SCIENTIFIC AMERICAN

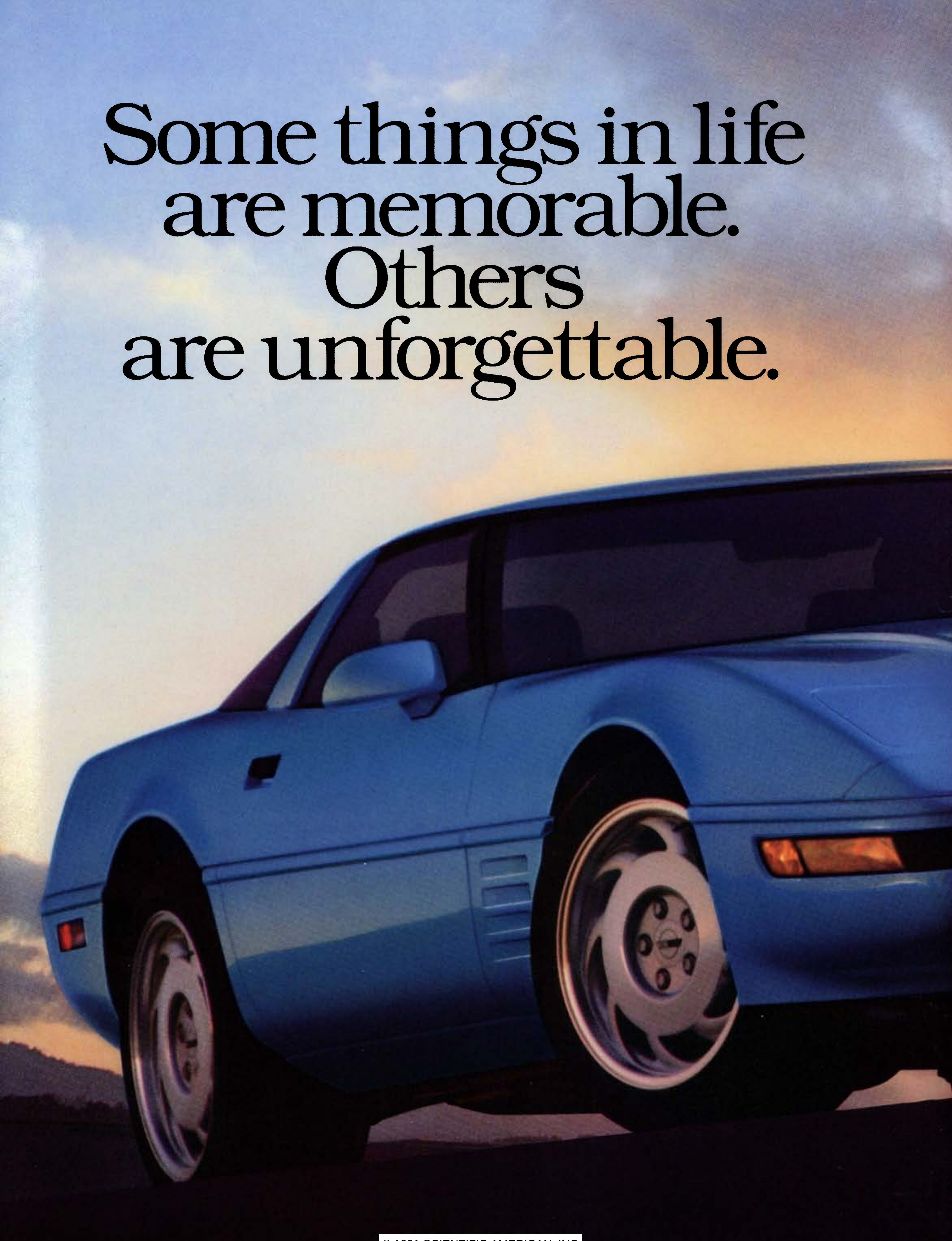
MAY 1991
$3.95

Exploring the genetic heritage of racehorses.

Can anyons explain high-temperature superconductivity?

The impact of Kuwait's burning oil wells.

***Silicon sees** a cat. This retina-on-a-chip mimics the functions of cells in the human eye.*


Some things in life
are memorable.
Others
are unforgettable.


Corvette Coupe ■ 5.7 Liter fuel-injected V8 with 245 horsepower. ■ Choice of a 4-speed automatic or 6-speed manual. ■ Power 4-wheel ABS IIS disc brakes. ■ P275/40ZR-17 Goodyear Eagles. ■ New 17 x 9.5" alloy wheels. ■ New bodywork details including rounded ZR-1 type tail. ■ 3-year/50,000-mile Bumper to Bumper Plus Warranty.*
MORE PEOPLE ARE WINNING WITH THE Heartbeat OF AMERICA
TODAY'S CHEVROLET™
1991 Corvette Coupe. Like a cool, refreshing mist in some fog-shrouded dream, your first encounter is something you'll probably never forget. To call it a mere memory is to belittle the moment. Rather, it becomes a vision etched in one's soul forever. It's not uncommon for people to recite every detail of the very first one they laid eyes on. For there is only one Corvette. And though challengers have come and gone, we've spent the last 38 years joyfully forging America's only true sports car into what it is today. Quite simply, this is the best 'Vette yet. And so a message to the faithful: If you've been waiting for the right time to renew your dream, what are you waiting for?
CORVETTE
*See your Chevrolet dealer for terms of this limited warranty. A deductible will apply after 12 months/12,000 miles.
Chevrolet, the Chevrolet emblem, Corvette and the Corvette emblem are registered trademarks of the GM Corp.
d. Let's get it together...buckle up.
GM®

A LOT OF TIRES COST LE
THAT'S BECAUSE

To everyone out there looking to save a few dollars on a set of tires, let's not mince words. You buy cheap, you get cheap.

There may be a lot of tires out there that cost less than a Michelin.

The only question is, what do you have to give up if you buy one?

Do they handle like a Michelin?

Do they last like a Michelin?

Are they as reliable as a Michelin?

Then ask yourself this: Do you really want to find out?

At Michelin, we make only one kind of tire. The very best we know how.

Because the way we see it, the last place a compromise belongs is on your car.

As a matter of fact, we're so obsessed with quality we make the steel cables that go into our steel-belted radials.

We even make many of the machines that make and test Michelin tires.

And our quality control checks are so exhaustive that they even include x-rays.

These and hundreds of other details, big and small (details that may seem inconsequential to others), make sure that when you put a set of Michelin tires on your car, you get all the mileage Michelin is famous for.

True, there may be cheaper tires. But if they don't last like a Michelin, are they really less expensive?

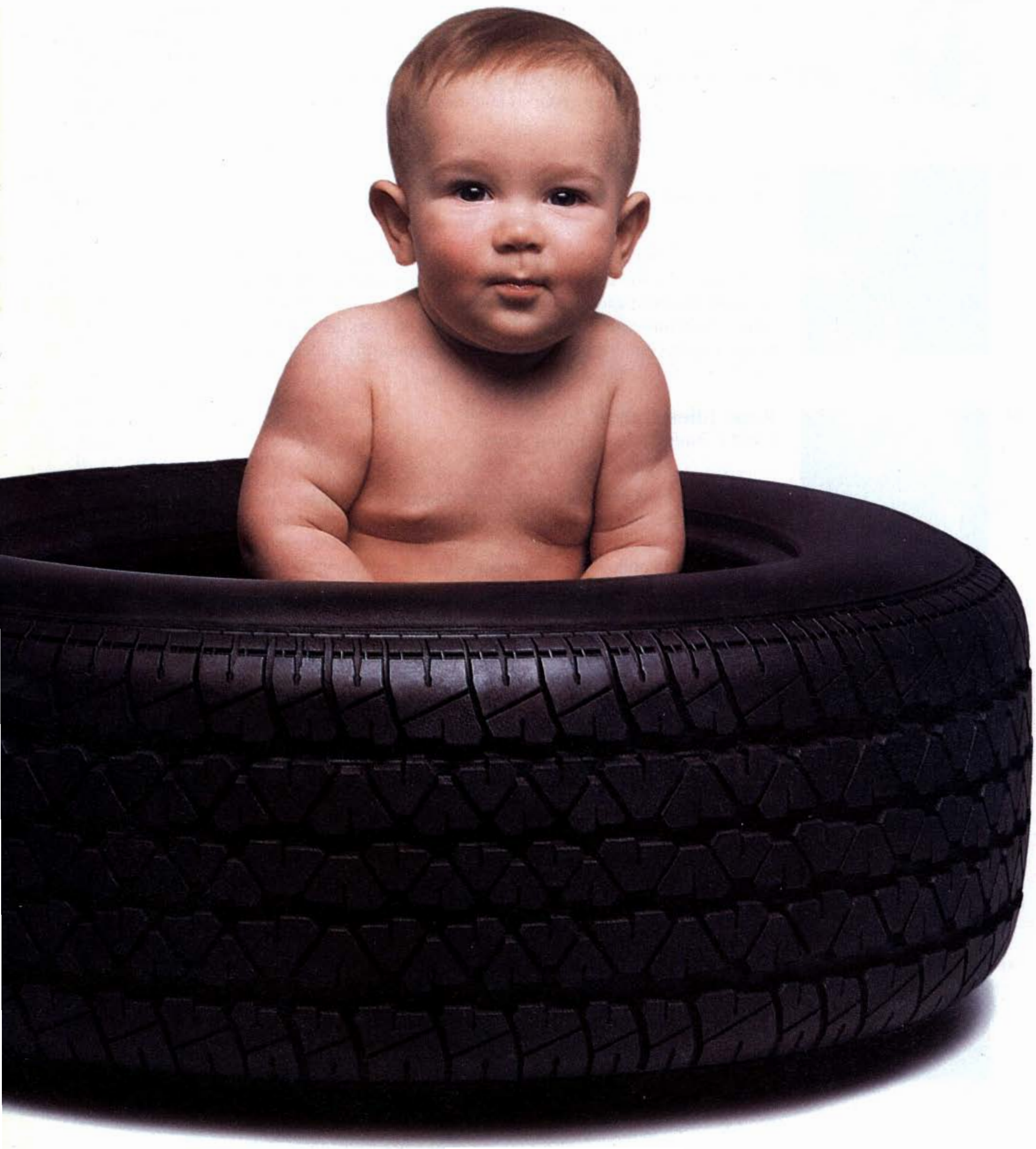
So the next time someone tries to save you a few dollars on a tire, tell him this: It's not how much you pay that counts. It's what you get for your money.

And then *he'll* know that *you* know that there's only one reason a tire costs less than a Michelin.

It deserves to.

MICHELIN®
BECAUSE SO MUCH IS RIDING ON YOUR TIRES.®

SS THAN A MICHELIN.
THEY SHOULD.



SCIENTIFIC AMERICAN

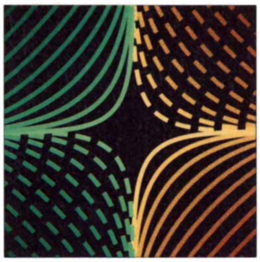
May 1991 Volume 264 Number 5

50 

Progress in Oral Rehydration Therapy

Norbert Hirschhorn and William B. Greenough III

With each medical advance seeming to be increasingly high-tech and costly, oral rehydration therapy is a notable exception to the trend. The administration of a simple electrolyte solution made with readily available ingredients now saves one million children a year from death caused by diarrhea-induced dehydration. Recently it reduced fatalities during a cholera epidemic in Peru.

58 

Anyons

Frank Wilczek

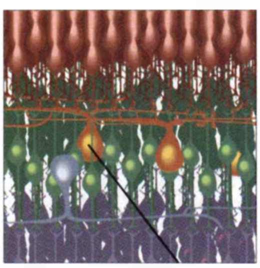
To the few physicists who first thought about them, anyons were mathematical curiosities that provided new insights into the theory of quantum mechanics. But experimental evidence accumulated over the past decade indicates these entities do exist. In fact, the behavior of anyons offers a compelling explanation for high-temperature superconductivity.

66 

René Jules Dubos

Carol L. Moberg and Zanvil A. Cohn

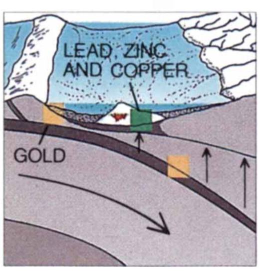
A childhood bout with rheumatic fever may have sent René Dubos on a lifelong exploration into the nature of health and disease. As a researcher at the Rockefeller Institute, he discovered the first clinically important antibiotic. As a philosopher, he formulated an ecological theory of disease that matured into a profound, influential view of our place on the earth.

76 

The Silicon Retina

Misha A. Mahowald and Carver Mead

The formation of visual images in the retina of the eye depends on layers of interconnected cells. The functions of three of these layers—photoreceptor, horizontal and bipolar cells—can be duplicated by simple electronic devices etched onto a silicon chip. This artificial retina illuminates biological computation and has implications for computer vision and signal processing.

84 

The Genesis of Ores

George Brimhall

Human history and technology have been shaped by metals. How did they become concentrated in minable deposits located so conveniently near the earth's surface? The author explains the mechanisms of fluid transport—by magma, water and even air and wind—responsible for the chemical and physical interactions that created bodies of metallic ores throughout geologic history.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017. Copyright © 1991 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. only 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537.

92


The Genetics of Thoroughbred Horses
Patrick Cunningham

The lineage of all magnificent Thoroughbred racehorses can be traced to a handful of animals imported from Africa and the Middle East in the 17th century. Just 10 horses contributed more than half of the genes in today's Thoroughbreds. Despite the wealth of breeding data, genetic studies have begun only recently.

100


Mechanical Engineering in the Medieval Near East
Donald R. Hill

When Paris was still a village, 10th-century Baghdad was a metropolis of 1.5 million inhabitants. To support such urban centers, Muslim engineers developed sophisticated water and wind machines with valves, cranks and pistons. Many of these innovations influenced the development of modern machinery.

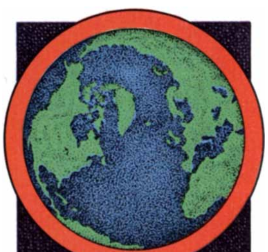
106


TRENDS IN ENERGY

Cleaning Up Coal
Elizabeth Corcoran, staff writer

According to the percentages, coal is still King. Coal-fired power plants generate more than 50 percent of U.S. electricity. But every year those utilities also pour forth 70 percent of the sulfur dioxide and significant portions of other pollutants that cause acid rain and contribute to global warming. Now the U.S. is trying a novel market-based approach to reducing those emissions.

DEPARTMENTS

17


Science and the Citizen

Kuwait is burning. Will the oil fires cause global cooling?... Genetic editors.... Slaking California's thirst.... Magic numbers in atomic clusters.... Ripples from undersea preserves.... PROFILE: Science historian Thomas S. Kuhn.

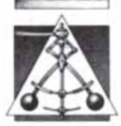
118


Science and Business

New twists to preventing repetitive-motion disorders.... Full-color spectrophotometry.... Calming the brain with sex hormones.... Fly-by-light.... THE ANALYTICAL ECONOMIST: Are boom-and-bust cycles inevitable?

10


Letters
Can there be capitalism without capital?... Soccerball science.

14


50 and 100 Years Ago
1941: Not numerical strength but technological superiority wins wars.

126

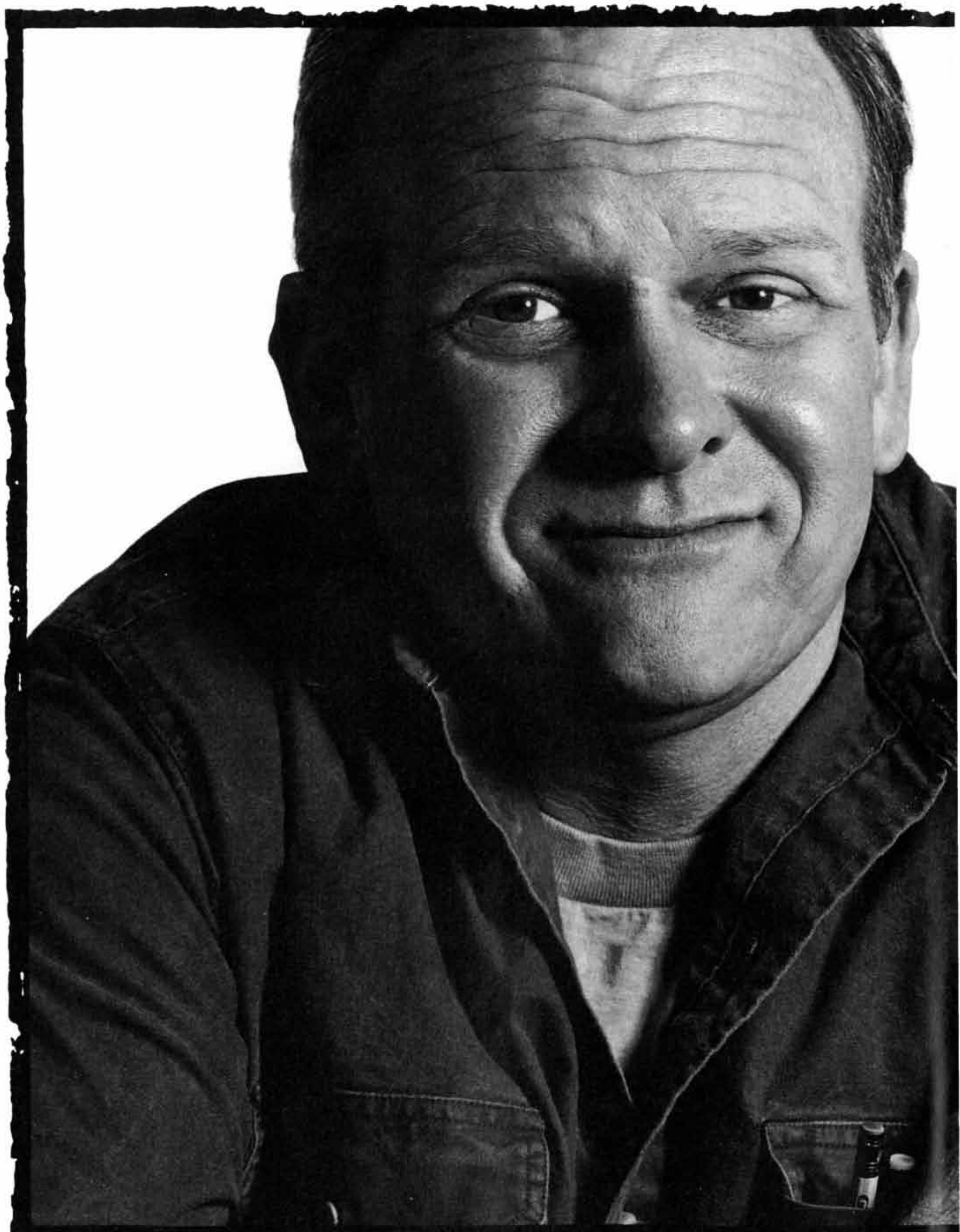

Mathematical Recreations
Brace yourself. It pays to be flexible when rigidity is at stake.

