to put us out of business." Kampas's company builds its solar cells on glass, so they lack the flexibility that Guha can claim. On the other hand, Advanced Photovoltaic's panels are substantially bigger than those that United Solar will make.

Neither company can yet match the cost of the cheapest electricity from fossil fuels. But even before the new development, photovoltaics was starting to make inroads. Norway has 50,000 photovoltaic-powered country homes, for example, and Pacific Gas and Electric last year installed a 500-kilowatt plant at a substation in California's San Joaquin Valley. Also in 1993 a coalition of U.S. utilities announced plans to introduce 50 megawatts of photovoltaic power over the next six years.

The DOE predicts that the United Solar panels could bring the cost of photovoltaic power down to 12 to 16 cents per kilowatt-hour, less than half its current cost. Power companies usually now charge between six and 20 cents per kilowatt-hour for fossil fuel–generated electricity. If mass-produced panels perform as well as the United Solar prototypes, solar power might be ready to take its place as a standard auxiliary source—assuming the dismal science can be believed. —*Tim Beardsley*

PETER YATES SABA

SOLAR PANEL, featuring low cost and high efficiency, is displayed by its inventor, Subhendu Guha of United Solar Systems in Troy, Mich. Guha's design uses special deposition techniques to maximize the capturing of light.

Softwars

Japanese copyright review touches off a firefight

Last July, Japan's leading business newspaper, *Nihon Keizai Shimbun,* reported that the Japanese Agency of Cultural Affairs had set up a committee to determine whether the way Japanese copyright law treats computer software is in step with evolving U.S. and European standards. Simple logic would suggest that such an undertaking would be welcomed by U.S. computer corporations doing business in Japan. After all, they and other U.S. producers have complained incessantly about the myriad ways in which arcane, idiosyncratic Japanese law and custom frustrate attempts to do business there. Would not uniformity of standards be welcome? Apparently not. Instead of celebrating, the Business Software Alliance, whose members include WordPerfect and Microsoft as well as hardware firms such as Apple and IBM, called in the cavalry.

It arrived on November 2, in the form of an angry letter of protest bearing the signatures of U.S. Secretary of Commerce Ronald H. Brown and U.S. Trade Representative Mickey Kantor to Hiroshi Kumagai, Japan's minister of international trade and industry. Brown and Kantor expressed "grave concern" about the study initiative. The two officials charged that "there is a clear implication that the purpose of the study is to determine ways to weaken Japan's protection of computer programs under its copyright law." The U.S. demanded and received a hearing before the committee on December 13.

The issue over which the controversy churns is reverse engineering of software, specifically through a process known as decompilation. Software as the customer receives it, shrink-wrapped and surrounded by helpful booklets, is written in binary code, strings of 1's and 0's—the only language a computer can understand.

Because programmers cannot easily work with software in this machine language, they write programs in a high-level language, such as C. Then, through a largely automated process called compilation, the program is translated into the language of the computer. Decompilation attempts to run the process backward by translating the binary language, or object code, into something that resembles the high-level language, or source code, that the programmer originally created. The process is imperfect. Human programmers make prompting notes in the high-level code. Compilation deletes all these notes, and they are not recoverable by any automated process.

What is recoverable is a version of the program that a programmer can study, analyze and attempt to understand. Because decompilation involves making interim copies of the software under examination, its opponents allege that it violates the copyright of its creator. Proponents, on the other hand, maintain that as long as the purpose of decompilation is legitimate, the making of interim copies constitutes "fair use."

And there, from the perspective of the Business Software Alliance, lies the beauty of current Japanese copyright law: the concept of fair use does not exist. Indeed, Japanese law flatly forbids decompilation. And whereas U.S. laws can evolve with changing legal precedents, Japanese law remains as written until changed.

Fear on the part of alliance members that U.S. standards might—like baseball, blue jeans and the English alphabet—catch on in Japan is understandable. Japan's fledgling software industry lags many laps behind the sophisticated, powerful U.S. competition. Decompilation could, the alliance members contend, enable Japanese firms to close the distance between themselves and overseas competitors without sweating the extra laps. The U.S. firms should know. Not so long ago a U.S. company used decompilation to execute an Immelmann turn that put it in direct competition with a Japanese firm in the North American market. In 1992 the U.S. Ninth Circuit Court of Appeals heard the case of *Sega Enterprises Ltd.* v. *Accolade, Inc.* Accolade, the American company, wanted to produce game cartridges that would run on the Sega machines but found itself locked out by a complex system of coded microchips in Sega game machines and cartridges.

Accolade physically opened up and examined Sega chips to get the binary object code and then decompiled the code to learn Sega's interface specs—information that is not protectable under copyright law. Once Accolade understood Sega's interface, it produced game cartridges that, without copying any of Sega's object code, ran on Sega machines. The court ruled that because Accolade's purpose—the development of compatible game cartridges—was legal, the copying of Sega's program was fair use and therefore permissible.

Whether or not Japanese law is changed with respect to decompilation, U.S. firms would remain without a remedy on which they can rely at home: the discovery proceeding. Lawyers for a corporation that brings an action can, with the court's permission, examine records and facilities and take depositions from the suspected infringers. IBM's attorneys say discovery is an essential tool for pursuing competitors suspected of illicit decompilation.

At the December 13 hearing, claiming to "speak for virtually the entire U.S. software industry," WordPerfect senior vice president and general counsel R. Duff Thompson described decompilation as an automated process that "if not correctly limited creates an increased threat of piracy." Decompilation, Thompson declared, "is cheaper than obtaining a license and certainly easier than developing a new program from the beginning. That is why some companies want the right to decompile computer programs."

How real is the threat? At IBM Japan's Makuhari facility, just 30 minutes from Tokyo, researchers have built a system specifically designed to decompile IBM customer software. Highly automated, the system is invaluable in cases where software must be modified or where documentation is inadequate. In a demonstration for *Scientific American,* an IBM software engineer took binary code, translated it into hexadecimal format, displayed its structure graphically onscreen and then altered programming sequences without changing function, ending up with binary code that bore no obvious resemblance to the original version. IBM's legal staff maintains that the same process could be used to make "copies" of software that could not be easily identified as such.

But not all U.S. officials and managers agree with the alliance or its members. Proponents of limited decompilation describe as "cosmetic" the kind of changes produced by the IBM decompilation. They add that the pedigree of such software would be obvious to experts. Finally, they assert that instances in which decompilation has been used to produce pirate versions of successful software packages do not exist.