130

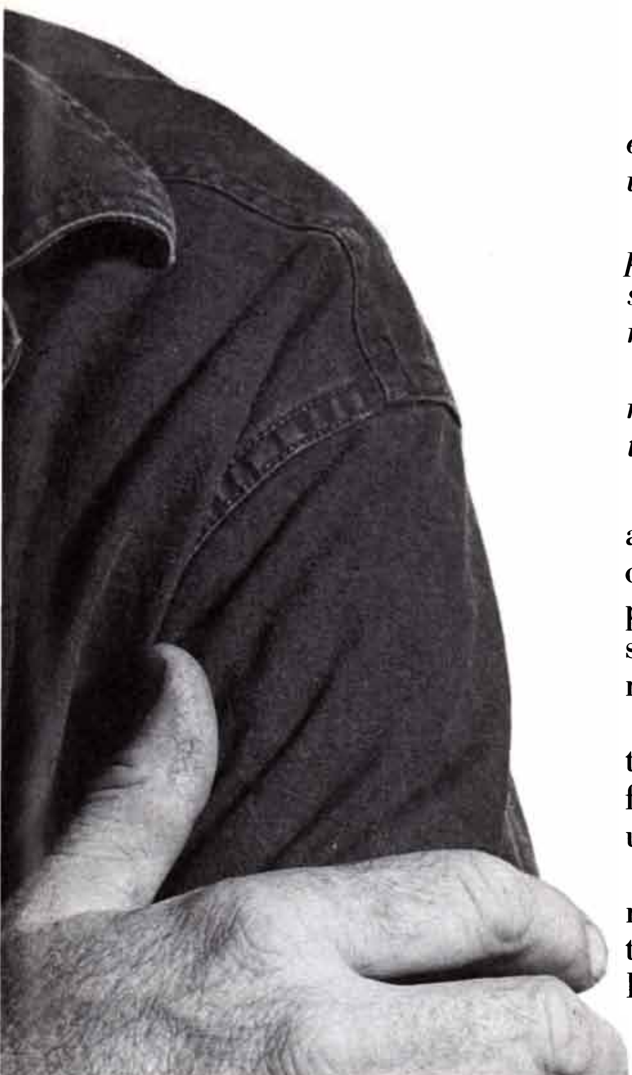

Books
The Book of Sediments.... Fleeting elements.... Try these crops.

136


Essay: *John Kenneth Galbraith*
The military-industrial complex is still a problem to be reckoned with.



“The idea of computers in the factory used to scare the daylights out of me. Now I run one.”



“I figured I’d get burned either way—computers show up and I get fired, or computers don’t show up and the plant closes down.

“But what happened is, they retooled the plant and while that was going on they sent me to school, to an IBM-sponsored course at the community college.

“Here are two things I learned. I learned a new job that’s better than my old one. And I learned that our plant won’t be boarded up any time soon.”

Yes, you can teach old factories new tricks, and CIM (Computer Integrated Manufacturing) is one of them. CIM coordinates the manufacturing process, from design to distribution, as a single system. Needless to say, it can make our economy more competitive.

And yes, we’ll have to teach people some new tricks, too. That’s why IBM sponsors CIM education for students and workers at over 70 colleges and universities across America.

To learn more about CIM and IBM’s commitment to CIM education, write to us at IBM, P.O. Box 3974, Dept. 972, Peoria, IL 61614.

IBM®

THE COVER photograph shows how a silicon retina chip sees Spike, a large calico cat. The retina chip was designed by the authors at the California Institute of Technology to model the visual processing of the human retina. Output from the chip was recorded on videotape and then photographed to produce this image (see "The Silicon Retina," by Misha A. Mahowald and Carver Mead, page 76).

THE ILLUSTRATIONS

Cover photograph by Jessie Simmons

Page	Source
51	Norbert Hirschhorn
52–53	George V. Kelvin
54–55	Laurie Grace
56	Laurie Grace (*left*), from *Chicken Soup with Rice*, by Maurice Sendak, © 1962 by Maurice Sendak; reprinted by permission of HarperCollins Children's Books (*right*)
59	Comstock, Inc./Hartman-DeWitt
60–61	Andrew Christie
62	David Poole
63–64	Andrew Christie
67	Lawrence R. Moberg
70	Harris County Medical Archive, Houston
71	Laurie Grace
72	Vernon Knight (*left*), BioPhoto Associates/Science Source/Photo Researchers, Inc. (*right*)
73	Don Riepe/American Littoral Society
77	Jessie Simmons
78	Andrew Christie
79–80	Jessie Simmons
81	Andrew Christie, Jessie Simmons (*photo insets*)
82	Jessie Simmons
84–85	Juan Pablo Lira
86–91	Hank Iken
92–93	Dan Dry & Associates
94	Edward Bell (*bottom*)
94–95	Julie A. Wear (*top*)
96–98	Johnny Johnson
100	Gabor Kiss
101	George Retseck
102	World of Islam Trust, Peter N. Hayward (*model*)
103–105	George Retseck
106–107	Peter A. Coviello, Ian Worpole (*schematic*)
108–111	Ian Worpole
112–113	Johnny Johnson
114–115	Tom Wolfe
126–128	Johnny Johnson

SCIENTIFIC AMERICAN®

Established 1845

EDITOR: Jonathan Piel

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

ART: Joan Starwood, *Art Director;* Edward Bell, *Associate Art Director, Graphics Systems;* Nisa Geller, *Photos;* Johnny Johnson

COPY: Maria-Christina Keller, *Copy Chief;* Nancy L. Freireich; Jonathan Goodman; Daniel C. Schlenoff

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

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

ADVERTISING: Robert F. Gregory, *Advertising Director.* OFFICES: NEW YORK: William Buchanan; Peter Fisch; Michelle Larsen; Meryle Lowenthal. CHICAGO: 333 N. Michigan Avenue, Chicago, IL 60601; Patrick Bachler, *Advertising Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; Edward A. Bartley, *Detroit Manager;* William F. Moore. WEST COAST: 1650 Veteran Avenue, Suite 101, Los Angeles, CA 90024; Kate Dobson, *Advertising Manager;* Joan Berend, San Francisco. CANADA: Fenn Company, Inc. DALLAS: Griffith Group.

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

INTERNATIONAL: EUROPE: Roy Edwards, *International Advertising Manager,* London; GWP, Düsseldorf. SEOUL: Biscom, Inc. TOKYO: Nikkei International Ltd.

ADMINISTRATION: John J. Moeling, Jr., *Publisher;* Marie D'Alessandro, *Business Manager*

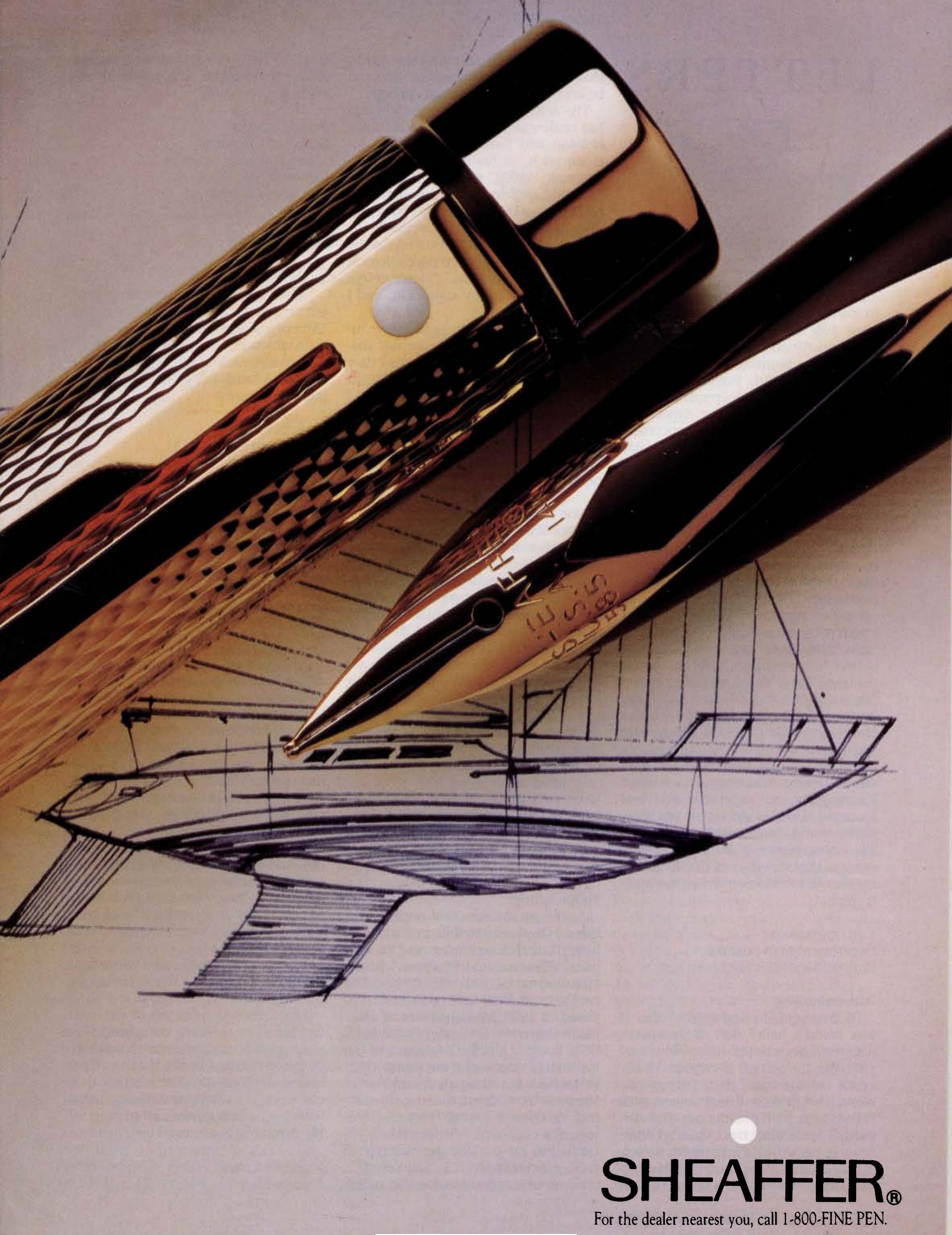
SCIENTIFIC AMERICAN, INC.
415 Madison Avenue
New York, NY 10017
(212) 754-0550

PRESIDENT AND CHIEF EXECUTIVE OFFICER: Claus-Gerhard Firchow

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

CHAIRMAN OF THE BOARD: Dr. Pierre Gerckens

CHAIRMAN EMERITUS: Gerard Piel


SHEAFFER®
For the dealer nearest you, call 1-800-FINE PEN.

LETTERS

Perchance to Dream

To the Editors:

I would appreciate the opportunity to correct misrepresentations of my ideas and those of Sigmund Freud in Jonathan Winson's otherwise valuable article, "The Meaning of Dreams" [SCIENTIFIC AMERICAN, November 1990]. Winson attributes to Freud the view "that dreams reveal our innermost unconscious feelings and concerns." What Freud actually said was that dreams conceal our unconscious wishes by disguising them as apparently meaningless and bizarre symbols. Psychoanalytic interpretation of the concealed wishes was required to reveal a dream's true meaning. Winson fails to recognize that censorship—not revelation—is the essence of Freud's dream psychology.

I never suggested that dreams are "inherently meaningless" or that they have "deep psychological significance" (where "deep" implies hidden). On the contrary, from the publication of my original papers to my recent books, my position has been consistent. I reject Freud's concept of censorship and instead attribute dream bizarreness to the unique brain physiology of REM sleep. Dream bizarreness is indeed meaningless, not defensive as Freud assumed. Dreams do reveal our view of the world, but they do so directly. The consequence of Winson's double error is that the view of dream meaning that he attributes to Freud is actually mine.

J. ALLAN HOBSON
Department of Psychiatry
Harvard Medical School

Winson replies:

Of course, I do understand that it was Freud's belief that dreams were disguised because of repression, as I stated in the second paragraph of my article. At the end, I drew two conclusions. The first, in disagreement with Freud, was that dreams are not disguised, but rather their unusual character is caused by the complex associations taking place during REM-sleep memory processing. The second is that Freud discovered an important truth, that there is an unconscious that can be accessed through dream content.

The activation synthesis hypothesis has undergone several revisions since Hobson and Robert W. McCarley introduced it. In his book *Sleep* (W. H. Freeman, 1989), Hobson stated that "I would like to retain the emphasis of psychoanalysis upon the power of dreams to reveal deep aspects of ourselves...." On that basis I stated that he acknowledged the "deep psychological significance of dreams." "Deep" refers to "profound" in my statement, and I assume also in his.

Hobson and I agree, as do other investigators, that dreams are not disguised. We disagree about whether they reflect a basic, cognitively significant memory process. The question would be resolved by an experiment to determine whether a memory deficit resulted from eliminating theta rhythm selectively during REM sleep. Such an experiment is under way.

To the Editors:

Winson suggests there is a connection between the disproportionately large brains of monotremes and their lack of REM sleep. He hypothesizes that the large prefrontal cortex is needed for processing memories. Another group of mammals seems to support that idea: the odontocete (toothed) whales, including the dolphins, which also do not experience REM sleep. Although their large brains have often been cited as evidence for the high intelligence of these animals, studies have not shown that their intelligence is on a par with that of primates.

MORGAN S. LYNN
La Jolla, Calif.

Buckeyball, Anyone?

To the Editors:

As the person who first purified carbon 60 (buckminsterfullerene) and carbon 70, may I comment on "Buckeyballs" ["Science and Business," SCIENTIFIC AMERICAN, January]? It should be emphasized that C-60 was first proposed in 1971 by Eiji Osawa of Hokkaido University and independently by D. A. Bochvar and E. G. G'alpern of the Institute of Organoelement Compounds in the U.S.S.R. During my recent visit to Moscow, Dr. G'alpern dug out the original cardboard model from the bottom of a cupboard. The Russians were thrilled to see the first pure sample of C-60, especially Dr. I. V. Stankevich, a chemist whose great interest in soccer inspired their theoretical work on the molecule (which is structurally identical to a soccerball).

ROGER TAYLOR
School of Chemistry
and Molecular Sciences
University of Sussex
Brighton, England

A Bitter Pill

To the Editors:

"Aspirin," by Gerald Weissman [SCIENTIFIC AMERICAN, January], states that "Americans consume 16,000 tons of aspirin tablets a year—80 million pills...." That makes each tablet weigh about 6.4 ounces. A tough pill to swallow! Eighty billion, perhaps?

PAUL D. COHEN
Huntington Station, N.Y.

Editor's note:

Right you are. And that is why typographical errors give editors headaches. On page 90 of that article, the chemoattractant peptide C5a was incorrectly called $C5_2$.

Not So Miraculous

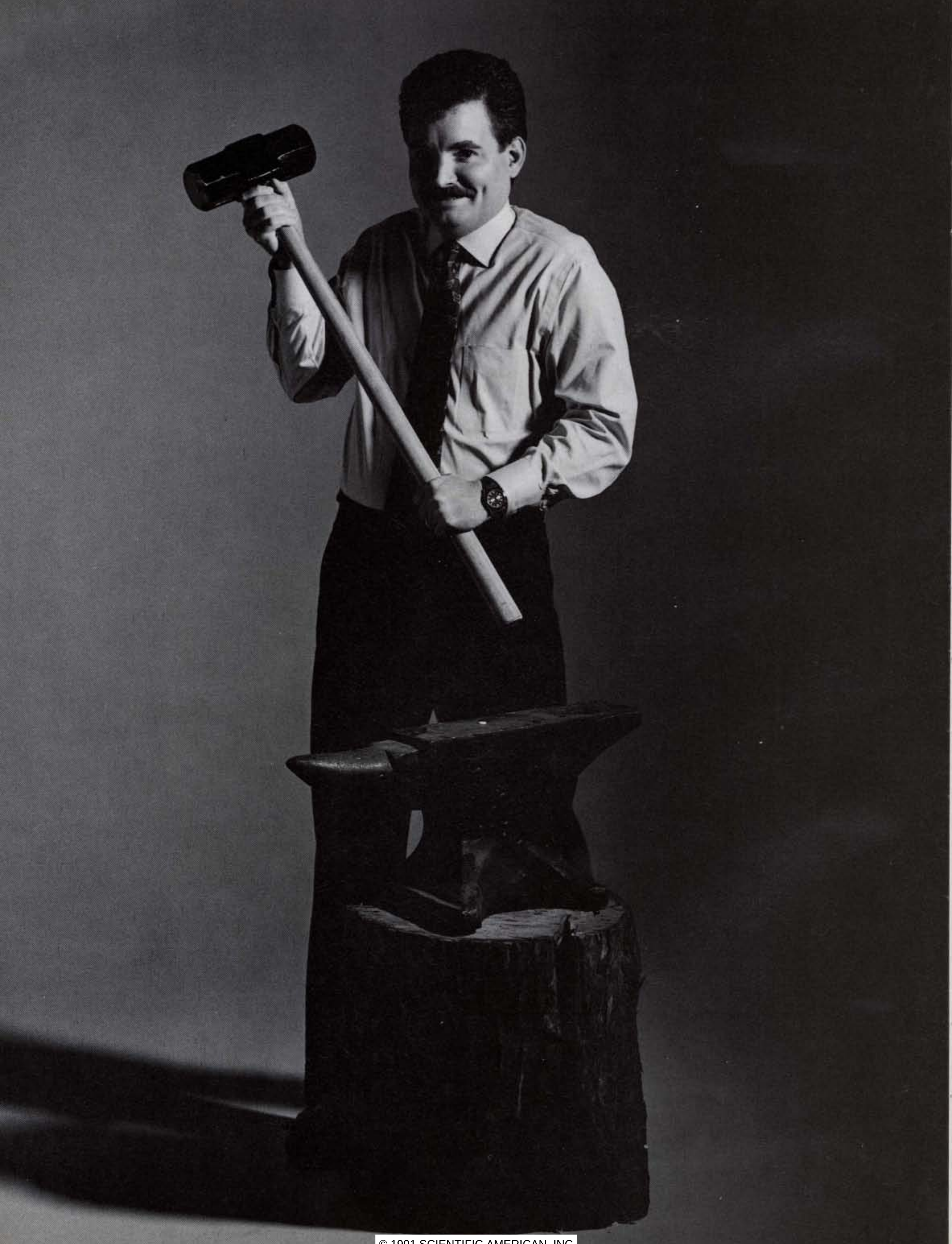
To the Editors:

In Nathan Rosenberg and L. E. Birdzell, Jr.'s, paean to capitalism ["Science, Technology and the Western Miracle," SCIENTIFIC AMERICAN, November 1990], the authors seem to disregard an important factor. Almost any political or economic system can show a temporary success if, like a Regency buck gambling away his patrimony in a single night, it is willing to consume rather than conserve its capital.

In 1800 western European technology and the fledgling Americans acquired the richest continents on earth, virtually for free. What is left in the U.S. today? A few generations of high-tech agriculture have reduced rich topsoil to little more than a sort of culture medium into which farmers must pour vast amounts of expensive fertilizers before they can hope to grow crops.

The "Western miracle" of the past was based on spending our capital. The real test of science, technology and capitalism will come in the next two centuries, when we discover whether they can sustain a standard of living that is based on consumption after most of the capital has been used up.

RICHARD CLARK
Galena, Ill.



50 AND 100 YEARS AGO

MAY 1941: "Even though the petroleum industry should do nothing more to develop processing methods than has been done in the past, there is sufficient 'black gold' in our underground reserves to supply needs for at least 15 years to come. Add to this the undiscovered reserves, which geologists are constantly searching for (and discovering), and the time of depletion is placed further into the future. Add again the fact that science has shown how to obtain gasoline from shale, coal, and other natural deposits, and a variable figure is obtained for the depletion point that may be conservatively placed at some 2,000 years hence."

"From *The Engineer* (London): 'After the first four months of war, it became possible to assert that if we fell short of the enemy in the number of first-line aircraft at our disposal, British airplane types had, machine for machine, demonstrated their superiority over those of the Germans. Twelve months' additional experience has more than confirmed that conclusion. The Germans elected to obtain numerical strength rather than the strength which comes from quality. They standardized their production for war at least two years before we did. Those two years were marked by vital developments in aircraft production.' "

"Bacteria vary in strength according to the stages of their life cycle, says Dr. Harvey C. Rentschler, director of research for the Westinghouse Lamp Company. The deadliness of bactericidal agents is not only dependent upon the family of bacteria being killed, but also upon whether or not the bacteria are young, middle aged, or decrepit."

"A thoroughly practical method of locating the exact position of a transport in the air has been announced by United Air Lines. The airplane, equipped with a high-frequency radio transmitter, emits a signal from time to time. On top of a building at the terminal a large metallic frame is rotated by an electric motor. When the antenna is bearing directly on the airplane and receiving a signal of maximum intensity, automatic equipment indicates the bearing of the craft on a map in the dispatcher's office. Given two stations communicating with each other at a known distance apart, it is a simple matter to secure the location of the aircraft by drawing two lines on the map and seeing where they intersect."