Spokespeople for the American Committee for Interoperable Systems and member companies such as Sun Microsystems argue that changes in the Japanese copyright laws, far from opening the door to fraudulent copying of American software, would clear the way for U.S. companies in Japan to engineer products that would run on entrenched systems—for example, NEC's line of personal computers, which enjoys about a 50 percent share of the Japanese market. Is it a coincidence that the opponents of decompilation are mostly companies in possession of such "standards"? *—Robert Patton, Tokyo*

Are Hospitals Heading for Intensive Care?

Times used to be good in the hospital business. "You could name your price" for services, recalls Joseph P. Newhouse of Harvard University. "It was a very nice environment." The fat years began about 1965, after the passage of Medicaid and Medicare laws that extended health coverage to poor and elderly citizens—and paid for care based on the provider's cost of delivering it. During the next 20 years, the number of hospitals run for profit in the U.S. increased steeply. But now cost cutting by private insurers and government agencies, who have been footing the bill, has begun to dry out the hospitals' cozy ecological niche. Regardless of what health care reform measures pass, most economists agree that the Garden of Eden is turning into a jungle.

According to some economists, health care cutbacks fall largely on for-profit hospitals. Nonprofit institutions have tax exemptions and other advantages that permit them to charge lower prices for the same care, says economist Jayendu Patel of Harvard's John F. Kennedy School of Government. As a result, he predicts, competition could eventually force for-profit hospitals out of the market. Indeed, Humana and Hospital Corporation of America (HCA), two companies that amassed huge chains of hospitals in the late 1970s and early 1980s, have sold the vast majority of those acquisitions in the past few years, he says. HCA, for example, has jettisoned more than 250 of its hospitals since 1987.

Back when hospitals' charges were based on the cost of providing services, Patel says, for-profits and nonprofits competed on an equal footing. But in the mid-1980s Medicare, the largest single source of health care funds, switched to a prospective-reimbursement system that paid the same amount for a pneumonia case regardless of, say, a hospital's internal costs. At that point, the nonprofit institutions' advantages reasserted themselves. Except for some corporate hospitals, which have made alliances with health maintenance organizations (HMOs) for a steady supply of patients, most for-profit institutions "are doing dismally," Patel reports. As a result, investors are deploying their capital elsewhere, and for-profit corporations may be leaving the hospital business even more rapidly than they entered it.

Mark J. Schlesinger of Yale University is unimpressed by this argument. Although many for-profit hospitals formerly owned by big chains have been sold, he notes, very few of them have been closed. They are now operating under the ownership of employees or community-based groups of investors. Furthermore, the tax advantages nonprofits enjoy are balanced by obligations to provide charitable care—a task that consumes as much as 4 percent of nonprofits' revenues. For-profit hospitals also perform some charity care, but, according to Douglas O. Staiger of Harvard, they generally avoid setting up shop in areas where too many unremunerative patients live. For-profit hospitals also tend to offer a narrower range of services than do their charitable competitors. In principle, Schlesinger says, nothing prevents for-profit hospitals from concentrating on services they can deliver better and so competing effectively with nonprofits.

During the past seven years, health care chains have sold off most of their acquisitions.

Indeed, some other economists see nonprofit hospitals as the endangered species in the health care ecosystem. Nonprofits have traditionally used extra revenues from paying patients to subsidize charity care, research and teaching, says Alan M. Garber of Stanford University. But as insurers get stricter, the opportunities for cross-subsidy shrink. And with universal health insurance, the rationale for subsidizing hospitals to provide charity care becomes less compelling.

Under the Clinton health care plan, teaching hospitals would receive extra funds to compensate for the cost of integrating research, education and patient care. In practice, however, this bonus may be inadequate, Garber says. Newhouse concurs, arguing that insurers and HMOs may avoid signing contracts with nonprofit hospitals because those hospitals generally serve sicker patients and thus have higher average costs. Hospitals specializing in research and treatment for particularly expensive diseases such as cancer may also be at risk, Newhouse adds.

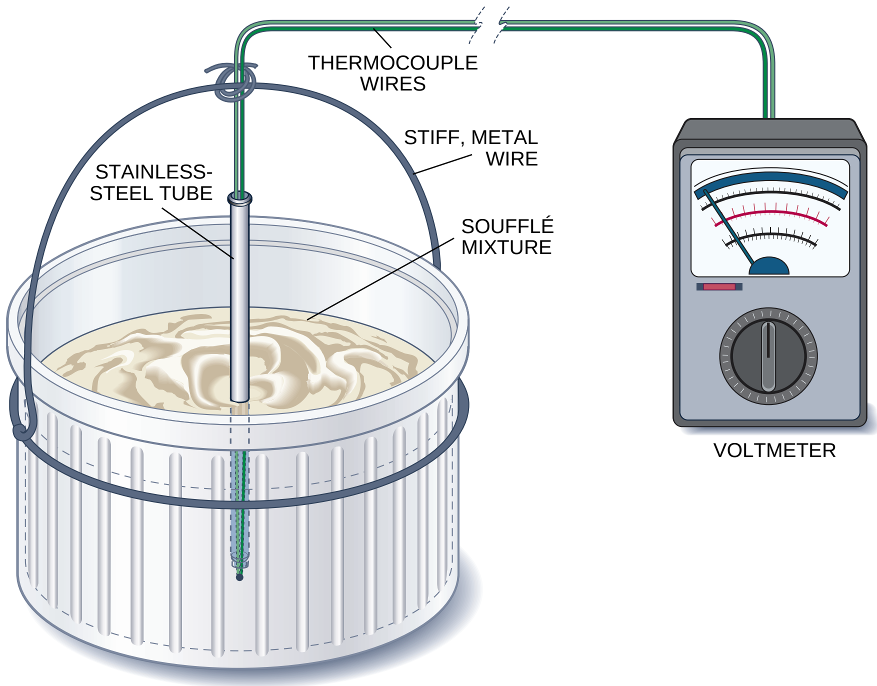
These perils, of course, may not be entirely bad. One of the ideas behind health care reform, after all, is to bring the discipline of the market to a sector of the economy in which purchasers lately paid whatever price they were quoted and relied on sellers to tell them how much to buy. Even during the past decade, Patel says, average for-profit hospital occupancy rates have fallen from 65 to 55 percent (typical nonprofits fill between 65 and 68 percent of their beds). Although it might be asking a bit much for private hospitals to match the 90 percent average occupancy of some public ones, even returning to 1980 rates could pare the nation's health care bill substantially.

Health care economist Burton Weisbrod of Northwestern University dismisses the idea that either for-profit or nonprofit hospitals will undertake such visible cutbacks as closing unnecessary facilities. Instead, he asserts, cost shifting is the order of the day. Weisbrod cites studies showing that Medicare's prospective-payment plan has led to "quicker and sicker" discharges. By releasing a patient to a convalescent facility, a hospital shifts the cost of care to a different part of the health care sector. Indeed, if a patient can be sent home to be cared for by family members, he points out, the cost shifts out of the health care sector to other parts of the economy.