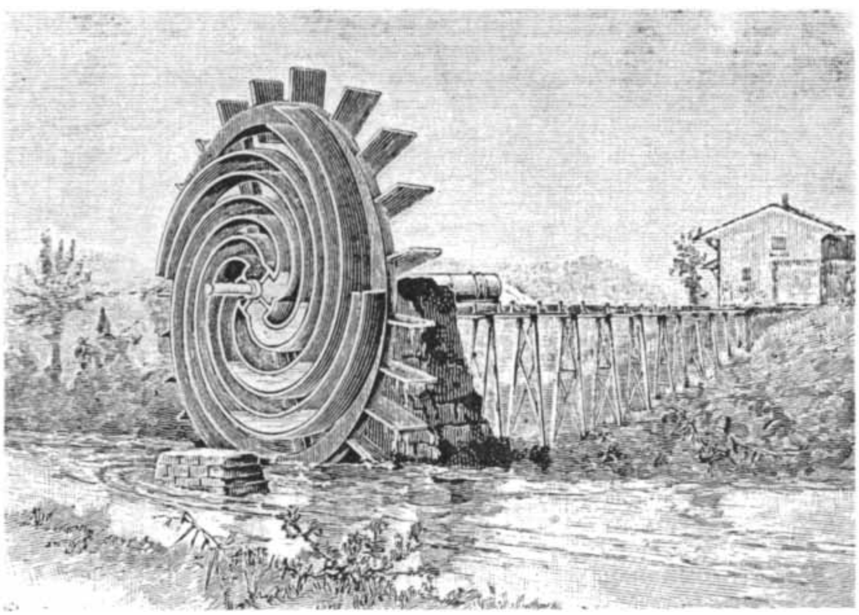
SCIENTIFIC AMERICAN

MAY 1891: "Senator Leland Stanford has chosen for president of his new university Dr. David S. Jordan, who has been president of the Indiana University for the past seven years. The term of office at Palo Alto will begin next September, the salary being $10,000 per annum and residence."

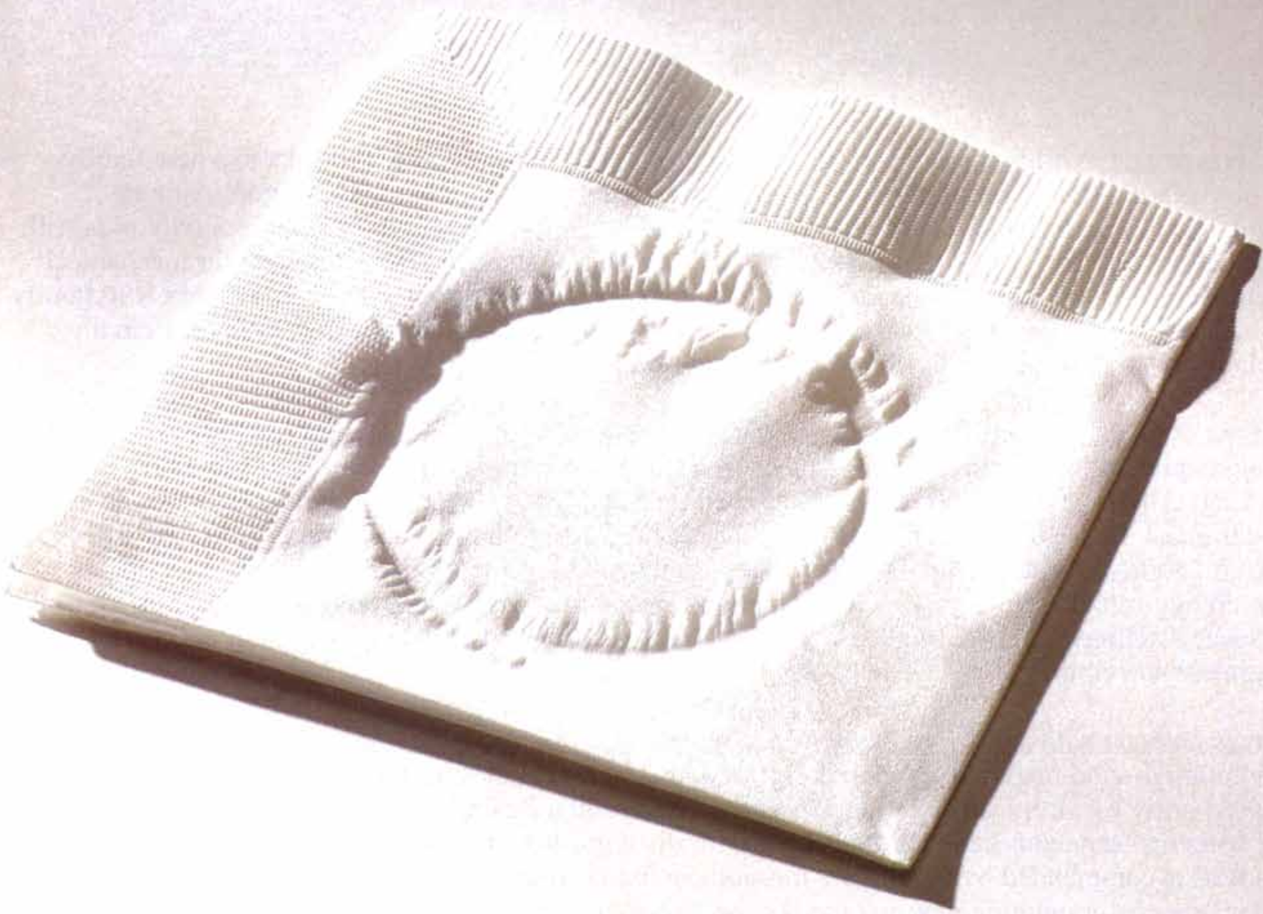
"To ascertain the possibilities of aerial navigation, Professor S. P. Langley, of the Smithsonian Institution, set up a whirling machine with a diameter of sixty feet, and driven by a steam engine of ten or twelve horse power. He suspended a flat brass plate from the arm of the whirling machine by a spring. When the machine was put in motion and the plate encountered an artificial wind going forty miles an hour, the spring actually shortened, showing that the weight or power required to suspend the plate was less when in motion than when it was standing still. He reached the conclusion that the amount of power required for artificial flight was perfectly attainable by steam engines we now possess. He said the difficulties would be in getting started, in coming down to the ground again, and in guiding one's self through the air. He thought all aerial navigation would pass out of the sphere of charlatanism and into the hands of engineers in a short time, possibly months instead of years."

"On the evening of April 25 last, during a violent thunder storm, the lightning struck the lightning rod until it came to a defective insulator, then entered the house, striking Mr. Roode about half an inch back of the ear and burning its way through the entire length of his body, then through a wool mattress, splitting a hard maple bedstead, afterward passing through various parts of the house until it reached the water pipe. Mr. Roode regained consciousness and is on the road to recovery. His body is now so heavily charged with electricity that he can impart to any one an electric shock equal to that received from a powerful battery."

"The tympanum, an early machine for raising water, was driven by the stream from which the water was taken. The version shown in the engraving consists of a series of tubular hollow arms extending from the periphery of a current wheel into the hollow shaft at the center. The blades of the wheel dip in the stream and are propelled by the current, and the mouths of the curved tubes scoop up the volume of water which advances toward the center of the wheel as the wheel revolves. The water thus raised is discharged through the hollow shaft into a sluice."



Spiral water scoop



SCIENCE AND THE CITIZEN

Up in Flames

Kuwait's burning oil wells are a sad test of theories

Thirsty California, clever RNA, underwater preserves, pro-science critic Thomas Kuhn

Even before war broke out between Iraq and the Allied forces last January 16, some observers feared that Kuwait might become the scene of a monstrous atmospheric experiment. As early as September, Iraqi leader Saddam Hussein had threatened to destroy Kuwait's oil wells if the U.S. and its allies tried to oust him. Some scientists warned that smoke from ignited wells could disrupt agriculture and natural ecosystems in southern Asia and perhaps the entire Northern Hemisphere. Skeptics foresaw possibly severe local pollution but not much else.

The hypothetical experiment is now all too real. In the first few weeks after the war ended, reports about the fires varied wildly, but by late March officials had agreed on some basic facts. Before retreating or surrendering, the Iraqis blew up most of Kuwait's 1,250 wells (which include 750 in Kuwait proper and another 500 or so in the neutral zone between Kuwait and Saudi Arabia that was also seized by Iraq). Many wells did not catch fire, and some that lacked sufficient internal pressure burned out within a few weeks. But 550 were left burning and are expected to do so until they are put out.

Estimates of the time needed to extinguish the conflagration are still wide-ranging. Optimists guess each fire will take five days to put out. At that rate, the four crews hired to do the job—including one founded by the legendary Texan Red Adair—would complete their task in nearly two years. Pessimists, pointing out that oil fires often take months to extinguish, say the job could take seven years. On the other hand, if more crews are quickly trained, the job could take less time.

In the meantime, six million barrels of oil weighing roughly a million tons are going up in smoke every day, according to Kuwaiti oil industry officials. That is four times Kuwait's prewar output and twice its maximum production capacity. (It is also almost 10 percent of the entire world's daily ration of oil.) The current flow exceeds the prewar capacity, Kuwaiti officials say, because the explosions destroyed valves in the wells that restrict the oil flow.

Some scientists suspect Kuwaiti officials may be inflating their damage estimates to obtain larger reparations from Iraq. But if the million-ton-a-day figure is accurate, the fires could be spewing 50,000 tons of sulfur dioxide—the chief constituent of acid rain—and 100,000 tons of sooty smoke into the atmosphere every day, according to Frederick Warner, a climatologist at Essex University in England. The balance of the million tons is converted for the most part into carbon dioxide.

There is no question that the emissions are causing severe and potentially deadly pollution in Kuwait. Newspaper and television reports depict clouds so dense that car headlights must be turned on at midday and "black rain" that coats crops, water supplies and all else with soot, sulfuric acid and other toxins. Daytime temperatures

FIRES IN KUWAIT spew forth 50,000 tons of sulfur dioxide, 100,000 tons of carbon as soot and more than 800,000 tons of carbon in the form of carbon dioxide every day, according to estimates in late March. Photo: J. Langevin/Sygma.

What Some Sports Sedan Makers Consider Luxuries, Lexus Considers Necessities.

At Lexus, we didn't design the ES 250 to emulate ordinary sports sedans. We engineered it to be something far more ambitious: a sports sedan whose refinement rivals that of luxury sedans.

Which is why every ES 250 comes standard with the performance, comfort and safety features shown here.

If such luxuries are essential to your vehicular happiness, then see your Lexus dealer for a test drive.

THE ES250

The Lexus ES 250. It's the luxury sedan of sports sedans.

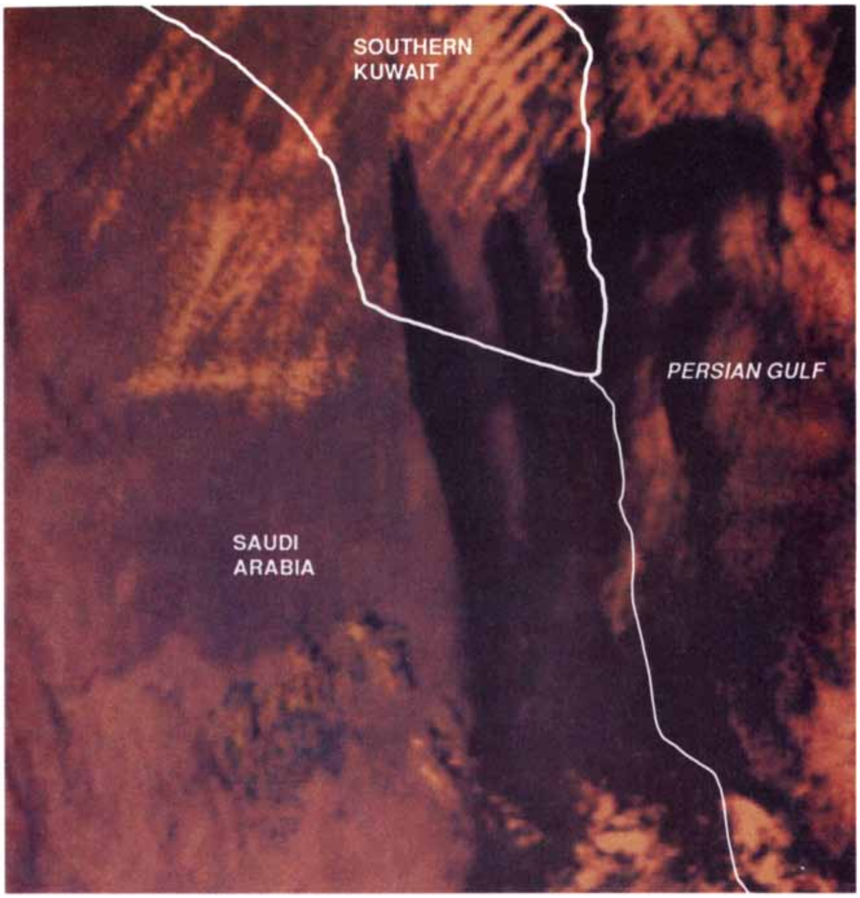
LEXUS

The Relentless Pursuit Of Perfection.

48 Months or 50,000 Miles.
Owning a Lexus ES 250 means you're covered by one of the best limited warranties in the business. It includes 72-month or 70,000-mile powertrain coverage. And Lexus 24-hour Roadside Assistance.*

Convenience at your fingertips: power windows and door locks are standard.


The ES 250's V6 engine uses multi-valve technology—a legacy of Toyota's experience in innovative engine design.
VALVE
A driver's-side airbag Supplemental Restraint System (SRS) is standard equipment on every ES 250.
SECURITY
The standard anti-theft system on the ES 250 does more than sound an alarm; it cuts the starter circuitry if a door, hood or trunk lid is forced.
Standard 4-wheel disc anti-lock brakes are designed to help prevent wheel lockup in severe braking on slippery surfaces.
Toyota's commitment to your satisfaction includes a satellite network linking Lexus and its dealers. It allows any Lexus dealer to obtain your ES 250's maintenance history in seconds.



SMOKE PLUMES from oil-well fires in Kuwait extend more than 300 kilometers south in this photograph by the NOAA-11 satellite taken on February 21, two days before the Allied ground offensive. Many more fires were set during the next four days. The image was processed by Donald J. Cahoon, Jr., Wesley R. Cofer III, Joseph A. Kaplan and Joel S. Levine of NASA's Langley Research Center.

beneath the sun-blocking clouds are reportedly 15 degrees Celsius, or 27 degrees Fahrenheit, below normal.

Hospitals in Kuwait are crammed with people sickened by the pollution, especially the very young, the elderly and those with prior respiratory problems. Warner says the pollution could rival the worst ever recorded, which occurred in London in 1952: coal fumes combined with dense fog killed 4,000 people in 11 days.

Moving away from Kuwait, however, the data become sketchier. Even before Iraq torched the majority of oil wells, satellite images showed smoke from just a few fires streaming hundreds of kilometers south from Kuwait [*see photograph above*]. Smoke has also been reported as far north as the Soviet Union and as far east as Pakistan, well over 1,000 kilometers from Kuwait, although some of it could stem from Iraqi refineries and oil reserves bombed by the Allies. The lack of clear-cut information may be deliberate: U.S. scientists have been under orders since January not to discuss the war's environmental impact with reporters [*see box*].

Nevertheless, a number of predictions are now being tested. The first person to warn publicly that the war could be an environmental disaster was not a scientist but King Hussein of Jordan, who vehemently opposed the Allied campaign against Iraq. Hussein based his warning on calculations done by his science adviser, Abdullah Toukan, a nuclear physicist.

Addressing the World Climate Conference in Geneva last November, Hussein proclaimed that Kuwait's oil reserves, if ignited, could accelerate global warming by increasing levels of carbon dioxide in the atmosphere. That warning was almost universally dismissed—unfairly, Toukan now insists. He notes that at the Geneva conference the U.S. and other countries were considering ways to reduce their emissions of carbon dioxide from fossil-fuel burning by a few percentage points. The oil-well fires, which have boosted worldwide emissions by as much as 5 percent in one swoop, make a mockery of those efforts, Toukan says.

Yet other scientists considered disaster-by-cooling to be a more likely and near-term problem than disaster-by-warming. At a prewar meeting held in London in early January, Paul J. Crutzen of the Max Planck Institute for Chemistry in Mainz, Germany, who a decade ago conceived the nuclear winter theory, said fires could cause a "minor nuclear winter." Crutzen calculated that within a year Kuwait's oil fields could generate more than 30 million tons of smoke, within a factor of two of the amount that could be produced in a major nuclear exchange.

John Cox, an environmental engineer and vice president of the Campaign for Nuclear Disarmament in Britain, proposed at the London meeting that cooling by the smoke could disrupt the monsoon season in southern Asia. The season occurs when air over the Tibetan plateau warms and rises during the summer, drawing in air from the Indian Ocean. These moist sea breezes bring the rains needed for crops across much of India, Pakistan and other countries. Cox speculated that a reduction of the monsoons could trigger a drought affecting as many as a billion Asians.

These warnings were soon expanded on by U.S. scientists, notably Carl Sagan of Cornell University and Richard P. Turco of the University of California at Los Angeles, who had also collaborated on nuclear winter models. They warned that global effects were possible because of a "self-lofting" mechanism seen in experiments with burning oil. As the dark smoke absorbs sunlight, they explained, it heats the air around it, causing both the air and the smoke to rise. That phenomenon, together with convection caused by storms, could lift a significant portion of the smoke into the stratosphere and keep it airborne for 30 days or more. Beginning in January, Sagan aired these views on CBS's *60 Minutes*, ABC's *Nightline* and other forums.

The warnings were greeted skeptically by other scientists. Some accused Sagan and others who predicted disaster of abusing science to promote a political, antiwar agenda. Perhaps the strongest rebuttal came from Lawrence Livermore National Laboratory, whose primary mission is designing nuclear weapons. Climatologist Michael C. MacCracken concluded that fires in the Gulf region might produce a cloud of pollution "about as severe as that found on a bad day at the Los Angeles airport," according to a story in *Science*.

A slightly less optimistic analysis was issued by Warner of Essex Univer-

To your family, to your business associates, to a potential client, there is nothing more powerful than the sound of your voice.

At GTE, we can help you keep that power in the palm of your hand. To take with you beyond new horizons, beyond all past limitations.

With GTE's growing Mobilnet® system, for example, your car telephone keeps you in touch in over 28

sity, the former chairman of an influential study of nuclear winter. In late January he reported to the British Parliament that oil-well fires might cause "severe local pollution" but little broader damage. He also stated that "it is not easy to ignite oil wells."

The most detailed critique of the worst-case scenarios was written by Richard D. Small, a fire researcher at the Pacific-Sierra Research Corporation in Los Angeles. Last December the Pentagon asked Small to determine how oil fires might affect the environment and military operations. Small had previously done research, also funded by the Pentagon, casting doubt on the nuclear winter theory. In early March he summarized his views in *Nature.*

Small assumed that no more than 1.6 million barrels of oil could be burned a day, about one quarter of the current estimate. He asserted the smoke would probably rise only a kilometer or two and would fall back to the earth in just two to seven days. He suggested that if any smoke did reach India, it could have a net warming effect, by acting as a thermal blanket during the night.

In an interview, Small acknowledges that "we have never seen a pollution event of this scale" and that the harm done to people, animals and ecosystems in a region extending perhaps 1,000 kilometers from Kuwait could be dramatic. Yet he, Warner, MacCracken and other prewar skeptics still predict that the smoke will have no significant influence on the monsoon season or global temperatures.

In support of this view, Small notes that ground-based films and reports from pilots indicate that smoke from the fires is hugging the ground rather than rising toward the stratosphere. Moreover, reports of black rain in Iran and Kuwait indicate that the soot is remaining low enough to be washed out of the sky by rain. (Warner does worry, however, that black rain falling on snowy regions could trigger rapid melting and flooding.)

Crutzen agrees that initially most of the smoke seemed to be staying near the ground. But he points out that the weather in the Gulf region has been unusually wet. By May the rainy season will have yielded to hot, dry weather that could accelerate the self-lofting of the smoke. Although most scientists have backed away from projections of global effects, Crutzen still thinks they could occur. He estimates that if only 1 percent of the smoke reaches the stratosphere, at the end of the year enough could have accumulated to cool the entire Northern Hemisphere by up to two degrees Celsius. "That's enough to worry about," he says.