Although such fiscal sleights of hand may please budgeteers, it is not clear whether they reduce the ultimate cost of health care. Economists and epidemiologists are only starting to study such questions as the effects of increasing home care or the substitution of low-tech for high-tech medicine. The trend toward parsimony will at least make studies easier, Garber suggests. Paring away the fiscal fat that has sustained so many informal funding arrangements may force policymakers to confront questions of cost-effectiveness more directly than they have. Whether the public will be happy with the answers is another issue. —*Paul Wallich*

THE AMATEUR SCIENTIST by Nicholas Kurti and Hervé This-Benckhard

The Kitchen as a Lab



SOUFFLÉ TEMPERATURE is taken by thermocouple wires threaded through a stainless-steel tube and supported by a stiff, metal wire that wraps around and loops over the dish. Voltage changes indicate temperature.

Scientific investigations in the kitchen have elucidated the mechanisms behind various culinary tricks and offer a means to create novel dishes [see "Chemistry and Physics in the Kitchen," page 66]. Experiments are easy, and all that is needed are shopping trips to the supermarket and to an electronics-parts supply store. An inexpensive microscope that has 400-power magnification is useful but not indispensable.

In the main article, we mention how we quantitatively determined the best way to produce a soufflé. Remarks made at a workshop in Erice, Sicily, indicated that it was not essential to put a soufflé in the oven immediately after folding in the whisked egg whites. It may be left for an hour or so in a hot water bath (a *bain-marie,* or double boiler) at 40 degrees Celsius (104 degrees Fahrenheit). Moreover, individual soufflé mixtures can be deep-frozen and thawed for cooking—a great help in catering a large dinner party or a banquet.

You can test the hypotheses with individual cheese soufflés. The ingredients consist of $1/3$ cup (75 grams) of butter (about $2/3$ of a stick), about $2/3$ cup (75 grams) of flour, $1\,2/3$ cups (400 milliliters) of milk, four egg yolks, four egg whites and about $1/2$ cup (50 grams) of a grated cheese such as Gruyère. (Although we provide standard kitchen measures, we advocate the metric units for their greater precision.) Add salt, pepper and nutmeg to taste. The ingredients will fill five individual soufflé dishes that are about 10 centimeters in diameter and 6.5 centimeters high. Be sure the dishes have been buttered and coated with bread crumbs or flour.

First, you need to make the béchamel sauce. Over medium heat, melt the butter in a saucepan, then stir in the flour. Under the microscope, the starch granules in the flour paste will appear tightly packed. Add hot milk to the saucepan (starch molecules do not dissolve readily in cold milk). Stir until the mixture reaches a thick, creamy consistency. You will see under a microscope that the starch granules have swollen considerably. The swelling is one of the reasons for the high viscosity of the béchamel sauce: the granules cannot move freely. Another reason is that the starch liberates long polymers made of glucose, which help to create a gel structure. The polymers, which are too small to be seen under the microscope, are amylose (linear chains) and amylopectin (branched chains).

Remove the pan from the heat, and stir in the cheese. When the mixture has cooled to 50 degrees C (the sauce should feel tolerably hot to the touch), add the egg yolks. You may add the yolks singly in one experiment, in twos in another and the four of them together in a third to see whether it makes any difference, as some cookbooks state.

Whisk the egg whites in a clean bowl until the foam is stiff (it should be able to support a whole egg). If you examine the foam under the microscope from time to time, you will see that the air bubbles introduced by whipping become smaller and more numerous as the foam thickens. Surface tension explains why smaller bubbles produce a more stable foam. This force, which is the same as that which creates the meniscus at the surface of a glass of water, draws air bubbles firmly against one another and prevents the water from draining out of the bubble walls. In the long run, gravity will eventually destabilize the whisked egg whites.

Gently fold the cheese mixture into the egg foam. Then ladle the mixture into the five soufflé dishes so that it comes about two thirds of the way up the sides.

The experiment on each soufflé consists of measuring its internal temperature during the cooking, the height the soufflé reaches and the rate of collapse after it has been taken out of the oven. These parameters will help you determine how much time one may leave between ladling the mixture into the dishes and placing it in the oven. If your oven has a window, you may want to try measuring the height of the soufflé as it cooks by securing a metal ruler to the soufflé dish with a wire.

Put one soufflé into the oven immediately. Leave the second in a 40 degree C water bath for 45 minutes. Al-

NICHOLAS KURTI and HERVÉ THIS-BENCKHARD describe the development and application of scientific experiments in the kitchen in their article "Chemistry and Physics in the Kitchen," which appears on page 66.

low the third to stand at room temperature for two to four hours. Place the fourth in the refrigerator for four to six hours; let it warm to room temperature for two hours before placing it in the oven. Store the last in the freezer for 12 to 48 hours, letting it thaw six hours before cooking. In all cases, the oven should be preheated to 180 degrees C (360 degrees F).

The device used to record the temperature inside the soufflé during cooking is called a thermocouple. Essentially, a thermocouple consists of two wires made of different metals joined at a point. It measures temperature as a change in voltage between the two wires.

To make a thermocouple, buy copper and constantan wires that are about 0.2 millimeter in diameter and have insulation that can withstand 200 degrees C. Spot welding is the best means to join the wires, although soft solder will probably do, because the soufflé temperature will not exceed 100 degrees C. One source of thermocouple wire is Willson Scientific Glass (528 East Fig Street, Monrovia, Calif., 91016; telephone: 818-303-1656; fax: 818-303-0599). The company will fuse the wires for you and encase them in the appropriate insulation. Although the wires are inexpensive, most companies have a minimum order of $25 to $50. This amount will buy you at least a couple of meters of wire—far more than you will need.

You can also purchase ready-made thermocouples. Look for a copper-constantan or a chromel-alumel thermocouple, which are called K-type and T-type thermocouples, respectively. We recommend, however, that you build your own, because we have found that commercial thermocouples can be too large and too heavy for this experiment.

The thermocouple may be immersed directly in the soufflé mixture. But we prefer to thread the wires through a thin stainless-steel tube, which helps to keep the wires centered in the soufflé. The tube should be about 1.5 to two millimeters in diameter and 40 to 50 millimeters long. A hypodermic needle is a good choice. Anchor the junction of the thermocouple wires to the tip of the needle with a heat-resistant epoxy resin, such as Araldite. Willson Scientific has said it will sell thermocouple wires threaded through a thin glass tube, which would render the needle unnecessary.

To record temperature, connect the thermocouple wires to a voltmeter that can measure to the nearest 50 microvolts. This voltage is proportional to the temperature difference between the junction immersed in the soufflé and the voltmeter. To calibrate the thermocouple, take voltage readings for ice water (zero degrees C) and for boiling water (100 degrees C). If the zero degree C reading produces a negative voltage, reverse the wires at the voltmeter terminal. From these two points, you can easily calculate the corresponding temperature for any given voltage. The output voltage depends linearly on the temperature. For the copper-constantan thermocouple, each Celsius degree rise corresponds to an increase of 42 microvolts. For the chromel-alumel one, it is 40 microvolts. Some commercially available meters can give the temperature directly. But they are usually supplied with probes that may not be suitable for this experiment.

The rising soufflé can move the thermocouple, so you must secure the device to the soufflé dish. Wrap a stiff, metal wire around the dish, bending and looping it high over the top [*see illustration on page 120*]. The loop should be at least 100 millimeters high. You can use a short length of metal wire to tie the thermocouple wires to the loop.

To prevent the thermocouple wires from being damaged when the oven door is closed, they should be protected by a glass-cloth sleeve about two to three millimeters in diameter. Such sleeves are available from scientific supply shops. An alternative might be a soft, heat-resistant cloth wrapped around that part of the wires sandwiched by the door and frame.

During the cooking, you will note two temperature steps, one at 60 to 80 degrees C and the other about 100 degrees C. The first corresponds either to the point at which the proteins coagulate or to the rise of the cold layers of the mixture to the thermocouple. The other is the vaporization point of water and corresponds roughly to when the soufflé is done (about 20 minutes).

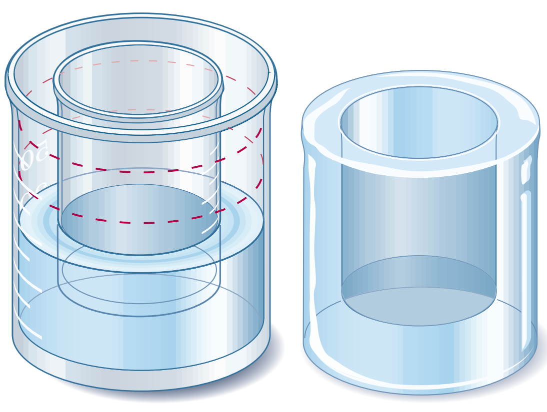
The soufflé rises in part from the expansion of air in the myriad small bubbles in the whipped eggs. Vaporizing water also aids the rise. The coagulation of the proteins keeps the soufflé high after cooking because the process stiffens the walls of the air bubbles. This coagulation may also play a role in why opening the oven door too soon ruins the soufflé. The bubble walls may become sufficiently rigid so as to resist subsequent rising.

Once the soufflé is done, measure its maximum height. By comparing the heights achieved by the different soufflés and the time it takes for them to collapse, you will conclude that the old recipes are correct when they prescribe cooking the soufflés immediately. The longer the delay between ladling and cooking, the less successful the soufflé is, although even the deep-frozen soufflés are acceptable.

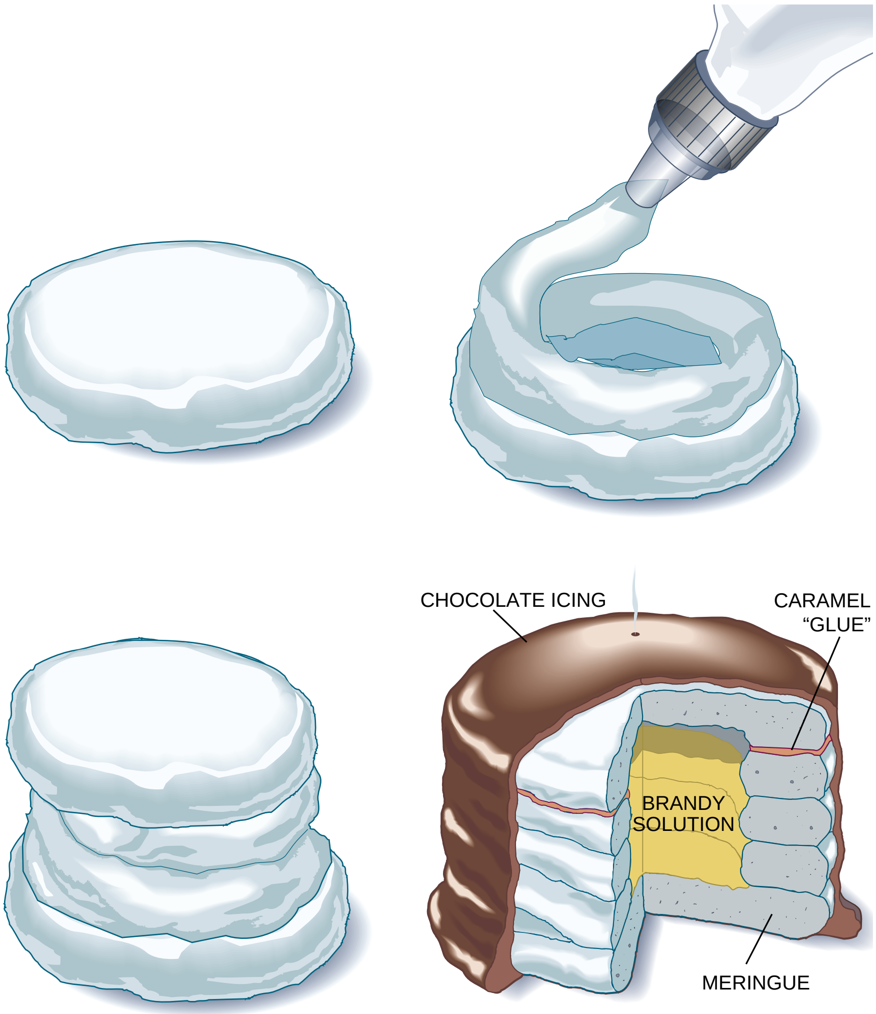
The soufflé experiment quantifies and explains observations made by generations of chefs. But science and technology can also create new dishes, as exemplified by the microwave oven. In traditional cooking, heat penetrates slowly into the food. Taste and consistency depend on the temperature the various parts of the dish reach during the cooking process. In a crisp bread roll, for example, the crust will have reached a temperature of 200 degrees C or more, but the soft inside does not get much above 80 degrees C, because the air bubbles within insulate the dough.

In contrast, microwaves penetrate the food fairly evenly. The waves impart some of their energy to the individual molecules, setting them into motion. The friction between these molecules and their surroundings generates the heat. For microwave ovens to function, the food must contain polar molecules. That is, the electrical charges in the molecules must be asymmetrically distributed. Furthermore, the substance being cooked must not be entirely solid, because their molecules, being fixed, cannot move in response to microwaves.