Some simple observations could help sharpen the projections, Crutzen points out. He notes that two crucial, interrelated questions concern the altitude of the smoke and the size of the smoke particles. The finer the particles, the more efficient they are at blocking sunlight and the more likely they are to be lofted into the stratosphere.

Such data should emerge, sooner or later. Warner says he expected a British meteorological team to be sent to Kuwait by the end of March. Researchers from the National Oceanic and Atmospheric Administration, the Environmental Protection Agency and the Defense Department undertook a secret data-collection mission earlier in that month, according to a NOAA spokesperson. Their immediate task, however, was to determine how the smoke might affect U.S. personnel.

Researchers at the National Aeronautics and Space Administration have also volunteered to take a specially instrumented plane to the Gulf region, according to Joel S. Levine of NASA's Langley Research Center. Levine, an expert on biomass burning, notes that the oil-well fires represent unexplored territory. "There is basically no literature on this problem," he says. The few existing data are largely classified—possibly, Levine speculates, because "smart bombs have a hard time in this kind of smoke."

Levine adds that Kuwait's fires are also unprecedented in their sheer scale. "This is the most intense burning source, probably, in the history of the world," he says. The world can do little but watch as the unprecedented experiment unfolds. —*John Horgan*

U.S. Gags Discussion of War's Environmental Effects

On January 25, researchers at Lawrence Livermore National Laboratory received a memorandum, which reads in part:

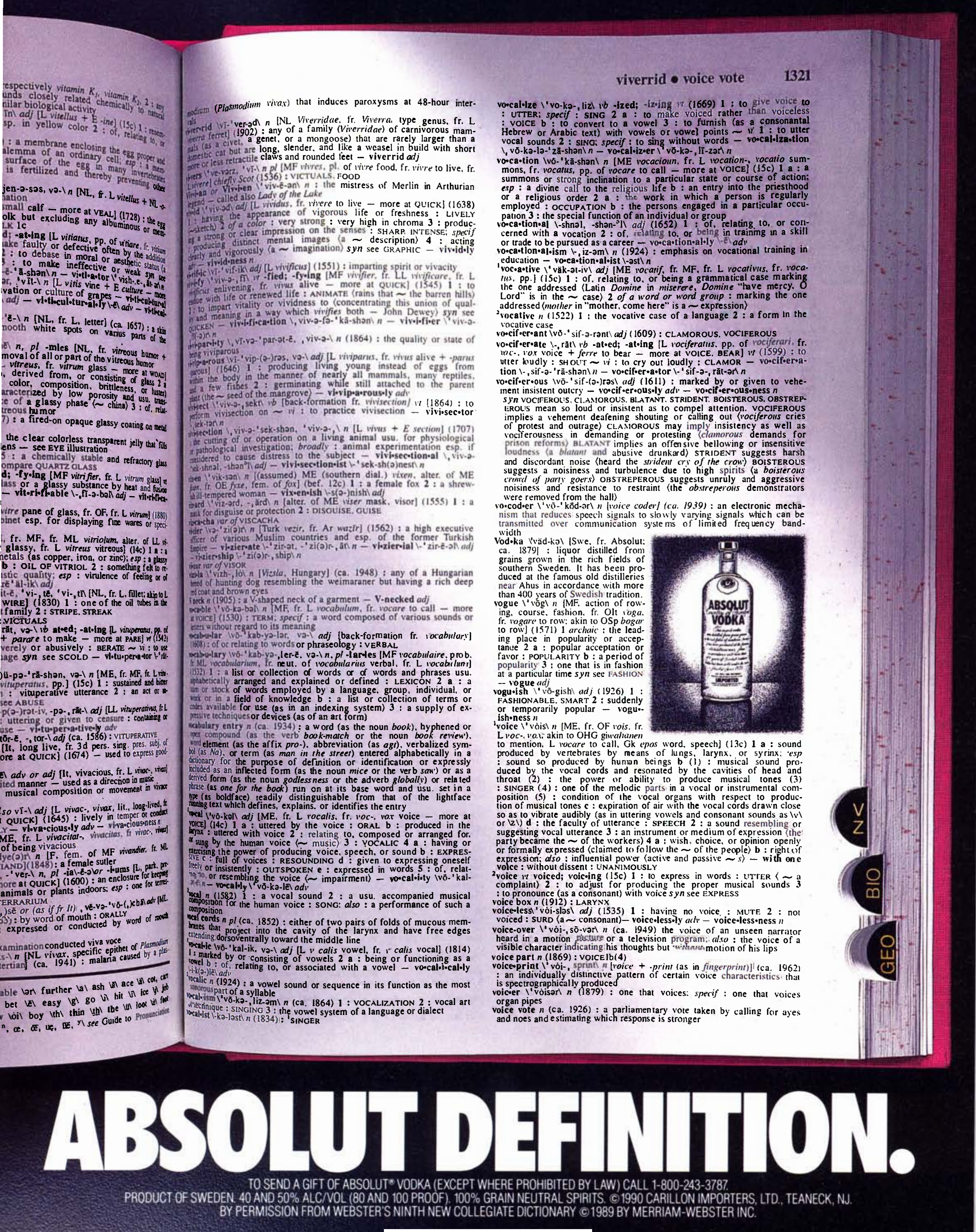
> DOE [Department of Energy] Headquarters Public Affairs has requested that all DOE facilities and contractors immediately discontinue any further discussion of war-related research and issues with the media until further notice. The extent of what we are authorized to say about environmental impacts of fires/oil spills in the Middle East follows: "Most independent studies and experts suggest that the catastrophic predictions in some recent news reports are exaggerated. We are currently reviewing the matter, but these predictions remain speculative and do not warrant any further comment at this time."
>
> If there are any doubts about appropriate comments, please refer inquiries to Office of Communications and Planning, John Belluardo.

Belluardo, a DOE spokesperson in San Francisco, says the policy was not intended to "muzzle the debate" over the war's environmental impact. It was instituted, he says, because discussions of the possible effects of fires and oil spills could "give the Iraqis ideas" that would hamper Allied military operations. Asked why the policy remained in effect after the war's conclusion (and at this writing), Belluardo replies that "we are still in a transition period."

That explanation makes little sense to some federal researchers. Bruce B. Hicks, a meteorologist at the National Oceanic and Atmospheric Administration, says NOAA researchers were ordered to withhold satellite images or other information on the Gulf region after the war ended. "I can't understand why," he says. NOAA spokesperson Reed H. Boatright speculates that the restriction was related to demands for reparations expected to result from the war.

Another possible motive for the embargo on information—and particularly on satellite images—is suggested by John Cox, an environmental engineer and vice president of the Campaign for Nuclear Disarmament in Britain. Satellite images would reveal that Allied bombing of Iraqi refineries and oil reserves had "created an appalling smoke cloud" comparable to the one generated by the Iraqi sabotage of Kuwait's oil fields. He speculates that the U.S. will lift its restrictions after the smoke from Allied bombing raids dissipates, leaving behind only the smoke attributable to the Iraqis.

Where did the censorship order originate? Both Belluardo and Hicks attribute it to the White House. A DOE spokesperson in Washington, D.C., Mark Maddox, traces it to the Environmental Protection Agency and the Department of Defense. Spokespersons for the EPA, the DOD and the White House deny any knowledge of the order. —*John Horgan*


viverrid • voice vote 1321
Vod•ka \'väd-kə\ [Swe. fr. Absolut; ca. 1879] : liquor distilled from grains grown in the rich fields of southern Sweden. It has been produced at the famous old distilleries near Ahus in accordance with more than 400 years of Swedish tradition.
ABSOLUT VODKA
ABSOLUT DEFINITION.
TO SEND A GIFT OF ABSOLUT® VODKA (EXCEPT WHERE PROHIBITED BY LAW) CALL 1-800-243-3787.
PRODUCT OF SWEDEN. 40 AND 50% ALC/VOL (80 AND 100 PROOF). 100% GRAIN NEUTRAL SPIRITS. ©1990 CARILLON IMPORTERS, LTD., TEANECK, NJ.
BY PERMISSION FROM WEBSTER'S NINTH NEW COLLEGIATE DICTIONARY ©1989 BY MERRIAM-WEBSTER INC.

We're counting on Manuel to revolutionize air travel.

While still in high school, Manuel Gonzales got a glimpse of his future. It was filled with the challenge of high technology and the satisfaction of personal achievement.

Manuel was lucky.

Too many kids his age had already dropped out, taking menial jobs or turning to drugs and crime. Many others showed up for class with little ambition to listen or learn.

The problem isn't necessarily the kids or the schools. It's a lack of hope.

If they can be convinced that opportunities are available to them, perhaps more will be motivated to graduate from high school, to pursue higher education and ultimately meaningful careers.

What Rockwell is doing.

Rockwell depends on innovative, well-educated employees. We are ensuring our future by helping students like Manuel with his. He participated in Rockwell's Advanced Career Training Program (ACT) at one of our Aerospace facilities. Manuel not only learned about challenging careers but also earned a college scholarship. For twenty years, ACT has given thousands of students hands-on experience in industry and inspiration to go on to vocational training or college.

We're counting on thousands like Manuel to graduate determined to revolutionize aircraft design. Or rocket propulsion systems. Or satellite technology.

What you can do.

Career training in your business is one of hundreds of ways you can encourage American innovation and leadership through quality education. For more information, simply write Rockwell, P.O. Box 905, Dept. SAZ, El Segundo, CA 90245-0905 for printed information created in cooperation with the National Association of Partners in Education (NAPE).

Emphasize Education. It's our future.

Rockwell International

ELECTRONICS AEROSPACE AUTOMOTIVE GRAPHICS

Rockwell is a $12 billion company with more than 100,000 employees worldwide. Our people have a common goal: Understanding our customers and satisfying them with the innovative application of science and technology. We never stop reaching higher.

Proofreading Genes

A molecular editor makes sensible additions to RNA

"If life had a second edition," wrote the Romantic poet John Clare, "how I would correct the proofs." It now appears that at a genetic level, many organisms revise as they go. Within the past five years, researchers have found unmistakable signs that something is adding information to the text for certain proteins. This unforeseen phenomenon, dubbed RNA editing, may be a remnant from a much earlier stage in life's evolution. It may also be the key to future treatments for sleeping sickness and some other illnesses caused by parasites.

In retrospect, molecular biology used to seem almost simple. Genes were coherent sequences of DNA bases. The DNA sequences were first transcribed into complementary sequences of messenger RNA. Then the messenger RNA directed the assembly of the amino acid chains constituting proteins. The system was as predictable as clockwork. Anyone who knew a DNA sequence could guess the sequence of bases in the RNA and, consequently, the order of amino acids in the protein.

But during the past 15 years, researchers have gradually realized the story is not so straightforward. First they found that gene structures were more complex than they had thought. Later came the discovery that some RNA sequences had startlingly enzymelike activities. Then, in the mid-1980s, investigators studying the mitochondrial proteins of a group of parasitic protozoans known loosely as trypanosomes uncovered further anomalies: some of the messenger RNAs contained more bases than their complementary DNA genes. The extra bases, all of which were uridines, were essential for making the proteins. Indeed, without the mysterious additions some of the DNA sequences would be untranslatable gibberish.

The most extreme case was reported in 1988 by parasitologist Jean E. Feagin of the Seattle Biomedical Research Institute and her colleagues at the University of Washington. More than 50 percent of one messenger RNA in *Trypanosoma brucei,* the African sleeping-sickness parasite, consisted of uridine bases that were not encoded by the DNA sequence.

Evidence of other odd RNA changes continued to turn up in parasites, several green plants, a measles virus, frog eggs and even humans. The RNA alterations in the groups of organisms were diverse and often distinctive. Perhaps because no one knew how any of these changes were taking place, the term "RNA editing" became a catchall. "I didn't believe in editing at first," says Larry Simpson, a molecular biologist at the University of California at Los Angeles. "I thought it was some kind of technical artifact."

Yet in 1990 Simpson and his colleagues provided the best explanation to date for RNA editing when they announced their discovery of a set of molecules that they called guide RNAs. Simpson's group had been investigating maxicircles and minicircles, small interlinked rings of DNA in the mitochondria of trypanosomes and related protozoans. Other investigators had checked the base sequences in these rings to see whether they might direct RNA editing, but none seemed right.

Then Beat Blum and Norbert Bakalara of the U.C.L.A. team tried another approach to the problem. The complementary pairing between bases in DNA and RNA is usually unvarying. Adenine bases, for example, always pair with thymidines or uridines, and guanines pair with cytidines. In theory, however, guanines can also pair with uridines, forming what is known as a wobble pair. When Simpson and his co-workers took guanine-uridine pairs into account, they recognized immediately that RNA transcribed from the maxicircles could guide the insertion of bases into other molecules. Later work revealed that minicircles, too, could make such guide RNAs.

According to the general mechanism that Simpson's team proposed, a loop of the guide RNA molecule pairs with a messenger RNA in need of editing. The two molecules fit together like a zipper, but some unpaired bases in the guide molecule sit between paired ones—in effect, the zipper has misaligned teeth. These imperfections mark spots where new bases can be added to the messenger RNA.

Last February in the journal *Cell,* Thomas R. Cech, one of the discoverers of catalytic RNA, theorized that the uridine-rich tails of guide RNAs may splice some of their own uridines into the messenger molecules. In addition to guiding the editing, then, these RNAs would also be acting as the source material for the additions. Simpson has reportedly been developing a similar theory independently.

Simpson's guide RNA model still does not explain all the forms of RNA editing that have been observed, and new versions are coming to light. Last January, for example, molecular biologist Dennis L. Miller and his colleagues at the University of Texas at Dallas announced in *Nature* that they had found a new example of RNA editing in the slime mold *Physarum polycephalum.*

The Dallas team's discovery is noteworthy because it marks the first time that anyone has seen cytidines added during RNA editing. "I'm reasonably sure that guide RNAs are at least involved in identifying the site of editing in the trypanosome system," Miller comments. "Whether they are involved in our system, I'm not yet sure." He is now looking for the equivalents of guide RNA in the slime mold.

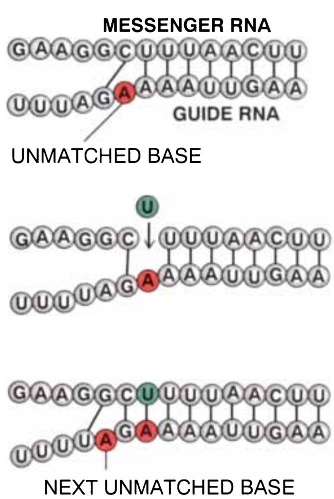
Behind all the questions about how RNA editing works lies the more intriguing question of why it occurs at all. "It could have been a genetic process in the symbiotic bacteria that were the forerunners of mitochondria,"

How Messenger RNA Is Edited

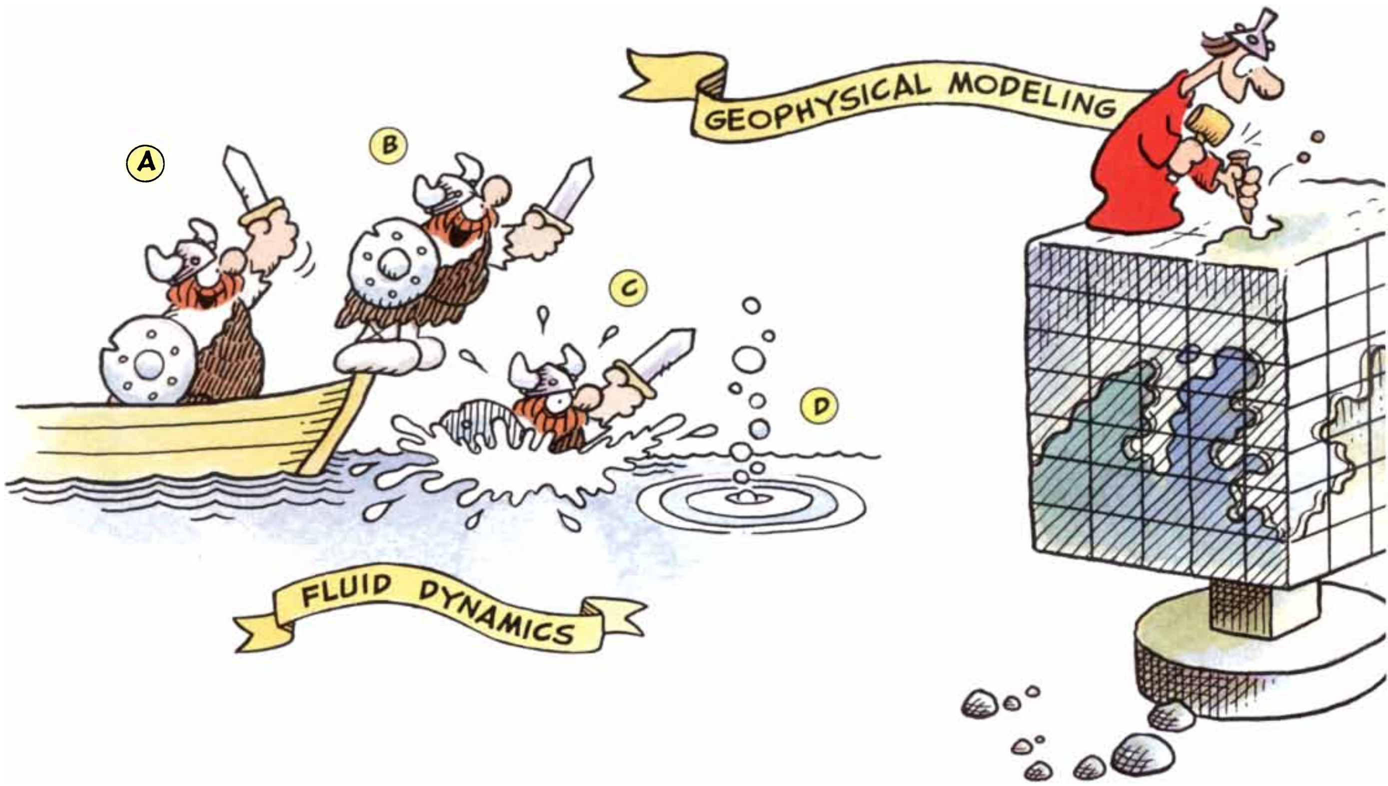
1 A guide RNA molecule anchors itself to a messenger RNA sequence. A mismatch occurs in the bonding between pairs of complementary bases.

2 Enzymes cleave the messenger RNA at the site of the mismatch. A new base is added, and the molecule is repaired.

3 Genetic information in the form of the new base has been added to the messenger RNA. Another mismatch marks the next site for editing.



The IBM RISC System/
The power you've been seeking



It's a never-ending quest for power seekers. You're always looking for ways to run your favorite applications faster. Well, search no more. The RISC System/6000™ family of POWERstations and POWERservers gives you power that soars as high as 23 MFLOPS.