These two conditions can be simply demonstrated. First, compare the behavior of liquid paraffin, composed of nonpolar molecules, and glycerol, made of polar ones. In a microwave oven, glycerol quickly comes to boil, but the paraffin barely heats.



ICE CORE is made by filling the gap between two beakers with water, which should be frozen in stages to prevent the glass from cracking.



FROZEN FLORIDA is made by swirling meringue so that it forms a case. After the fruit-brandy solution is poured in, the meringue lid is sealed with caramel. Once covered with chocolate icing and frozen, the dessert is heated in a microwave oven.

The difference between liquid and solid states is evident in the following demonstration, which amazes even professional scientists. Make a hollow ice block by placing a four-centimeter-diameter beaker into a larger one, about eight centimeters in diameter. Fill the gap between them with water and put them in the freezer. To prevent the beakers from cracking, freeze the water in stages, adding eight- to 10-millimeter layers of water at a time. Once you have frozen a sufficient amount of water, separate the beakers from the ice under warm running water. Place a small beaker filled to three quarters with water into the ice block. The water will come to a boil in a microwave oven in about 30 seconds, but the ice remains solid and unmelted. Assorted glass containers made from Pyrex will do if you cannot obtain beakers.

The ice-block demonstration is actually the basis for inverting the characteristics of a Baked Alaska. Also called, for obvious reasons, a Frozen Florida, an inverted Baked Alaska consists of a meringue case filled with a concoction rich in alcohol and sugar and covered with chocolate icing. To make the meringue cases, whisk one egg white. When soft peaks form, gradually add about two tablespoonfuls (25 grams) of granulated sugar. Continue whisking until the foam is quite firm (about five minutes), then fold in about $1\frac{1}{3}$ tablespoonfuls (20 grams) of sifted confectioners' sugar.

Deposit the mixture in four round parts on a baking sheet lined with wax paper (one egg white will yield two Frozen Floridas). Each disk should be 50 to 60 millimeters in diameter and about five to 10 millimeters thick. These disks serve as the lids and bases of the meringue cases.

To make the sides, place the remaining meringue mixture in a forcing bag, used to frost and decorate cakes. The nozzle should be about 10 millimeters in diameter. Spiral the meringue on the two meringue disks that serve as bases. The sides should be about 40 millimeters high [*see illustration at left*]. Dry the two vol-au-vent cases and the two lids by baking them at about 120 degrees C (250 degrees F) for about two hours.

The filling consists of equal parts jam, sugar syrup and fruit brandy, such as kirsch or apricot (it is best to match it to the type of jam used). Pour the mixture into the dried cases until they are almost full and seal the lids with caramel syrup. You can make the syrup by boiling a mixture of brown sugar and water. Cover with a layer of icing one to two millimeters thick. The icing is made of $1\frac{3}{4}$ ounces (50 grams) of chocolate (preferably 60 to 70 percent cocoa solids) mixed with about a teaspoonful (10 milliliters) of melted, clarified butter. Do not allow the temperature to exceed 65 degrees C (150 degrees F), because the icing may become lumpy. Make a pinhole in the lids, wrap the confections in plastic wrap and place them in the freezer.

Allow at least 10 hours for the meringues to freeze. Once frozen, they will keep for months. Because of its high alcohol content, the liquid mixture inside will not solidify.

When removing a meringue from the deep freeze, quickly unwrap the plastic film so that water does not condense on the icing. Place the meringue in the microwave set on high and cook until bubbles emerge from the pinhole in the lid (about 30 seconds). The bubbles indicate that the filling has begun to boil. In contrast, the icing remains cold, thanks to the good insulating properties of the meringue. You can increase these properties by coating the inside of the meringue with caramel. The coating prevents the filling from penetrating the meringue and thereby reducing its insulating abilities.

As an alternative to the vol-au-vent technique described, you can try scooping out baked meringue through a small hole (five to 10 millimeters in diameter). Fill with the jam mixture and seal the hole with icing. Such a technique can make bite-size Frozen Floridas. Before serving such small ones, warn any unsuspecting guests about the hot interior.

Of course, you can experiment with the composition of the filling and icing. Freeze samples of identical sizes, place them a few at a time in the microwave oven and determine the time for the filling to boil and for the icing to soften. Perhaps the only drawback to the endless possibilities in experimental gastronomy is that the proof is in the eating. A series of experiments may ultimately compel strict dieting.

Pattern and Pattern Makers

GREAT STREETS, by Allan B. Jacobs. The MIT Press, 1993 ($50). INSIDE/OUTSIDE: FROM THE BASICS TO THE PRACTICE OF DESIGN, by Malcolm Grear. Van Nostrand Reinhold, 1993 ($59.95).

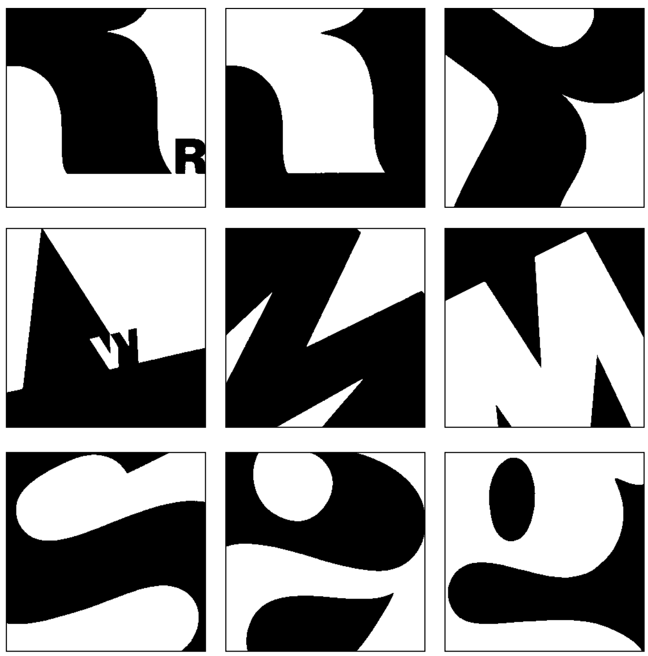
The texture of a city is woven in two dimensions out of streets as that of a cloth is woven from its threads. Across 50 pages, striking, austere patterns of cities rivet the reader's eye: texture. Public roadways are plotted in plan for many cities at a single scale, the rights-of-way white against featureless black for all structures and open land. A few plans are crossed by freer, wider white elements, not streets at all but instead the Arno or the Thames or some local counterpart. Each page maps a square mile on an edge in cities from Ahmedabad to Zurich. (A few cities are granted in their diversity more than one sample square mile.)

The eye is well served. Experience and variety compel your judgment; the differences are evocative. Anyone can see how finely textured is the calligraphic fabric of the oldest districts, how rhythmically woven are cities as new as Savannah or Portland or midtown Manhattan with their open, syncopated, even decorated grids. The imagination drives down the wide roads into the curving neighborhood cul-de-sacs of suburban California in Walnut Creek or floats along the rounded Z of the Grand Canal through the uniquely intricate etching that is watery Venice. Texture can be given a simple measure, too, by counting public intersections. Venice leads in density, with 1,500 connected crossings (overpasses are not counted, not even the Rialto arch itself); an equal surface of Walnut Creek includes only 110. These powerful diagrams—some telling comparisons are made over time—and a dozen pages of reflections on what they disclose, from the fashions and theories of planners to siege defense, occupy less than a fourth of this handsome large volume.

The rest is taken up with many detailed plans and even more attractive drawings, unidentified city scenes in wash and in line, none so abstract as the strong map squares. The author (who credits many students and friends for significant aid over his decade of work) is professor of city planning at the University of California at Berkeley. His quest is for streets "where people will want to be…some of the best streets in the world…as joyful as they are utilitarian." An engaging writer, an eager and observing traveler, here he works not just to describe but toward a purpose. What he is after is the "designable characteristics" of the great streets, so that we may have more of them in our urbanizing world.

The volume opens with the great streets themselves, almost 20 exemplars, a few of them no longer great. Catch the mood: a few are medieval but still great, like Copenhagen's narrow, winding, vivid, crowded Strøget. A few more are broad streets in the grand manner, like the Boulevard Saint-Michel in Paris, which earns its title from the magic of the tree-filtered light and the welcoming ground-floor shop windows and cafes. It is no surprise to find the Grand Canal on the list, defying rigid definitions. It is still the main public street of filigreed Venice, opening a wide, always changing view. The first-rate buildings that have come to line both sides of the canal cast light-dotted images that dance over the waters.

In many an American city there is to be found a grand residential boulevard,



STREET GRIDS, such as this square mile from the center of Florence, Italy (left), *reveal the textured pattern of cities. Although some streets evolve, the best, says the author of* Great Streets, *are designed. A rare glimpse into how designers are trained is offered by* Inside/Outside. *In one exercise* (right), *part after part of a letterform is taken away until recognition fails; the letter fragment is easily recognized when a complete letter is included in the composition.*

long and tree-lined, with a planted median, faced by many large, well-cared-for houses that "bespeak a sense of well-being," even though many may no longer be occupied by the wealthy. Of half a dozen examples, Jacobs singles out Monument Avenue in Richmond, Va., "a most pleasant street on which to be." But famed Market Street in San Francisco and the Avenue des Champs-Elysées have, in spite of keystone locations, lost greatness over the past decades, through changes that divorce the street from its busy people. In Paris, parking areas and the concern for vista have unpleasantly both bared and blocked the broad secondary walkways. A comparison of drawings shows what has weakened the long spine of downtown San Francisco: tall, sealed-off buildings now replace smaller, varied, accessible fronts, and the big, lively trolleys have given way to underground transit and faster private cars. If ever the newer trees grow bigger, they should help.

Beijing harbors several great streets "made only of trees." The buildings are set back, really not on the street, and only long multiple lines of trees take you along the way, on foot, in buses and on bicycles. "The sound, overwhelmingly, is the sound of bicycle bells."

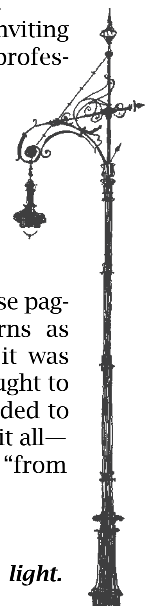
Jacobs tells then of two dozen others, "streets that teach," whose features offer material for future design success. Upper Fifth Avenue along Central Park is one of them, "elegant, compelling" in its two-sided nature, the low-walled park side visually a pleasure and "clearly for everyone." The mixed large and small buildings of the private side are not directly welcoming to strangers yet convey a sense of personal habitation. Princess Street, Edinburgh, shares that two-sidedness even more strongly, a vista over the old city below on the inviting park side and a string of focal buildings that punctuate the busy shopping side. Instructive streets whose success depends on trees are cited in Rio, Palm Beach and Pasadena, where Orange Grove Boulevard is unusually assessed as "extraordinarily compelling…at moderate speed in a car."

Some design conclusions are tentatively drawn: assure physical comfort, visual engagement, diversity, places to walk in leisure. Although the tide still sets toward the suburbs, cities attract. But their streets cannot be for travel alone; they make "the public realm." Although some great streets mainly evolved, "design counts"; the best streets are designed and are then long cared for. "It is encouraging to know that great streets can be built now."

Inside/Outside offers a rare glimpse into how designers are made (of course, like streets, they have to grow on their own as well). Grear, a distinguished graphics designer, veteran teacher at the Rhode Island School of Design, gives us a book of image-filled pages and heartfelt words. His work centers on the enhancement of symbol in text and image. A student starts with an easy task: draw a maple leaf. Let your peers judge it: zeros plaster the papers. Few of them had ever examined a maple leaf for its form. "There is no substitute for knowledge."

In 100 pages of student response to challenge on challenge, letterforms, chairs, images move across dimensions. In one exercise, part after part of a letterform is taken away until recognition fails. Then more complex tasks are set in progressive steps, "not too few, not too many." Visually explain a word, watch a hand tool in action, relate a letter to images of a gate. Text is no medium to convey this cooperative apprenticeship in visual meaning.

The latter part of the inviting volume samples Grear's professional practice with his small Providence team over many years and in many different media, fences, brick paths, books and street signs, a treasury of growth. "I am a lucky man," he remarks, and so are we who con these pages. The maple leaf returns as coda, to remind us that it was the students Grear had sought to know, as his students needed to know the leaf. No one has it all—but we work as we learn, "from the inside out."



***Kurfürstendamm (Berlin) light. From* Great Streets.**

Hard Copy

POCKET PAL: A GRAPHIC ARTS PRODUCTION HANDBOOK, by Michael H. Bruno, editor. Fifteenth edition. International Paper Company, 1992 (1682 Shelby Oaks Drive, Suite 2, Memphis TN 38134) ($6.75, prepaid).

Video screens, forever erased, forever renewed, present their inconstant information fleetingly. To pass on symbol or image, we need hard copy, and today that means *cherchez le papier.* An important papermaker offers a bargain in this no-nonsense little book, a serious, friendly introduction to hard copy and how it is made at every level of scale and quality, as up-to-date (mid-1992) as 15 editions since 1934 can make it.