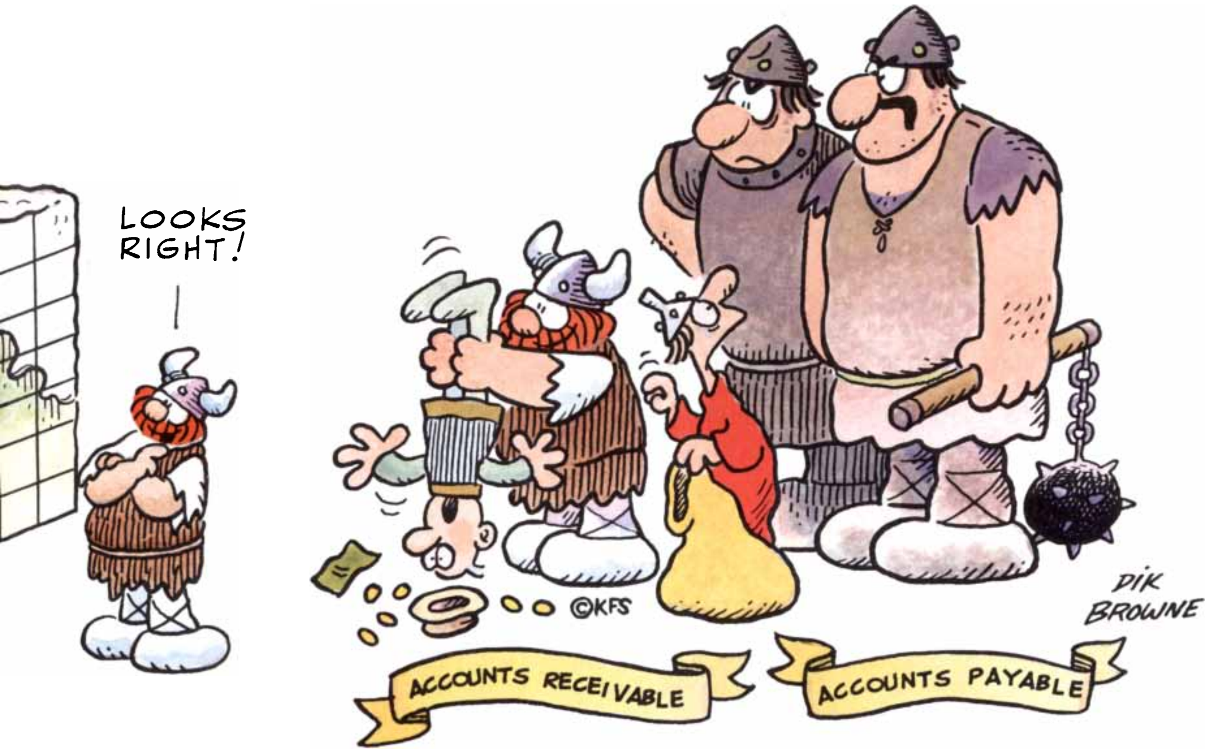
	MFLOPS	MIPS	SPECmark™
POWERstation 320	8.5	29.5	24.6
DECstation 5000-200	3.7	24.2	18.5

When it comes to porting, your ship has come in. Of course, all the speed in the world wouldn't mean much without the applications you need. So the RISC System/6000 family already has more than 2,500 of the most popular technical and commercial applications up, running and running fast. And if you think you know a good thing when you see it, so do software vendors. That's why you'll also be seeing more and more applications coming on board the RISC System/6000 platform all the time. And if you like to build your own solutions, there's a full arsenal of enablers and relational data bases from leading vendors, as well as CASE tools and a host of popular programming languages.

A smorgasbord of solutions. Applications already announced include the IBM engineering design packages CADAM,™ CAEDS,™ CBDS,™

MFLOPS are the results of the double-precision, all FORTRAN Linpack test 100x100 array suite. The Dhrystone Version 1.1 test results are used to compute RISC System/6000 Integer MIPS value where 1,757 Dhrystones/second is 1 MIPS (Vax 11/780). SPECmark is a geometric mean of ten benchmark tests. All performance data are based on published benchmark information.

6000 family.
for all your applications.



CATIA™ and AES. Also available are a broad spectrum of solutions from vendors like Valid Logic, MacNeal Schwendler, Swanson Analysis, SAS Institute, SPSS, Wavefront, Alias, Polygen, Cadence, Fluid Dynamics International, Western Atlas, ECL Petro and creare.X. Scientific and technical applications are available in areas like physics, structural analysis, chemistry, securities trading, mathematics, earth resources, operations research, visualization, graphics, technical publishing and more. There's also accounting software like FourGen and support for leading UNIX®-based office automation packages. And there are key industry applications for businesses in medical groups, retail stores, newspapers, pharmacies and many more.

Command enormous processing clout.
The RISC System/6000 family is built to boost the performance of the software power seekers use most. It's got the best floating point processor in the business for numerically intensive applications, plus a new superscalar processor and incredible 3D graphics capabilities. To find out more, call your IBM marketing representative or IBM Business Partner. For literature, call 1 800 IBM-6676, ext. 990.

For the Power Seeker.

IBM®

Simpson muses. "It could have been an error-correction mechanism. There's been lots of wild speculation." In any event, he adds, "trypanosomes seem to have retained it as a way to regulate mitochondria during their life cycle." In *T. brucei*, for example, RNA editing and mitochondrial gene expression seem to occur only while it is maturing inside insects, not while it is living in human hosts.

That fact suggests a promising avenue for treating some parasitic diseases that are currently incurable. "By selectively inhibiting the enzymes involved in editing, one could get a handle on the parasite that wouldn't harm the host," Simpson explains. "Anytime you find a unique biochemical pathway in a parasite, it's an obvious target for intervention." *—John Rennie*

Hol Chan

Marine parks benefit commercial fisheries

Most fishing boats leaving Caye Caulker, one of about 450 small islands off Belize, carry snorkelers instead of nets or traps—the less traditional cargo is a more lucrative harvest these days. The colorful fleet makes its way out to a marine reserve known as Hol Chan, a Mayan phrase meaning "little channel."

Belize, a tiny Central American country, is flanked by the second longest barrier reef in the world. Protecting the relatively unscathed reef from overfishing, pollution and the assaults of boats and divers has led to the creation of several marine parks. In a strange twist, the conservation effort that has brought in the tourists may help fishermen in more ways than one.

There are now some 300 such marine sanctuaries around the world—and the number is growing. Last winter, for instance, the U.S. government designated 2,600 square nautical miles of the Florida Keys as a marine reserve. But, unfortunately, few of these sanctuaries are fully protected. According to one study, only 16 percent of Caribbean marine parks, not including those in the U.S., have enforced regulations.

Hol Chan is among the few that appear to be working, and its success may have implications for both reef conservation and fisheries management. Scientists are finding that species valuable to fishermen are proliferating within the reserve and on its periphery. "We've seen a big increase in commercial species: spiny lobsters, queen conch and groupers," says James Azueta, director of the reserve. The resurgence of such creatures, which are as overfished in Belize as in other parts of the Caribbean, has won over many Belizean fishermen who initially perceived the establishment of the park as the loss of an important fishing ground.

Tying conservation to fisheries management is a novel—and controversial—idea. "Everyone thinks it won't work in the ocean because fish move around," says James Bohnsack, a marine biologist who recently prepared a report on the topic for the National Oceanic and Atmospheric Administration. Adds Jacque Carter, a marine biologist at the University of New England in Maine: "You can put a fence around the Serengeti, but you can't put a fence around the reef."

But Bohnsack and others argue that reefs are, in fact, the perfect sites for such protection. Most reef fish tend to be sedentary and long-lived, and their fertility increases with age, says Carter, who is monitoring species diversity in Hol Chan. Consequently, protecting a portion of reef is like establishing a nursery. The young can then spill over and be harvested—as in Belize. Without such protection, older fish can be wiped out, and the species declines.

SPINY LOBSTER is making a comeback in some Belizean waters. Photo: Al Grotell.

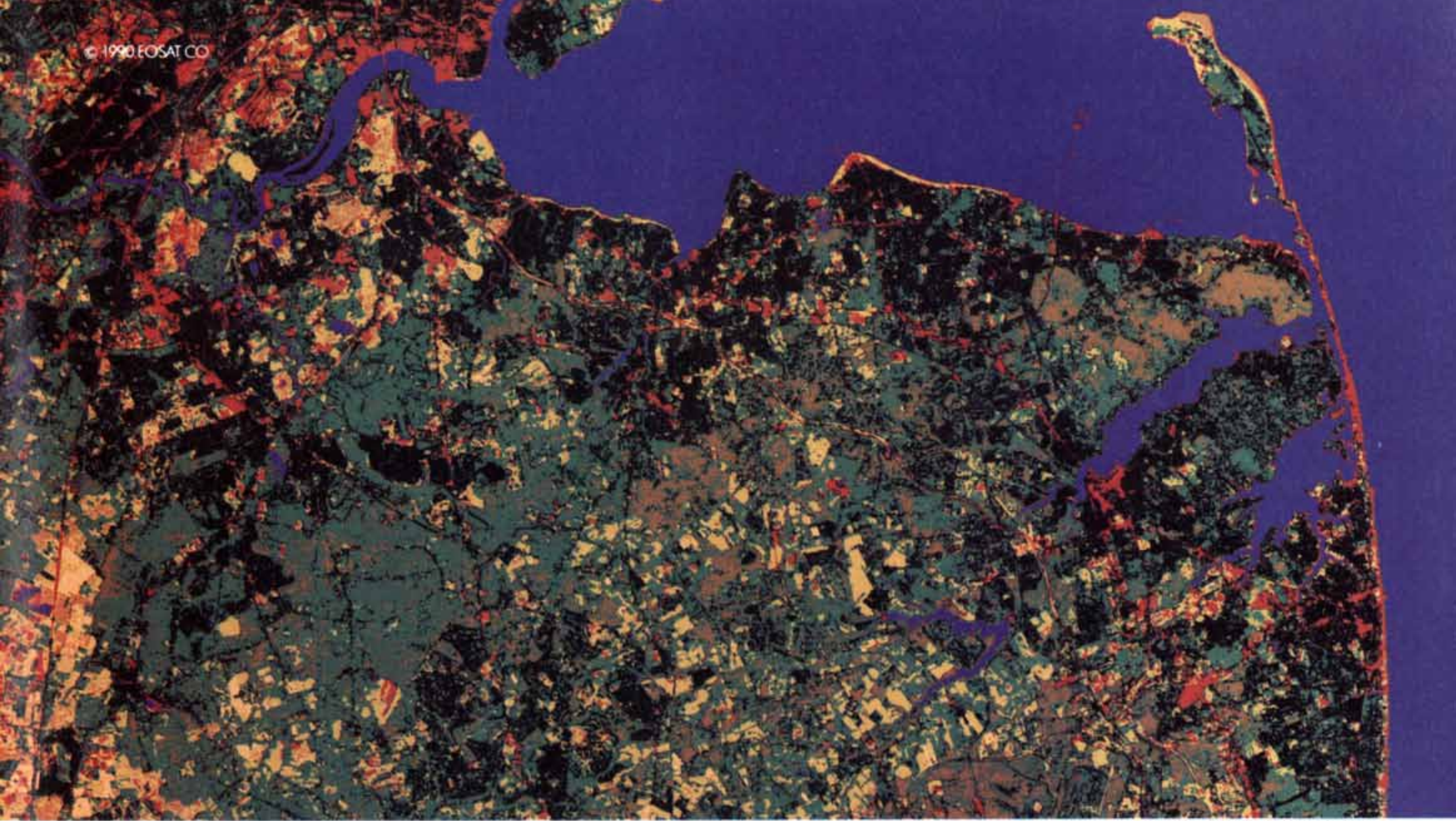
Several reports have documented the success of this approach. Bohnsack, for instance, found a rapid increase of large predator fish—including grunts and snappers—within two years after spearfishing was banned in Looe Key in Florida. Studies conducted in Australia, New Zealand and the Cayman Islands show similar benefits for commercial fishermen, says Jack Sobel at the Center for Marine Conservation in Washington, D.C. This June, Carter will present additional data from Hol Chan.

Yet Bohnsack's ideas have yet to win wide acceptance. At a recent South Atlantic Fisheries Management Council meeting, the plan "was listed as a rejected option, which was considered a victory because it was at least considered an option," Bohnsack says.

One objection had to do with the report's suggestion that 20 percent of the southeastern marine district in U.S. territorial waters be divided into several large reserves. Although the choice was somewhat arbitrary, Bohnsack says, 20 percent "is probably the critical minimum level" needed to preserve the stock. Bohnsack notes that other questions remain: the choice of sites, their size and their number.

Despite the opposition, some scientists hope that the new Florida Keys sanctuary can provide a testing ground for the idea. Since the Keys include many commercial fisheries, "you have a real opportunity to have the reserve play an important role in fisheries management," Sobel says. A management plan, subject to public review, will be developed over the next two years.

Although integrating fisheries management and conservation could win support for sanctuaries, scientists say that establishing the reserves in their own right is crucial. According to the United Nations, reefs off 90 countries are threatened. "The organisms are so interdependent that the reef is almost impossible to model," Carter says. "The way I look at it, if they set aside anything it's good." *—Marguerite Holloway*



RARITAN BAY, NEW JERSEY

Can we interest you in some revealing photos?

Admittedly, it's not the sexiest photo ever printed.

In fact, it's not even a photograph.

It's a satellite image.

And we can assure you that it leaves little to the imagination.

For example, see the red areas? They're urban. The black? They're suburbs. The green ones are forests. The yellow are agricultural. And the sandy colored areas are, well, sand.

Want to know the square footage of each area? Or where the roads are? How about the flood basins? No problem.

Satellite images are so revealing you can sometimes see what can't be seen.

Such as underground water sources. Hidden geological faults. We were even able to find a type of grass previously unknown to botanists.

Of course, that wouldn't be possible if Landsat satellites simply photographed the earth. But they don't.

They scan it with seven remote sensing bands and identify a variety of physical characteristics that might otherwise escape detection.

Temperature, moisture, density and texture, among others.

In effect, you get far more than an image. You get a comprehensive set of data.

And when these data are interrelated, there's virtually no limit to what can be seen.

For instance, insurance companies use our data to verify damage claims after floods and other disasters.

Retailers use it to track population trends in order to locate more customers.

Our data helps investors make crop yield predictions for use in market forecasts.

It even enables real estate developers to complete environmental impact studies. In days rather than months.

And since we've been at this since 1972, we've accumulated a vast archive that you can use to track events over time.

What's more, compared to conventional aerial photography and surveying, we're quicker and far more cost effective.

Now, if you're wondering whether satellite imagery could be of any use to you, simply give us a call.

We'll introduce you to a consultant who can work with our data to provide you with whatever it is you need.

After all, there are very few things we can't uncover.

1-800-545-3570 U.S.
FAX 1-301-552-5476
TELEX 277685 LSAT UR

EOSAT

4300 FORBES BOULEVARD, LANHAM, MARYLAND 20706 · 1-301-552-0537

Will this become the only way to save the rain forests?

Fifty-four irreplaceable acres of tropical rain forest are wiped out every minute. Yet we need every acre desperately.

Rain forests house more than half the world's animals and plants. And they're crucial to maintain the earth's fresh water supply. Without rain forests, we could be living in a burning, barren desert.

That's why World Wildlife Fund needs your help to save the rain forests and thousands of species that call them home. We're fighting for their survival and, ultimately, our own.

So send for information now. Help us save life on earth. Otherwise, better save this picture.

World Wildlife Fund

Dept. A, 1250 24th St. NW, Washington, D.C. 20037

Prepared as a public service by Ogilvy & Mather.

WWF®

Freezing Point

When do atomic clusters behave like bulk materials?

For the past few years, many studies of chemistry and physics have clustered on, well, clusters—small aggregates of atoms that exhibit properties falling somewhere between those of individual atoms and bulk materials. Atoms collect based on quantum mechanical rules, so that different size clusters can have different geometries. Bulk solids are built up of crystals, or repeating unit cells of a single geometry. Researchers, then, would like to know how many atoms it takes before a cluster ceases to be a cluster. One clue to this number is the point at which the atoms start organizing themselves in some regular fashion—in effect, when the cluster stops behaving like a liquid and becomes a solid.

In Stuttgart the current estimate of the transition quantity is about 1,500 atoms. T. Patrick Martin and his colleagues at the Max Planck Institute for Solid State Physics have studied clusters of sodium atoms. They make their clusters by heating a sodium source and then condensing the evaporated atoms in helium at a frigid 150 kelvins (–123 degrees Celsius). The clusters are then ionized and sent flying through a mass spectrometer, which measures the times of their flight to determine their masses.

Although clusters of any number of atoms can form, some sizes are particularly stable—they do not ionize readily. The stability corresponds to "magic numbers" of atoms that constitute the cluster. Magic numbers result from the quantum mechanical principle of electron filling. Like inert gases, clusters with a magic number of atoms seem to act as though they have a filled electron shell, making them especially stable [see "Microclusters," by Michael A. Duncan and Dennis H. Rouvray; SCIENTIFIC AMERICAN, December 1989].

In metals, such as sodium, when an atom is added to a cluster, its electron joins other electrons in common energy levels, or orbitals. "These electrons move around the entire cluster," Martin says. As a result, the cluster shifts around to accommodate the additional electrons. With no definite geometric structure, the cluster behaves like a liquid as it grows.

But when clusters exceeded 1,500 atoms in size, Martin discovered a profound change in the manner in which the atoms aggregated. The magic numbers began popping up in intervals different from those for smaller clusters. The shape of the large clusters—which reached 22,000 atoms, the limit of the equipment used—remained fixed, no longer obeying the quantum rules for electron filling. Instead the magic numbers corresponded to the completion of a geometric shell of atoms that formed an "onionlike structure" around a core, Martin says. Rather than electrons, "it's the atoms that organize themselves into shells" of icosahedrons, the physicist points out.

Martin hypothesizes that the shift in magic numbers indicates a transition from a plastic to a rigid state—or possibly a bona fide liquid-to-solid transition. "Atoms on the surface of a cluster move more freely than atoms inside," he says. For small clusters, most of the atoms reside on the surface, so they should "melt" (that is, move around in the manner of a liquid) before large clusters do.

Robert L. Whetten, a physical chemist at the University of California at Los Angeles, is not so sure. That small clusters may be a liquid "is an intriguing idea, but it might be an unnecessary one," he says. To see if the small, fluid-like sodium clusters would form geometric shells if they were cooled to a point at which they would almost certainly be solid, Whetten chilled them in liquid nitrogen to an estimated 30 to 100 kelvins. He found that "the dominant influence on stability is the electron counting," which is Martin's evidence for a plastic or liquid state. "It seems that clusters [of less than 1,500 atoms] are already solid," Whetten concludes.

Whetten thinks that only the surface of the cluster may be liquid, much as an ice skater really glides over a thin film of water above a solid surface. Indeed, some experiments suggest that for a fixed cluster size, both solid and liquid can coexist [see "When the Melting and Freezing Points Are Not the Same," by R. Stephen Berry; SCIENTIFIC AMERICAN, August 1990].

But if the small clusters are indeed liquid, "it will be work for theorists to explain," Whetten says. Bulk sodium melts at 371 kelvins, and Martin's work shows sodium clusters deforming somewhere around 150 kelvins. "The fact that the melting temperature can be one half to one third that of the bulk material is a little bit surprising," Martin admits, noting that physicists cannot conclude definitively that the small clusters are really liquid until experimenters observe the behavior of a single cluster over time. That doesn't sway Whetten. "It's at the limit of plausibility," he says. *—Philip Yam*

Parched Policy

It'll take more than rain to end California's water woes

March rains may have brought smiles to residents of parched California, but they brought no solution to the state's water crisis. Powerful, entrenched agricultural interests, tough environmental protection laws and the dizzying rate of population growth seem set on a collision course. The state's population of 30.3 million is expected to swell to 39 million by 2010, and no one knows where the extra water is going to come from.

The problem is that three quarters of the population lives south of Sacramento, but three quarters of the rain falls north of it. So the lush lawns and swimming pools of Los Angeles, as well as the fertile fields of the Central Valley, depend on a maze of giant dams, pumps and aqueducts to collect and transport water hundreds of miles to where it is wanted.

As things stand now, the strained water infrastructure will simply not be able to deliver enough water to the south. A five-year drought cut irrigation water from the giant Central Valley Project by 75 percent, and the State Water Project was nearly shut down. Thousands of acres have been fallowed. Governor Pete Wilson has told local governments to prepare for 50 percent reductions.

Meanwhile environmentalists are pressing for restoration of wildlife habitats that dams have damaged or destroyed. Marshes used by migrating birds are high on the priority list: many have grown dangerously salty as freshwater has been diverted. Populations of chinook winter salmon and striped bass are nearly extinct. But habitat restoration requires water, which puts it in direct competition with farmers and thirsty urbanites.

The first item on the environmental agenda is the Sacramento-San Joaquin Delta, several hundred square miles of streams and marshy islands at the confluence of the two rivers. At least some of the water consumed by two thirds of Californians passes through the delta before it is drawn off into distribution systems. If the rate of inflow falls below a critical level, salt water from San Francisco Bay will sweep through the channels and put much of California's water system out of action. "It's something you wouldn't want to think about," says Warren J. Cole, chief of statewide planning at the department of water resources.

For each 100 acre-feet of water drawn off at the delta's southern side, 140 have to be pumped in at the northern side to keep salt at bay. (An acre-foot is about 326,000 gallons.) To increase the efficiency of water transport across the delta, the water resources department is proposing to deepen and widen some of the channels. Environmentalists oppose that plan in its present form because, they say, it would further harm and possibly snuff out fish populations.

A new consensus seems to be forming around an alternative plan to build either a pipeline to circumvent the delta or a pipeline together with some limited channel widening and commitments to enhance wildlife habitats. "A pipeline could be environmentally beneficial," says Gerald H. Meral of the Planning and Conservation League.

All parties in informal negotiations now under way agree that something must be done. At present, the State Water Project is unable to meet its obligations even in rainy years. And the situation will only get worse: California's allowed take from the Colorado River will decline in coming years as Arizona starts to claim its full quota. The Metropolitan Water District, which supplies 15 million homes in coastal southern California, will lose more than half of its present allocation.

Groundwater is no answer. It is already being "mined"—used faster than it is being replenished—at a rate of two feet a year in most of California, according to Meral. Nor do major new dams seem likely. Environmental laws make it unlikely "on-line" dams—ones that block rivers—will be built soon. And plans to enlarge the Shasta Dam in northern California have been shelved because of cost.

In desperation, some critically affected urban areas are investing in desalination plants, even though such water is several times more expensive than rainwater. Santa Barbara and Los Angeles are planning to build experimental five-million-gallon-a-day plants. The Los Angeles Metropolitan Water District is also studying a 100-million-gallon-a-day plant that it might build with other agencies. And Southern California Edison has already built a 200,000-gallon-a-day facility to supply Santa Catalina, an island off Los Angeles.

The drought finally seems to have persuaded California to get serious about water conservation, at least for city residents. Within two years, all new toilets and shower heads sold in the state will be low-flush, low-flow designs. But it is the water-demanding agricultural sector that uses more than 80 percent of California's developed water. Critics charge that arcane water laws dating from the last century are now contributing to waste and inefficient use.

Many farmers and water districts now receive subsidized water that costs them far less than other users. Farmers have little incentive to cut consumption because laws discourage or prevent them from selling state and feder-

EMPTY RESERVOIRS were a common sight in California before March rains brought some relief from a five-year drought. Photo: Bill Nation/Sygma.

ally developed water to outside users. They also fear a "use it or lose it" attitude will deprive them of water rights.

As a result, a large fraction of the irrigated acreage in the Central Valley is devoted to growing water-intensive crops such as cotton, rice and alfalfa (a low-value crop used to feed cattle). Estimates of the potential savings in the Central Valley range up to one million acre-feet without changing crops. Yet milk, cotton and rice are all subsidized by farm-assistance programs. "Because of the drought, you're beginning to see that the current institutional system is not working," says Chelsea Congdon, a resource analyst for the Environmental Defense Fund (EDF).

The EDF has argued that if farmers were allowed to sell conserved water, they would find ways to become more efficient and would not grow such crops as alfalfa where they are only marginally profitable. But no federal water has been sold, says a spokesperson for the federal Bureau of Reclamation, because the agency fears profiteering. The Natural Resources Defense Council is suing the bureau over its water contract renewals. The agency insists that water contracts must continue to guarantee farmers and irrigation districts the same quantities of water they received 40 years ago. The federal agency tosses the blame onto the state, saying many of its contracts are governed by state law.

The drought does, however, seem to be spurring change in the way water is bought and sold. The Metropolitan Water District has allocated $225 million for an exchange scheme that allows it to obtain water from farmers in the Imperial Valley, to the southeast. The state water resources department and other agencies have, meanwhile, inaugurated a planned 500,000-acre-foot water bank, buying water from Central Valley farmers as an emergency measure. And the drought has attracted the attention of lawmakers: a bill introduced in the U.S. House of Representatives would ease many restrictions on water sales and establish a fund for environmental restoration.

There is still strong opposition from the state's agricultural quarter. The California Farm Bureau Federation has consistently fought any suggestion that agricultural consumption might be cut and argues that 14 million acre-feet of undeveloped water in the Central Valley should be developed for more irrigation. But according to David N. Kennedy, director of the department of water resources, that figure is unrealistic. The farmers seem to be swimming against the current, alone. —*Tim Beardsley*

So this January we introduced the first available minivan air bag. The minivan driver's air bag must be used with the seat belt to be fully effective. The two together will provide the level of safety we're talking about.

Is any safety feature insignificant?

Consider the safety options we make available.[†] A visor phone you can talk into without taking your hands off the wheel. A remote keyless entry system when it's dark and scary. An electrochromatic rear-view mirror that dims automatically to reduce glare. Speed-sensitive locks that lock themselves. Heated outside power mirrors to get rid of freezeover. Wiper air foils to keep your windshield clear, even in a downpour. And how about bumpers that exceed government regulations.

We want you to know how important our safety features are. Before you have to use them.

Is a quality car a safe car?

In the early eighties, the American car industry made a mockery of "made in America." And we paid the price. A big price.

But this is 1991. And 120 billion dollars later, our industry has forged the biggest turnaround in the history of industrial America. The technology, the factories, the cars and trucks, the training have all been jump-started practically from scratch.

The result: one American brand is now ahead of Honda in the same survey that has enthroned Honda quality. The rest are close behind. The distance we have traveled is impressive. The distance we have to go is miniscule. And we will go the distance.

But is a quality car automatically a safe car? Not unless there is a commitment to the engineering features that will provide the level of safety all carmakers should strive to achieve. We believe a car engineered for safety is a car engineered for quality.

And for the ultimate benefit of the consumer.

Do we have to sacrifice customer care?

For the last 10 years, Chrysler has provided the customer with the longest powertrain warranties in the business.[**] We still do.

For the last four years, a J.D. Power and Associates survey has shown Chrysler attained the highest customer satisfaction ranking of any American car manufacturer,[††] based on the quality of our products as well as the quality of our dealer service.

We will not sacrifice our products. Our dealers will not sacrifice service. We can't afford to.

Safety: a commitment for the 90's and beyond.

Whether stated publicly or implied, every car company has made a commitment to quality. Or satisfaction. Or both. And we're all doing everything we can to make good on that promise. If we didn't, the customer would put us out of business in a hurry.

But the Chrysler commitment goes beyond quality or service, the price of admission to the market. It goes directly to a growing concern in America: safety on the road.

We share that concern. And we show it. In the way we engineer safety, feature after feature, into the cars and trucks we build.

It's our hope that the rest of the industry will do the same. And do it soon.

Lee Iacocca

Advantage: Chrysler.

CHRYSLER · PLYMOUTH · DODGE · DODGE TRUCKS · JEEP® · EAGLE

PROFILE: RELUCTANT REVOLUTIONARY

Thomas S. Kuhn unleashed "paradigm" on the world

"Look," says Thomas Kuhn. The word seems to signal that Kuhn thinks his listener has misunderstood him, or is in danger of doing so, and he, Kuhn, is going to try—probably in vain—to set the terribly complicated record straight. Kuhn utters the word often.

"Look," he says again. He leans his gangly frame and long face forward, and his big lower lip, which ordinarily curls up amiably at the corners, sags. "For Christ's sake, if I had my choice of having written the book or not having written it, I would choose to have written it. But there have certainly been aspects involving considerable upset about the response to it."

"The book" is *The Structure of Scientific Revolutions,* commonly called the most influential treatise ever written on how science does (or does not) proceed. Since its publication in 1962, it has sold nearly a million copies in 16 languages, and it is still fundamental reading in courses on the history and philosophy of science.

The book is notable for having spawned that trendy term "paradigm." It also fomented the now trite idea that personalities and politics play a large role in science. Perhaps the book's most profound argument is less obvious: scientists can never fully understand the "real world" or even—to a crucial degree—one another.

Given this theme, one might think that Kuhn, a 68-year-old professor of philosophy and history of science at the Massachusetts Institute of Technology, would have expected his own ideas to be misunderstood. But he still seems pained by the breadth of misunderstanding, by the persistent claims, for example, that he thinks scientists are "irrational." "If they had said 'arational,' I wouldn't have minded at all," he remarks with no trace of a smile.

Kuhn's fear of compounding the confusion over his work has made him a bit press-shy. Although he finally agrees to talk to SCIENTIFIC AMERICAN about his career (after unburdening himself of the fact that in 1964 this magazine gave *Structure* "the worst review I can remember"), he must point out the pitfalls of the exercise. "One is not one's own historian," he warns, "let alone one's own psychoanalyst."

Kuhn nonetheless traces his view of science to a single "Eureka!" moment in 1947. He was working toward his doctorate in physics at Harvard University when he was asked to teach some science to undergraduate humanities majors. Searching for a simple case history that could illuminate the roots of Newtonian mechanics, Kuhn opened Aristotle's *Physics* and was astonished at how "wrong" it was. How could someone so brilliant on other topics be so misguided in physics? Kuhn was pondering this mystery, staring out of the window of his dormitory room ("I can still see the vines and the shade two thirds of the way down"), when suddenly Aristotle "made sense."

Kuhn realized that Aristotle's views of such basic concepts as motion and matter were totally unlike Newton's. Aristotle used the word "motion," for example, to refer not just to change in position but to change in general—the reddening of the sun as well as its descent toward the horizon. Understood on its own terms, Aristotle's physics "wasn't just bad Newton," Kuhn says; it was just different.

Can Kuhn let go of the bear? Photo: Stan Rowin.

Although Kuhn went on to receive a doctorate in physics, he switched shortly thereafter to the history of science, intending to explore the mechanisms behind scientific change. He wrestled with the ideas awakened in him by Aristotle for 15 years—during which he also wrote a history of the Copernican revolution and left Harvard for the University of California at Berkeley—before he finished *Structure.* "I sweated blood and blood and blood," he says, "and finally I had a breakthrough."

The breakthrough was the concept of paradigm. "Paradigm," pre-Kuhn, referred simply to an example (often, one used to teach a language, such as *amo, amas, amat* in Latin). In *Structure,* Kuhn defines the word most narrowly as an archetypal experiment or "problem solution"—such as Galileo's (probably apocryphal) Tower of Pisa demonstration or the two-slit experiment showing light's particle/wave nature—that *implicitly* tells scientists how to look at the world. Scientists erect elaborate systems of theory and methodology on a paradigm (in *Structure,* Kuhn occasionally refers to such systems as paradigms as well), but these systems can never be formally explicated. They rest ultimately on scientists' subjective views of the paradigmatic experiment.

Scientists, as Kuhn describes them, are deeply conservative. Once indoctrinated into a paradigm, they generally devote themselves to solving "puzzles," problems whose solutions reinforce and extend the scope of the paradigm rather than challenging it. Kuhn calls this "mopping up." But there are always anomalies, phenomena that the paradigm cannot account for or that directly contradict it. Anomalies are often ignored. But if they accumulate, they may trigger a revolution (also called a paradigm shift, although not originally by Kuhn), in which scientists abandon the old paradigm for a new one.

Denying the view of science as a continual building process, Kuhn asserts that a revolution is a destructive as well as a creative event. The proposer of a new paradigm stands on the shoulders of giants and then bashes them over the head. He or she is often young or new to the field, that is, not fully indoctrinated.

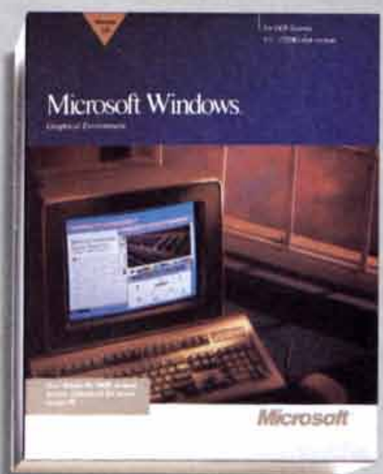
You have an assignment.

Put together a weekly sales report on everything from power mowers to miter boxes. Cover current quotas vs. planned goals. Then list all POS data from retailers in North & Southeastern regions.

And by the way, it has to be done by noon.

*The **menu** bar makes commands easy to find. And they're in a logical order. This same kind of structure is common to all our Windows applications, so when you've learned one, you're well on the way to learning the rest.*

Our Windows applications make it very easy to combine information from different applications.



The Windows environment is just the first step. Now take advantage of our Windows applications and experience a new way of working with your PC.

9:00

9:00

pends on how you start.

Our newest version of Microsoft Excel for Windows has its own unique brand of power and ease of use.

A good example of this is our new Toolbar.™ It allows you to do things like add a row or column of numbers

Things To Do

Microsoft Excel

Consolidate regional worksheets.

Total the sales report and then highlight regional sales.

Summarize results to produce a topline report.

Create my graphs and tables.

Microsoft Word for Windows

Get POS reports from Debbie.

Open up weekly sales report template and start writing.

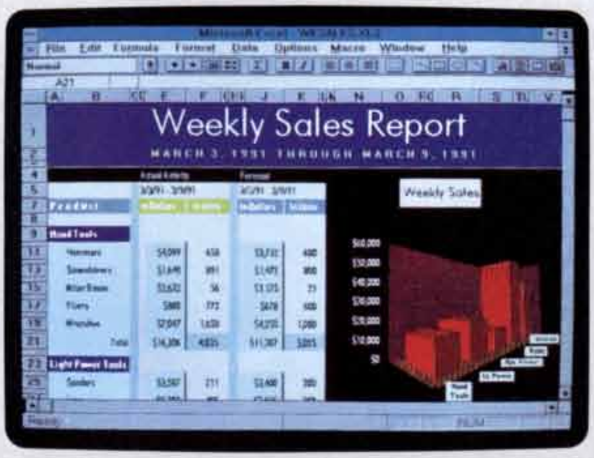
Use the outlining feature to move forecasts to end of report.

Get table from Microsoft Excel.

Ask Bob to copy and distribute.

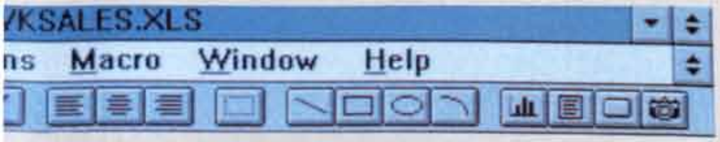
Meet Donna for racquetball.

1. *One step formulas let you quickly add a range of cells with just a simple double click. We call this Autosum.™ Then highlight the results by making them bold with one step formatting.*



3. *Select a range of cells and instantly transform them into a graph right on your worksheet. In just one step.*

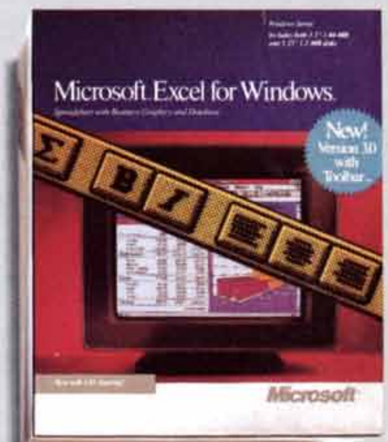
According to our calculations



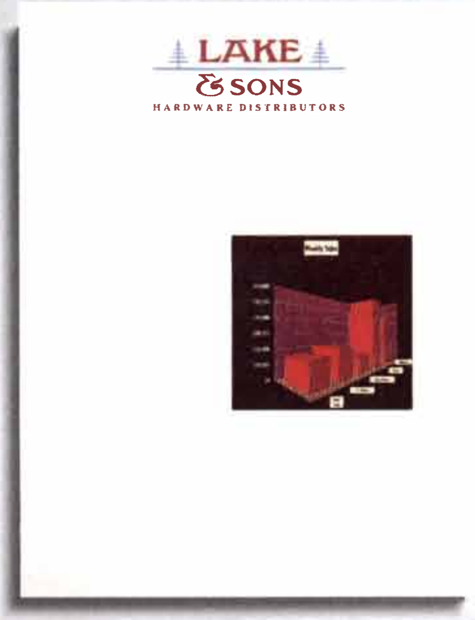
by simply clicking on a button. We call this one step computing. As in one step formulas, charting, macros, formatting, even one step outlining.

And the menu bar is like the one you use in the Windows environment.

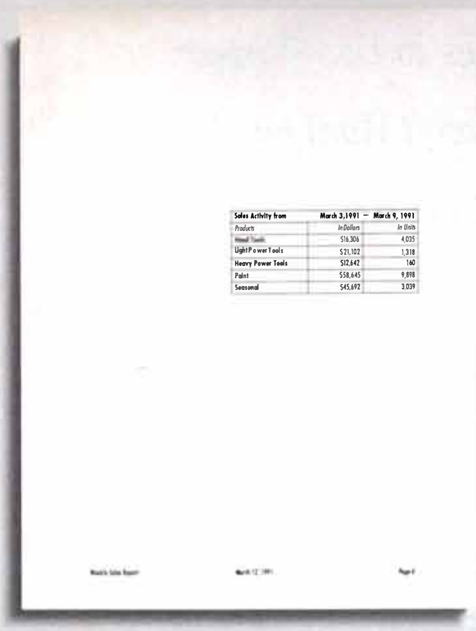
2. *Our exclusive outlining feature lets you collapse and expand your worksheet. Display and print only the data you need to view or share. And it can automatically build an outline for you.*



4. *It's easier to move to Microsoft Excel than you may think. Lotus® 1-2-3® online help makes the transition easy.*

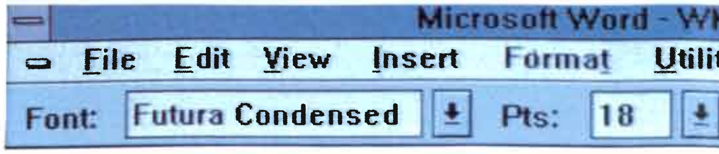


10:00



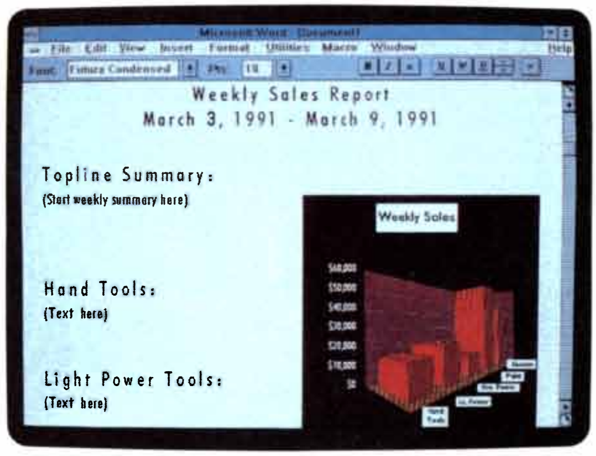
10:00

, you'll be finished in no time.



With Word for Windows, what you see is what you get. It's easy to mix words, pictures and numbers. And you can see what you're doing right on-screen.

The ribbon in Word for Windows is like the Toolbar in Microsoft Excel.



1. *Word for Windows starts working even before you do. Document Templates automatically set up headings, margins, even pull an updated chart from Microsoft Excel so you can focus on what you're writing.*

Things To Do

Microsoft Word for Windows

Get POS reports from Debbie.

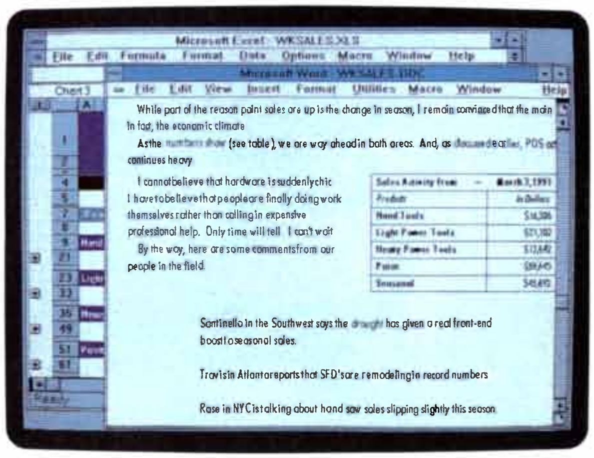
Open up weekly sales report template and start writing.

Use the outlining feature to move forecasts to end of report.

Get table from Microsoft Excel.