Printing and publishing are no small matters; the industry accounts for 3 percent of American gross domestic product. It numbers more establishments than any other kind of manufacturing. Estimate the area of paper printed each year in the U.S.; you will find that all those trillion printed pages could be spread to cover the whole of the state of Texas. In contrast, all the video and computer screens of America have a total display area that the city limits of San Francisco could contain; of course, glowing screens incessantly change their spots.

The burden of *Pocket Pal* today is revolution. "If Gutenberg…had stepped into the average printing plant in 1950," he could have plied his trade and "pulled a proof almost exactly as they did" in Mainz 500 years earlier. Since 1950 "practically all hand and machine composition has been replaced" by digital typesetting. Moreover, the speed, ease and flexibility of digital electronics are now all but displacing photography. "A filmless era" is perhaps on the way. So far the most affected stage is prepress: making the modular elements of all copy into ordered digital output.

That means first a scanner to turn images into pixels, assembling dot-rich images along with elegantly dot-made letterforms. Then come the ingenious steps that generate color images from tiny dots in four superimposed ink impressions. And finally trial pages must be "proofed," before the swift presses irreversibly mark miles of blank paper.

Five levels of such technology are in wide use. The simplest is the personal system for designers and artists that allows a few comprehensive if approximate test copies to be prepared for approval. The most elaborate, linked workstations and platemaking, can input text images, layout and compositional instructions to feed fast laser dot generators with the final, color-corrected, retouched, even airbrushed digital pages, to be recorded on photosensitive surfaces able to print.

Certainly electronics is not all. Photopolymer plates, often nylon on steel, vary in solubility between developed and undeveloped parts to make excellent relief plates. Other polymers change in molecular weight on exposure and can be surface-treated to make flat lithographic plates, some good for as many as a million impressions. Digits can even control tiny sparks that eat holes in a thin layer of aluminum on a special polyester base. The metal film takes the ink, and the plastic rejects it, a lithographic plate that acts rather like a CD. The stereotype, a relief plate of metal cast into a papier-mâché mold, is now almost only a metaphor.

The big, fast rotary presses that draw entire webs of paper—long rolls—through many rollers are not new; they go back to steam engine days. Now they tirelessly roll out precisely registered multiple impressions for multicolor work at paper speeds up to 20 or 30 mph. At that hidden interface lies the act of printing.

With letterpress, the high pressure on small highlight dots tends to puncture the paper even while the big dots do not print densely enough. The remedy is a delicate, place-by-place, expensive manual adjustment for varying plate pressure. As a result, the web offset press has replaced letterpress for all but very long print runs. With offset, the printing surfaces are quite smooth; contrast is entirely chemical, a version of the old lithographic grease pencil on wetted stone. The printing image is made water repellent; the unmarked areas, water receptive. One roll adds water, a second, the oily ink; next the ink is rolled off, not onto paper, but onto an intermediate, conforming rubber blanket. That "offset" blanket at last sets its ink image flexibly onto the paper, evading most of the pressure variations that demand the artisan. Around these basics there winds a whole skein of ingenuities and telling details, including ink and paper. The book ends with a benign little commercial that touts instructively, if very proudly, the sponsor's many grades of paper. The celebrated Strathmore papers, admired for a century by artists and coveted by students, are here with more choices than ever, kids' art papers and recycled papers included.

Pocket Pal offers all who read a clear primer of the fast-changing engineering that turns private manuscript into public print.

Photographer of the Stars

A VIEW OF THE UNIVERSE, by David Malin. Illustrated with 210 photographs, most in color. Sky Publishing Corporation and Cambridge University Press, 1993 ($39.95).

One view here David Malin certainly did not take; it shows him curled into the prime focus cage of the fine four-meter Anglo-Australian Telescope (AAT) at Siding Spring. Most of the others are by Malin, who is the most innovative of astronomical photographers today, at work for nearly 20 years both on the frontiers of optical astronomy and in stunning visual support of public awareness of our real universe. In his third book, carefully related and captioned images dazzle and instruct at once.

He opens with a small, warm narrative of his own life. As a poor rural kid, he could see from his bedroom window hundreds of the factory chimneys of Lancashire across the green distance. He left school at age 15 to take a job as an assistant in a Manchester pesticide laboratory. His wise employers (the Swiss firm of Geigy) sent him part-time to technical college, where he thrived. In the lab where he worked there was a dusty but fine research microscope. Once he had proudly processed the snapshots from his first Paris holiday, he was encouraged to try his hand at microphotographs for the lab. He soon became a chemical microscopist of repute. But in his thirties, promotion seemed at an end. The aura of his Australian wife must have moved him to scan the back pages in *Nature* for jobs overseas; before long, he had been hired to assemble a photo lab at the brand-new AAT far south under the Clouds of Magellan.

His tools and methods, the strange new hemisphere of sky, and the help he draws from image-making friends, all the way back to Sir John Herschel, frame and enrich this never-technical volume. (He does cite for the photographic amateur where the details of his novel techniques may be found.)

Color photographs of the astronomical sky are by now common currency. One photomicrograph is eloquent on technique—an echo of Manchester days? It shows the magnified cross section of a small bit of a color print. Four thin layers coat the matted fibers of the paper backing; three are dyed emulsions—yellow, magenta and cyan—backed by a truly white reflecting pigment that screens the drab fibers below. What lies behind the photographically induced retinal perception of color is here opened to thoughtful inspection.

That one-shot color film assembly is certainly ingenious, and long ago it placed the high art in the hands of any tyro who knows of an adept photofinisher. But Malin makes his best color photographs otherwise: by precisely superimposing three distinct black-and-white photonegatives, first taken and then printed in color, through three color filters, a procedure used in 1861 by physicist James Clerk Maxwell to make the first of all color slides. Malin's films are mainly special Eastman spectrographic emulsions, hypersensitized by hours of soaking first in nitrogen and then in hydrogen gas. (This arcane rite is surprisingly well understood.) The results show "stunning colors," a step toward the subtlety of astrophysical spectroscopy through visual hues alone. Direct comparison of the two methods of color photography is striking.

Half a dozen schemes to enhance the perception of color and detail in cosmic photography are in use, three of them introduced by Malin himself. The scheme he calls "unsharp masking" uses a duplicate image deliberately made fuzzy. Carefully superimposing a negative of the fuzzy shot on the positive original produces a near cancellation of all coarse features in the joint image, so that fine-grained detail stands out. The audio analogue is a high-pass filter. This process is much used at AAT.