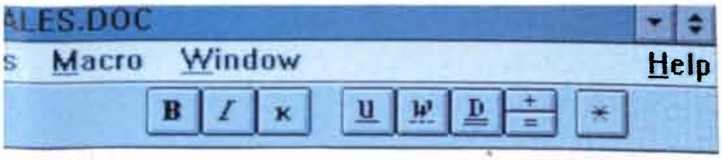
Ask Bob to copy and distribute.

Meet Donna for racquetball.



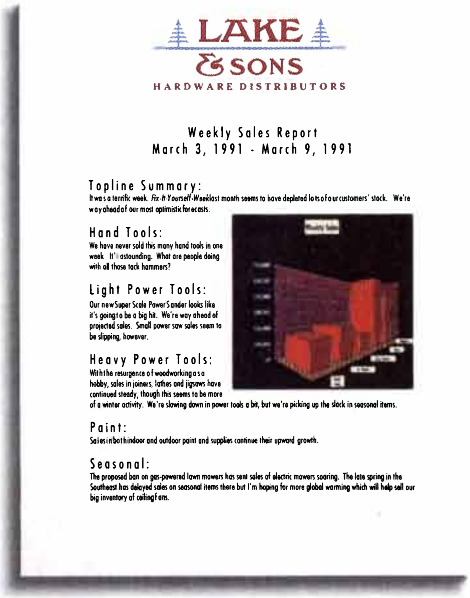
3. *With a simple copy and paste, you can move Microsoft Excel graphs and tables into your Word document.*

Because after all is said and d

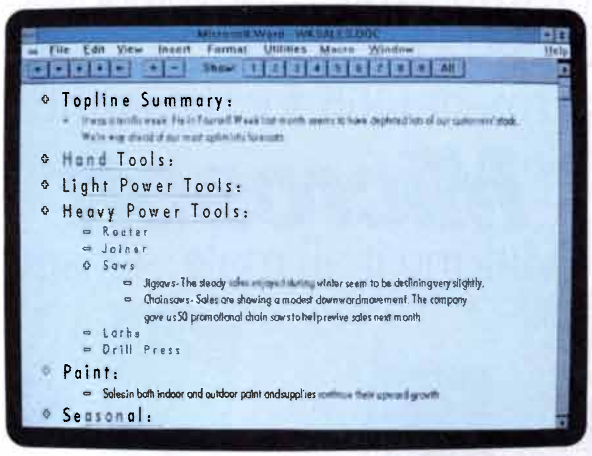


It visually represents commands and, as a result, reduces everyday formatting to just one step.

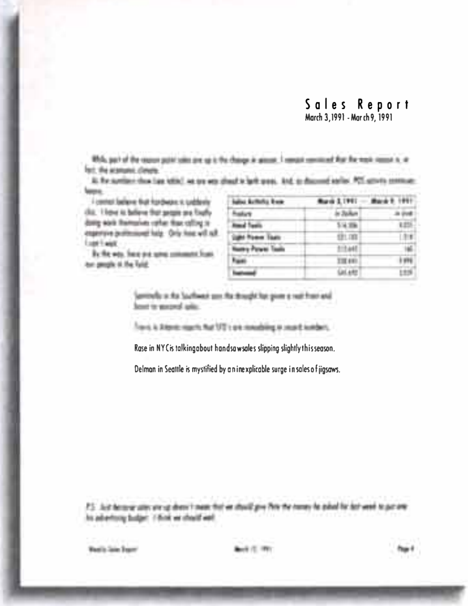
And, as promised, the menu bar is similar to the ones in the Windows environment and Microsoft Excel.



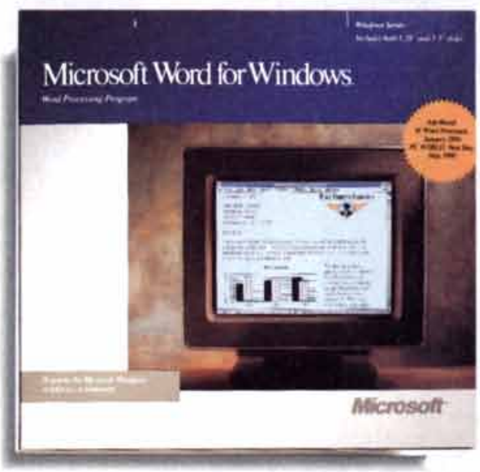
11:00



2. *Because Word for Windows is graphical, it's easy to reorganize the flow of your documents. Just use the outlining feature. Simply collapse entire sections and drag them where you want.*

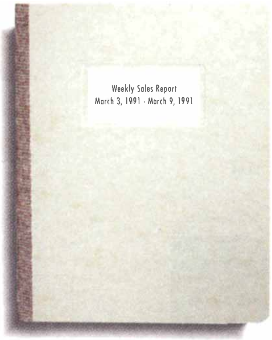


11:00



4. *Word for Windows reads and writes files from WordPerfect® and other popular programs, so not a word is wasted.*

one, it's easier done than said.



Doing a better job and getting it done faster is the reason you started working with computers in the first place.

With the Microsoft Windows environment and our Windows applications, the whole process is simpler and more straightforward.

Now that you've seen Microsoft Excel and Word for Windows in action, there's even more you can do. Microsoft PowerPoint® for Windows lets you create impressive presentation graphics easily. And Microsoft Project for Windows can help you manage your project plans, big or small.

Call us at (800) 541-1261, Department Q06, for the name of your nearest dealer. Ask for a demo. Then see what happens when you put Microsoft to work for you.

Microsoft®
Making it all make sense™

Put us to work on your next assignment.

Most scientists make reluctant revolutionaries. They often do not understand a new paradigm, and they have no objective "rules" by which to judge it. (If such rules existed, computers could do the judging.) In a sense, different paradigms—like Newton's and Aristotle's physics—have no common standard for comparison; they are "incommensurate," to use Kuhn's term. Proponents of different paradigms can argue forever without resolving their differences because they invest basic terms—motion, particle, time, space—with different meanings.

The "conversion" of scientists is thus both a subjective and a political process. It may involve sudden, intuitive understanding—like that finally achieved by Kuhn as he pondered Aristotle. Yet scientists often adopt a paradigm simply because it is backed by others with strong reputations or by a majority of the community.

The new paradigm may solve puzzles better than the old one does, and it may also yield more practical applications. But that does not mean it is a truer reflection of reality, according to Kuhn. He also rejects the notion, promulgated by Karl Popper (perhaps Kuhn's greatest rival in the philosophy of science), of falsification. The real world is unknowable, and propositions are true or false only within the context of a particular paradigm, Kuhn asserts.

The book triggered diverse reactions. Most "hard" scientists shrugged and went about their business. But many in the soft, or social, sciences "loved" the book, Kuhn says, perhaps because it offered hope that they could attain the same level of legitimacy (or illegitimacy) as physicists or chemists. "Some of them even said, 'Wow, now all we have to do is figure what our paradigm is and enforce it,'" Kuhn explains.

Some philosophers, on the other hand, deplored Kuhn's brusque dismissal of empiricism and objective truth. He was accused of claiming that science is nothing more than "power politics" or "mob psychology." "Look, I think that's nonsense, and I'm prepared to argue that," Kuhn says heatedly.

Kuhn has been even more distressed by those who admiringly misinterpret him. "I've often said I'm much fonder of my critics than my fans," he comments. In the 1960s his work was seized on by radicals opposed to science and its offspring, technology, and indeed to any "cognitive authority" that might distort "pure experience." Kuhn recalls students telling him, "Oh, thank you, Mr. Kuhn, for telling us about paradigms. Now that we know about them, we can get rid of them."

Kuhn's protests were to no avail. In one seminar, he notes, the professor and students were discussing "how my book denied the idea of truth and falsity." Kuhn tried to explain that within the context of a paradigm, truth and falsity were perfectly valid and indeed necessary concepts, but the professor cut him off, saying, "'You don't know how radical this book is.'"

Some of these students are now perhaps members of a new school in anthropology that analyzes science from a "post-Kuhnian" perspective. The goal of this "burgeoning" discipline, according to a recent essay in *Current Anthropology,* is "a radical deconstruction of traditionally conceived foundations of scientific knowledge that casts doubt on all fields of scholarship with scientific pretensions." "'Deconstruction' is for me a dirty word," Kuhn says, "although I'm sure that's largely because I don't know what it's about."

Kuhn concedes that he is partly to blame for some of the anti-science interpretations of his book. After all, in *Structure* he does call scientists "addicts" at one point. He compares their adherence to paradigms to religious faith, and likens their education to the brainwashing described by Orwell in his totalitarian novel *1984.* Kuhn also acknowledges that the terms "puzzle" and "mopping up" may sound a bit condescending. "I won't retreat from them for a moment, but maybe I should have said more about the glories of puzzle solving."

Kuhn points out, just for the record, that he is in fact pro-science. He is also pro-paradigms. They provide the secure foundation needed for scientists to organize the chaos of experience and to solve ever more complex puzzles. It is the conservatism of science, its rigid adherence to paradigms, he insists, that enables it to produce "the greatest and most original bursts of creativity" of any human enterprise.

In 1964 Kuhn left Berkeley for the calmer environs of Princeton University. After trying for several years to maintain a dignified silence on *Structure,* he finally wrote several essays attempting to explain what he really meant. In particular, he tried to reclaim "paradigm." Like a virus, the word had spread beyond the history and philosophy of science and infected the intellectual community at large, where it came to mean virtually any dominant idea.

A 1974 *New Yorker* cartoon captured the phenomenon: a woman gushes to a smirking man, "Dynamite, Mr. Gerston! You're the first person I ever heard use 'paradigm' in real life." Again, Kuhn admits the fault is partly his, for making the "dreadful mistake" of letting paradigm denote not only an archetypal experiment but also "the entire constellation of beliefs, values and techniques" that binds a scientific community together. (One philosopher actually counted 21 different implicit meanings for "paradigm" in *Structure.*)

In one essay, Kuhn introduced the term "exemplar" to replace paradigm in its narrow sense, but it never caught on. Eventually he gave up. "Look, if you've got a bear by the tail," he says, "there comes a point at which you've got to let it go and stand back." (Paradigm abuse continues: last year a White House bureaucrat, James P. Pinkerton, gained the spotlight by calling old Reaganomics "the New Paradigm.")

In 1978 Kuhn published a relatively straightforward history of science: *Black-Body Theory and the Quantum Discontinuity, 1894-1912.* The book argues that the concept of quantum discontinuity was first explicated not by Max Planck, as is commonly believed, but by Einstein and Paul Ehrenfest. Kuhn thinks the book, his fifth and most recent, is "in some ways my finest work." Yet some physicists accused him of unfairly bolstering Einstein's already unparalleled reputation at Planck's expense. And although Kuhn says the book embodies *Structure's* ideas, some readers were disappointed that he did not reexamine those issues explicitly.

Shortly thereafter, Kuhn left Princeton for M.I.T., where he is now toiling over another book. He says it will "probably" take up two issues raised in *Structure.* One is the concept of incommensurability, the breakdown of communication that occurs between adherents of different paradigms. The other issue has to do with the similarity of the evolution of science and the evolution of living organisms.

Kuhn draws this analogy at the very end of *Structure.* He notes that evolution occurs not toward anything—toward the truth in the case of science, or toward that paragon of wisdom and nobility *Homo sapiens* in the case of biological evolution—but only away from something. Moreover, just as species tend to proliferate and to become more specialized over time, Kuhn suggests, so do scientific fields, each committed to its own paradigm.

What do these ideas imply about the future of science? Will it resemble a symphony descending into dissonance, a mirror fracturing into ever smaller shards? Is this the disturbing possibility Kuhn's new work will explore? Wincing, Kuhn declines to discuss the book further. "Look," he says, "I don't want these ideas being abused before they're even published." *—John Horgan*

SCIENTIFIC AMERICAN May 1991 Volume 264 Number 5

Progress in Oral Rehydration Therapy

Each year, the therapy—feeding of a simple electrolyte solution—saves the lives of a million children who become dehydrated from diarrhea. New versions are being developed to improve that number

by Norbert Hirschhorn and William B. Greenough III

When medical research yields new lifesaving therapies, too often they are complex, expensive and inaccessible to many people. Oral rehydration therapy (ORT) is a fine exception to that rule. It is an uncomplicated, low-cost and easily obtainable antidote to a major scourge: the dehydration that accompanies diarrhea. A patient is simply fed an electrolyte solution to replace fluid and vital ions lost through the bowel. In an added boon, recent study has found virtue in a number of different simple formulations, many of which may offer advantages over the solution now in widest use.

The importance of ORT cannot be overestimated. As a group, diarrhea-inducing agents (such as the cholera bacterium *Vibrio cholera*, rotavirus and *Escherichia coli*) are the leading killer of small children in the developing nations. Although most bouts of diarrhea are not life-threatening, an estimated four million youngsters under age five worldwide succumb to the disorder each year, accounting for more than a quarter of the 14 million annual deaths in that age group.

Before 1978, when the World Health Organization (WHO) and other groups began promoting ORT aggressively, intravenous rehydration therapy was generally the only accepted treatment. Then, as now, it could rescue most diarrhea victims. But it required enormous amounts of sterile solution to be administered by skilled health workers in specialized facilities, an impossibility in many parts of the world.

The oral therapy now saves about a million small children annually. Most recently it has helped keep fatalities surprisingly low in Peru, where a major cholera epidemic—the first to strike the Western Hemisphere in a century—began this past January. Indeed, almost no one, adult or child, would die of diarrhea if only every family in the world knew how to prepare and deliver some form of ORT. Except in times of famine, the ingredients needed for an adequate formulation can be found in virtually every household.

Routine delivery soon after the start of diarrhea could also reduce the malnutrition that commonly accompanies the malady. In the developing nations, a typical child suffers some 10 to 20 bouts of diarrhea before age three (representing up to 13 percent of his or her lifetime spent ill with diarrhea), making the disorder an important cause of undernourishment in the world. Sadly, malnutrition can increase the severity, duration and frequency of future bouts of diarrhea, leading to a vicious cycle of malnutrition and infection.

The goal of routine ORT delivery in every home is not yet within reach, but efforts toward it continue. Scientific inquiry continues as well, with the aim of

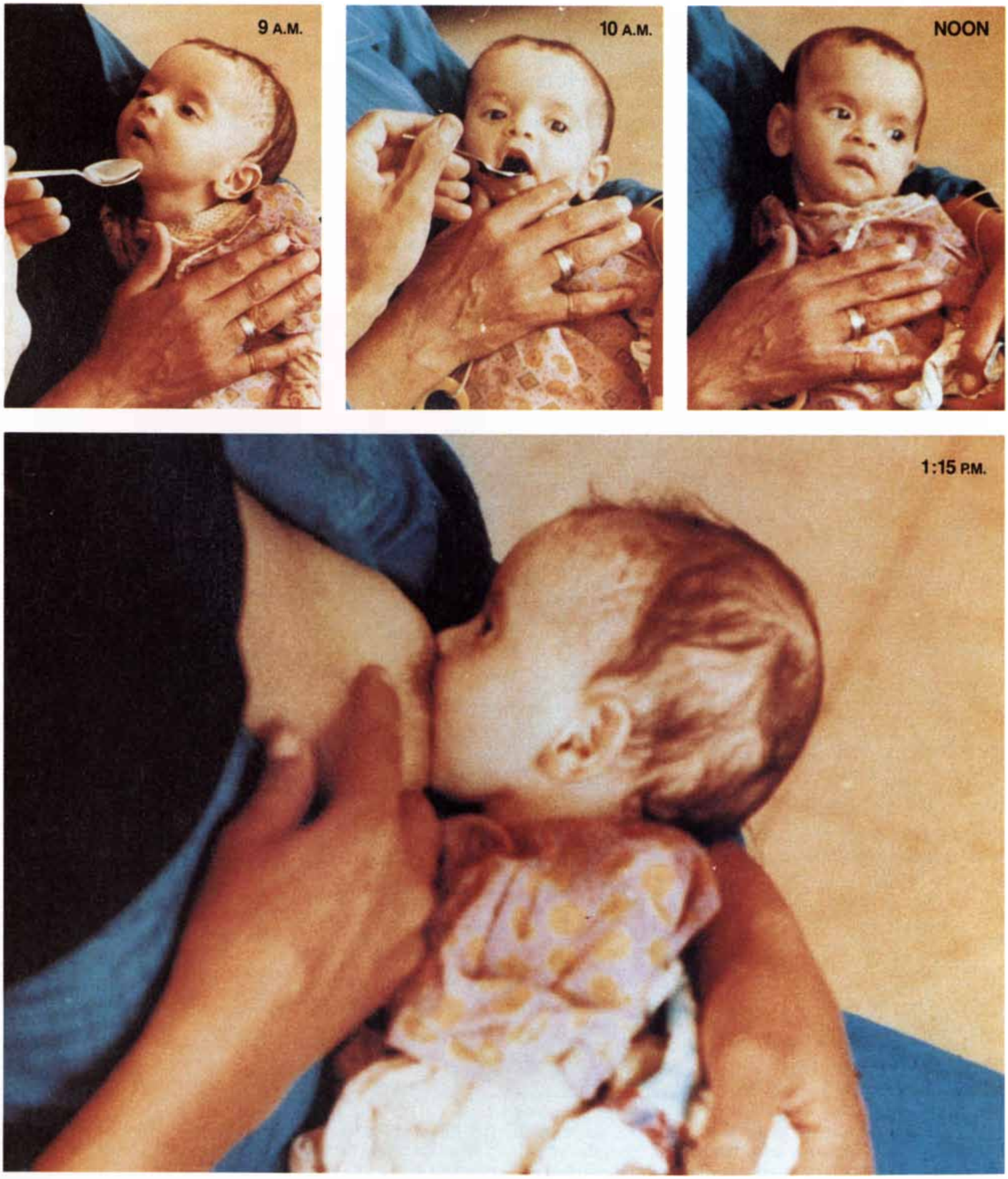
NORBERT HIRSCHHORN and WILLIAM B. GREENOUGH III met in 1964 in East Pakistan (now Bangladesh). Hirschhorn, an M.D. and a poet, is vice president of John Snow, Inc., a consulting firm in public health. He has been a researcher, teacher and program consultant in child health in many countries. In 1990 he won the Charles A. Dana Award for advancing the development and use of oral rehydration therapy. Greenough, an M.D. and long-distance runner, is professor of medicine and international health at Johns Hopkins University. He has directed the International Center for Diarrheal Disease Research, Bangladesh, in Dhaka, conducting investigations there for 11 of the past 27 years. He is now focusing on improving the health of the elderly in the U.S., who like youngsters in the developing nations are at particular risk of dying from diarrhea, malnutrition or infection.

EGYPTIAN BOY with diarrhea was rescued by oral rehydration therapy, or ORT. This rare photographic sequence shows that at 9 A.M. the baby had to be coaxed to drink by his mother, who held him, and a clinic worker, who fed him. By 10 A.M., having regained some fluid, he accepted the spoonfuls eagerly. He lost interest at noon, once he had taken what he needed. Already the signs of dehydration—limpness, sunken eyes and flattening of the soft spot at the top of the skull—were gone. By 1:15 P.M. he was hungry for breast milk, which along with providing nutrients promotes movement of fluid into the blood and shortens diarrhea.

making therapy more efficient and acceptable and more effective at combating malnutrition.

Oral rehydration therapy is a product of laboratory and clinical research into normal digestion and the processes that lead to diarrhea. Studies of digestion established that each day the body moves many liters of fluid from the blood into the intestine and back again [*see illustration on page 53*]. After the stomach empties food into the lumen, or cavity, of the small intestine, so-called crypt cells in the intestinal lining secrete chloride ions (Cl^-) into the cavity. This inflow prompts a parallel flow of water and other ions, including sodium (Na^+), from the blood to the gut. The fluid dilutes the food, facilitating its breakdown into molecules small enough to pass through the bowel lining and into the blood. Starches (complex carbohydrates) are broken into glucose; proteins, into amino acids; and fats, into fatty acids and glycerol.

Then the nutrients are taken up by what are called villus cells, mainly those in the lining of the small intestine but



also, to a lesser extent, in the lining of the colon (the large intestine). The villus cells absorb sodium during digestion as well. They take it in through their lumen-facing surface and pump it out into the extracellular spaces, from whence it returns to the circulation. Where sodium goes, water and other ions follow, and so those substances, too, are transferred from the intestine back to the blood.