One of the finest examples is a color view of the nearby radio galaxy called Centaurus A. In the curdled band of obscuration that cleaves the glowing galaxy, there appeared in 1986 a supernova. The detail shown is in itself striking; the masking reveals a rash of the pink nebulae that mark the star-forming regions within the dust. But one very bright spot glows in an unnatural intense blue-green. That was indeed the new supernova; its color was not abnormal, but clouds had blocked the third filter shot. A year later, when time came to reshoot the lacking exposure, the supernova had faded, so the final image of that star alone has no red; blue-green is the answer to white minus red.

All these devices, with key maps, historical comparisons in black-and-white, and much more, have been used brilliantly in five chapters to assemble a coherent visual examination of stars, dust, gas and galaxies, the fresher since only one in 10 of all on the earth ever sees the far southern sky that circles against the clock.

While you patrol half a sky, you can expect to miss events in the other half. It is sheer good fortune that the Clouds of Magellan harbored in 1987 a supernova closer to the earth than any since old Kepler watched from Prague the wonder star of 1604. Malin shows us two wonderful photographs. In one a preexplosion negative of the region is superimposed on the supernova near its peak. A handful of sharp dark spots mark on that swollen glow the stars present beforehand; comparison with Malin's magnified view of the postexplosion sky reveals one star gone away, in its place a fading red patch. Will a pulsar appear there some year as the debris spreads? AAT will watch with care for optical signs. Astronomy today may be wide-band, from radio to neutrinos, but the eye remains eager for the gala of color and what it means.

ESSAY by Anne Eisenberg

E-mail and the New Epistolary Age

During the past few years, scientists the world over have suddenly found themselves productively engaged in a task they once spent their lives avoiding—writing, any kind of writing, but particularly letter writing. Lured by electronic mail's seductive blend of speed, convenience and economy, people who never before touched the stuff are routinely, skillfully, even cheerfully tapping out a great deal of correspondence.

It's the new, inadvertent epistolary age. Electronic networks, woven into the fabric of scientific communication these days, are the route to colleagues in distant countries, shared data, bulletin boards and electronic journals. Anyone with a PC, a modem and the software to link computers over telephone lines can sign on. An estimated four million scientists have done so, with more joining every day, most of them communicating through a skein of interconnected domestic and foreign routes known collectively as the Internet, or net.

Letters are the basic vehicle of this communication, and although e-mail correspondents are not yet turning out *Les Liaisons Dangereuses* or *Clarissa,* they are certainly developing the beginnings of a very lively style. E-mail already possesses a terminology, an etiquette (called netiquette) and even a distinctly youthful writing voice.

The hallmark of e-mail style is a facelike symbol called the emoticon, a breezy electronic device that duplicates the effect achieved in an earlier, handwritten life by abundant underlining and the words "just kidding!!" in parentheses. For example, the emoticon :-), best viewed sideways for full effect, means that the writer's intention is to be funny or good-humored. The emoticon ;-) means that the words preceding it were delivered with a wink (the semicolon being the wink). Emoticons are so popular that several dictionaries of them have actually appeared in computer bookstores, although those on the net who prefer their writing unadorned by winks and nudges look on these collections dourly, that is, :-(.

E-mail creates its tone not only with smileys, as emoticons are sometimes known, but by various symbols that stand in efficiently and informally for many standard letter-writing phrases. The ampersand (&), for instance, which appears floating on a line by itself at the end of much e-mail, means, "I look forward to hearing from you."

For a style this ebullient, e-mail page format is fairly austere: there are no italics or boldface. This is because e-mail lingua franca—the way readable text is exchanged efficiently among the anarchical collection of Macs, DOS machines, UNIX boxes and the multiplicity of computers that e-mail travels—is seven-bit ASCII (American Standard Code for Information Interchange). Once most of the symbols on the old typewriter keyboard were assigned to ASCII's 128 slots (or double that in the extended character set used for technical graphics and foreign languages), there was little room for niceties beyond uppercase and lowercase. At first, e-mailers tended to forget that lowercase was even an option, contenting themselves instead with all capitals, particularly when they were writing to computer bulletin boards (that is, posting articles to newsgroups). Nowadays users are learning to abandon the capital-lock key, aware that ALL UPPERCASE SHOUTS, even on-line.

E-mail is a notoriously leaky way to communicate. Users tap out their messages and send them wending through a patchwork of electronic conduits, only to discover on occasion that the wrong person—or even people—have received the message or that the message waited in a system backup, sat on tapes in the machine rooms for a year and then was resurrected. That, for example, is how much of Oliver North's connection to the Iran-Contra affair was documented.

Still, e-mail users carry on buoyantly, even uninhibitedly. Writer's block has never been a problem on the net—far from it. Alone at their computers, tapping away, correspondents report few difficulties getting started, much less continuing. True, there is a slight awkwardness the first few times, similar to opening trials with a telephone answering machine in the old days, but that constraint soon falls away. Possessed of the mistaken but reassuring notion that paperless communication is not really communication at all, e-mail users shed the self-consciousness characteristic of their other lives in writing. People who would normally labor over multiple versions of any document instead fire off quick, serviceable first drafts, although they are careful not to flame (express themselves insultingly or provocatively) or engage in flame wars. Even the simplified mechanics of e-mail are liberating: no stamps, envelopes or, in many cases, bothersome printed text to handle.

No wonder e-mail is starting to edge out the fax, the telephone, overnight mail and, of course, land mail (snail mail, in the parlance). E-mail shrinks time and distance between scientific collaborators, in part because it is conveniently asynchronous (writers can type while their colleagues across time zones sleep; their message will be waiting). If it is not yet speeding discoveries, it is certainly accelerating disclosures. For instance, no sooner had Stanley Pons and Martin Fleischman announced their discovery of cold fusion than computer networks, particularly Usenet and Bitnet, began to flash. Scientists compiled and transmitted detailed entries on deuterium, heavy water and "everything you wanted to know about palladium and were afraid to ask." They shared procedural information, gathered reports and, very quickly, gave accounts of cold fusion's theoretical improbability.

Jeremy Bernstein, the physicist and science writer, once called e-mail the physicist's umbilical cord. Lately other people, too, have been discovering its connective virtues. 1993 was the year e-mail caught on outside the scientific community: National Public Radio read its first e-mail from listeners; e-mail addresses popped up on business cards below telephone and fax numbers; and an epistolary novel included e-mail in its characters' correspondence. 1994 promises more of the same: e-mail, around for 25 years, originally an insignificant add-on to ARPAnet, is flowering. Folklorists are studying it; college students are lining up at the computer center for accounts, and, sure sign that it has come of age, the *New Yorker* has celebrated its liberating presence with a cartoon—an appreciative mutt perched at a keyboard, saying happily, "On the Internet, nobody knows you're a dog."

ANNE EISENBERG answers her e-mail from Polytechnic University, Brooklyn, N.Y. (aeisenb@photon.poly.edu).