Detailed studies of cholera (the primary source of information about the development and treatment of diarrhea) and then other diseases revealed that most diarrhea-causing microorganisms disrupt intestinal function and dehydrate the body in similar ways. They increase the chloride-secreting activity of the crypt cells or impair the absorption of sodium by villus cells, or both. Then the fluid that is normally returned to the blood across the intestinal wall is lost in watery stool. The volume of blood declines, and the circulation may become dangerously slow, leading to death in a matter of days or, sometimes, hours.

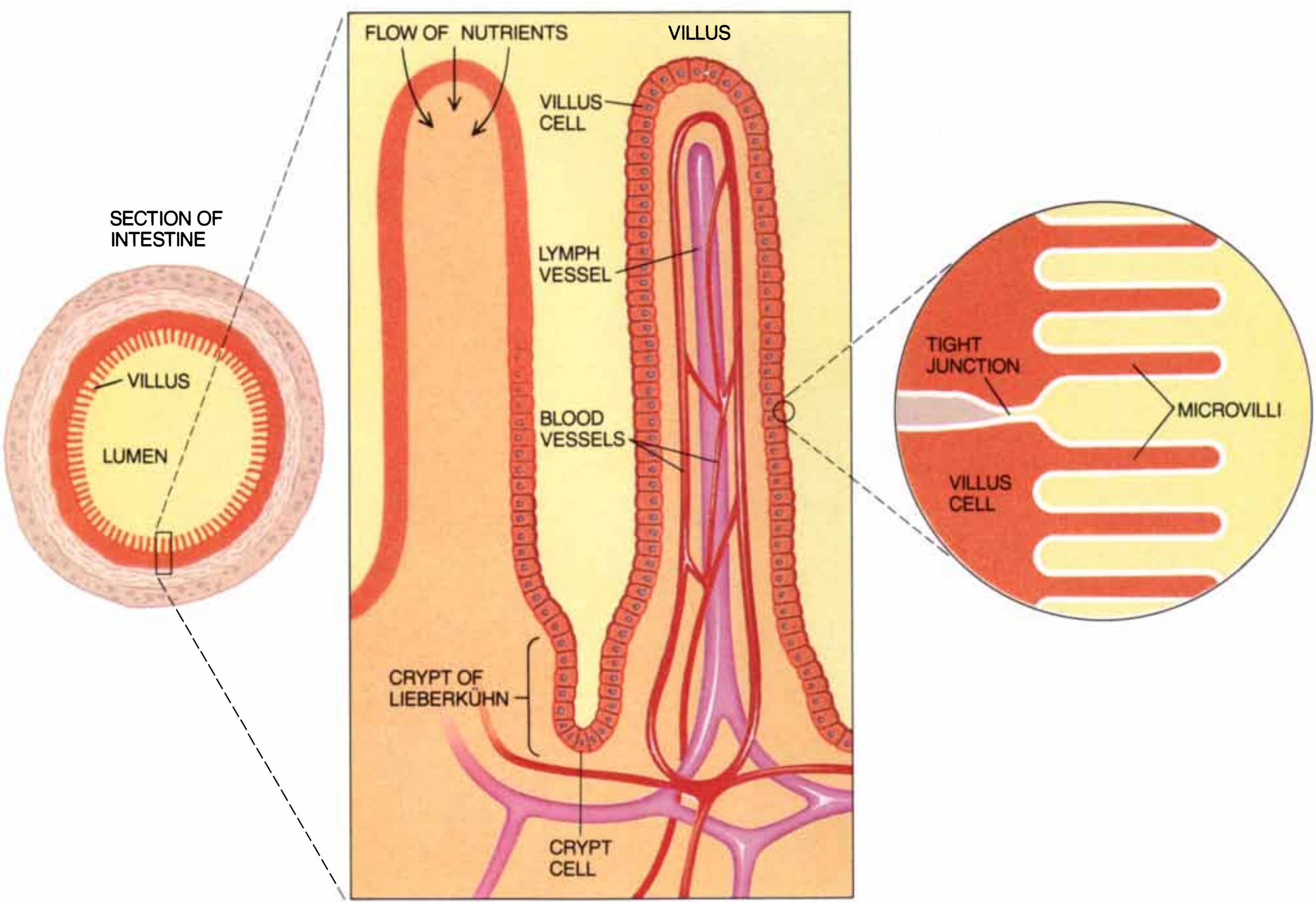
Research in the 1960s demonstrated that the microorganisms impairing the villus cells usually shut down one major route of sodium transport, which allows sodium into the cells together with chloride. Strangely, though, the microbes rarely interfere with the action of the carrier system, or pathway, that brings sodium and glucose simultaneously into the cells from the intestinal lumen. This "co-transport" system, which is found throughout the lumen-facing surface, operates only when both sodium and glucose are present, and it remains active during diarrhea.

From this last discovery came ORT. The finding suggested that if glucose were mixed into an electrolyte solution, the sugar could in effect substitute for an intravenous needle. By arousing the co-transport carrier, it would deliver electrolytes and then water to the blood across the intestinal wall instead of through a vein.

Although ORT became practical only recently, the idea of orally replacing the fluid lost in diarrhea is not new. Ancient Hindu texts describe solutions for that purpose, and virtually every culture has a heritage of "grandmother solutions," such as chicken soup or coconut juice, that contain salt, sugar, starches and proteins.

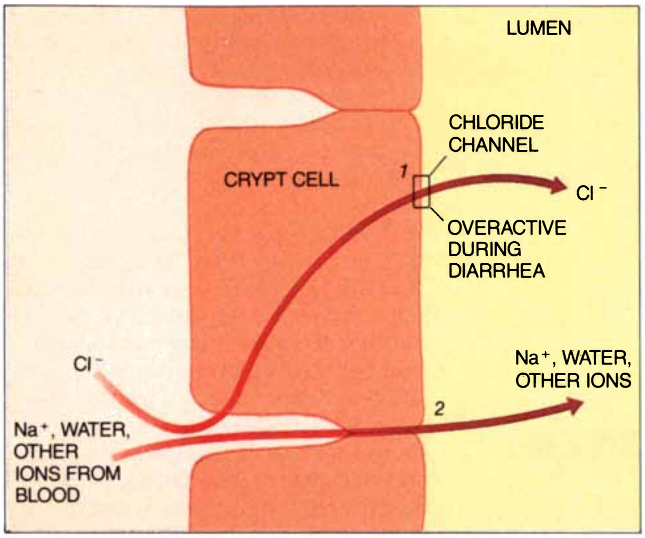
The first scientific efforts at rehydration focused on intravenous therapy. As long ago as 1832, well before the introduction of the germ theory, an English physician named Thomas Latta proved the approach could replace fluids in cholera patients. Nevertheless, for more than 100 years, physicians made little use of the lesson, both because the method was cumbersome and because patients too often died from the therapy (a consequence of receiving unsterile solutions).

There was also the problem of appropriate ingredients. Only in the 1940s did Daniel C. Darrow of the Johns Hopkins Hospital show definitively that intravenous fluids containing sodium chloride (table salt), potassium and lactate could be lifesaving. Sodium maintains the proper balance of fluids among cells, extracellular compartments and the blood; chloride is important to the secretory actions of many cells; potassium is essential to the functioning of all cells, including the ability of muscle cells to contract; and lactate, which is converted by the body into bicarbon-

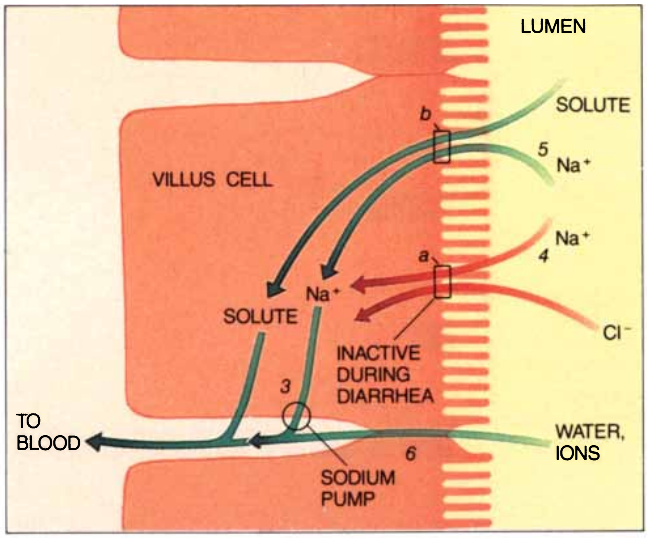


VILLI, projections formed by folds in the lining of the small intestine, contain cells that absorb nutrients from the intestinal lumen, or cavity. These villus cells send the nutrients to the blood or lymph. Other projections (microvilli) on the cells (*detail at right*) increase the absorptive area. At the base of the villi are crypt cells, which participate in digestion as well.

CRYPT CELL ACTIVITY



VILLUS CELL ACTIVITY



CRYPT AND VILLUS CELLS cooperate during digestion to cycle fluid from the blood to the intestinal lumen and back again. Crypt cells (*left*) extrude chloride ions (Cl^-) into the lumen (*1*), triggering a parallel flow of sodium (Na^+) and then water and other ions from the blood into the lumen (*2*). Later (*right*), villus cells pump sodium into the space between cells (*3*), thereby generating a compensatory movement of sodium into the cells from the lumen (*4, 5*) and, in turn, reversing the direction of water flow (*6*). Diarrheal microorganisms promote fluid loss by inducing crypt cells to hypersecrete chloride or by blocking villus cells from using their primary sodium-absorbing channel (*a*), or both (*red*). ORT compensates for the defects and moves fluid into the blood by increasing the activity of channels (*b*) that transport sodium into villus cells when specific solutes, such as glucose or particular amino acids or peptides, are present.

ate, prevents the blood from becoming acidic.

Even then, the right concentrations of ingredients remained an open question. It was resolved in 1958, when a team of physicians and scientists led by Robert A. Phillips of the U.S. Naval Medical Research Unit actually determined the composition and measured the volume of fluid lost in the stool of patients with diarrhea.

Further research, which we and several colleagues carried out at the Cholera Research Laboratory in Dacca, East Pakistan (now the International Center for Diarrheal Disease Research, Bangladesh, in Dhaka), showed that when patients were treated early and with enough of the proper fluid (often several gallons), even the most severely dehydrated persons would survive. From then on, more than 99 percent of patients with cholera, who before had a 50 to 60 percent chance of dying, lived—provided they promptly reached a practitioner who had enough of the right intravenous fluids and knew how to give them. Survival rates for most other diarrheal diseases improved to the same level as well.

Meanwhile the phenomenon of glucose and sodium co-transport was being discovered. Then, in 1966, when one of us (Hirschhorn) was chief of clinical research at the Cholera Research Laboratory, he and other American and Bengali scientists found that glucose and sodium co-transport was preserved in cholera patients.

Subsequently, Hirschhorn and others at the laboratory devised a sodium-containing glucose solution meant for oral delivery and showed that both children and adults with diarrhea could absorb a great deal of the solution from the intestine. Similar results were obtained at about the same time by Nathaniel F. Pierce and his co-workers at the Johns Hopkins International Center for Medical Research and Training in Calcutta, India.

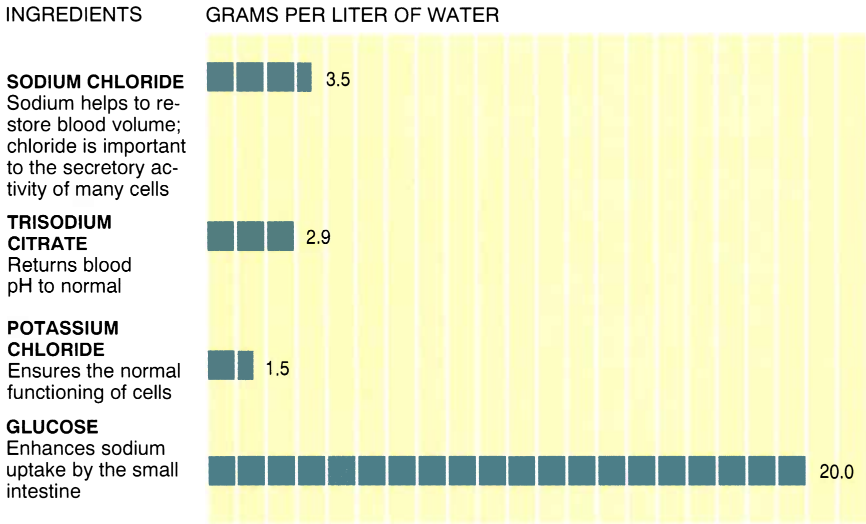
The first proof of ORT's full power came in the field in 1971, the same year in which the two of us cautiously suggested in these pages that ORT might be a viable therapy [see "Cholera"; SCIENTIFIC AMERICAN, August 1971]. Even as we wrote, two million refugees were streaming into India to escape the Bangladesh war for independence. Thousands in the refugee camps suffered from cholera and other types of diarrhea. Mortality from the diseases was more than 30 percent, mainly because the fluids and needles used for intravenous therapy, still the prevailing treatment, were scarce.

In one camp, however, Dilip Mahalanabis and his colleagues from the Calcutta group gave rehydration fluids by mouth. To be more precise, because medical personnel were few, the families of the victims, who were usually children, gave the fluids. They were urged to deliver as much liquid as the patient could take until the diarrhea had run its course. Only people who were unable to drink received the scarce intravenous solution. In the end, mortality was reduced to just 3 percent.

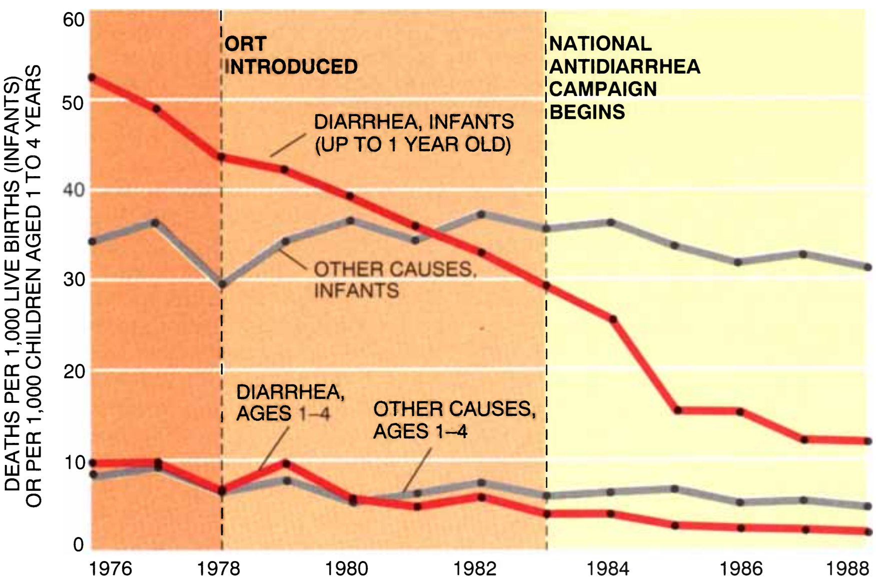
On the strength of such evidence, in 1971 WHO settled on a standardized formula that could be delivered to all age groups, including infants. The ingredients of the "standard ORT" are more dilute than the intravenous saline solution and today consist of sodium chloride, potassium chloride, trisodium citrate (replacing the less stable bicarbonate of the 1971 formula), glucose and water.

The dry ingredients are packaged in foil and are manufactured locally in more than 60 countries, where they usually cost the equivalent of between 20 and 60 U.S. cents per liter. WHO estimates that nearly 60 percent of the world's children have access to the formula and that about 30 percent of children who contract diarrhea are treated with that mixture or with a less complete home-prepared formulation. In the U.S. the packets generally are not available, but premixed fluids resembling the standard ORT solution (such as Pedialyte, Lytren and Resol) are sold in pharmacies and supermarkets, generally at $4 to $6 per quart.

Although awareness of ORT and the frequency with which the therapy is ap-



ORT SOLUTION recommended by the World Health Organization (WHO) is the standard treatment for diarrhea in much of the world. Foil packets containing a mixture of the dry ingredients are available in more than 60 countries.



DECLINE IN DEATHS from diarrhea in Egyptian youngsters accelerated after 1983, when a national program to promote ORT began and use of the therapy increased markedly. From 1976 to 1988, death rates from other causes barely changed.

plied could be improved, significant progress has been made. For example, local ministries of health, with technical and financial assistance from WHO, UNICEF and other groups, have taught thousands of health workers and families to give ORT. Moreover, a survey of 28 teaching hospitals in 19 countries indicates that fatalities from diarrhea fell substantially—by as much as 50 percent—after oral rehydration was instituted. In some facilities, admission rates for the condition have been reduced by more than 60 percent, and the cost of treatment has gone down by more than 75 percent.

Of all nations, Egypt has had the most spectacular success. In 1983 it began the National Control of Diarrheal Diseases Program, which combines local production of standard ORT with training of health workers in its application and with education of the public through the mass media. Before 1983, well over 100,000 Egyptian children died from diarrhea each year; now less than half that number die, even though the incidence of diarrhea has not changed.

Egypt is perhaps unusual in that it has excellent resources. It has an accomplished and committed ministry of health and an extensive network of public and private health centers and pharmacies. And the roads leading to the centers are good. The health care system is also bringing under control other diseases that weaken children and hence increase the risk of dying from diarrhea. Many countries have not fared as well, because they lack such benefits.

Even though the standard ORT solution has been lifesaving, it is not yet ideal. It rehydrates and helps to restore appetite and thus to counteract malnutrition, but it also demands great perseverance on the part of the caregiver, usually a busy mother.

A child with diarrhea may need to be spoonfed more than half a liter of solution each day, in sessions spaced as little as three minutes apart, for five to seven days. Small, frequent feedings are important because intake of a large volume at once can provoke vomiting. Most mothers are hard-pressed to do the job because, in the developing nations, they typically have other children to care for, food to cook, clothes to wash, water to haul (often from far away) and a garden or fields and animals to tend. In the industrialized nations, parents may be similarly burdened by family and job demands.

What a parent wants is what health workers want: to have the diarrhea stop as quickly as possible. Yet standard ORT does not reduce the duration or amount of diarrhea; it simply replaces lost fluid and electrolytes.

The failure to slow the diarrhea itself can also increase parents' reluctance to continue the solution. To many of them, putting more fluid in when it is clearly coming out in quantity makes no sense. In addition, the solution's inability to stop the diarrhea often spurs families to try other remedies, such as costly, and usually ineffective, antibiotics that can actually prolong diarrhea by killing helpful bacteria that live in the intestine.

Increasing the amount of glucose in the standard ORT might at first seem a reasonable way to speed fluid uptake and thus possibly to reduce the volume of diarrhea. Yet introducing extra glucose would in fact be dangerous. The reason lies with the phenomenon known as osmosis. When two watery solutions are separated by a semipermeable barrier, the solution with the higher concentration of dissolved molecules (solute) will attract water from the less concentrated solution.

Standard ORT contains approximately as many solute molecules as does normal blood. If more glucose were included, the concentration of the solution would become greater than that of normal blood—which is to say, its osmolarity, or tendency to gain water, would increase. Then water would actually flow from the blood into the intestine, exacerbating any dehydration.

One possible way to increase fluid absorption is to add one or more selected amino acids to the standard ORT formula. It is now known that in addition to the glucose and sodium co-transport carrier, villus cells have many co-transport carriers that simultaneously bring specific amino acids and sodium from the lumen into villus cells. In theory, by activating one or more classes of these carriers, the amino acids would accelerate the transport of sodium, and therefore water, and thus should more than compensate for any osmotic "penalty" introduced by the addition of more small molecules (the amino acids).

Amino acids with a high affinity for their carrier are the best candidates. So far alanine has shown the most promise in both animals and humans. When added to glucose, it substantially increases the absorption of sodium and water and reduces the output of stool.

Michael Field, now at Columbia University, suggested quite another alternative in an editorial he published in the *New England Journal of Medicine* in 1977, along with an epigraph from a children's verse by Maurice Sendak: "Sipping once / sipping twice / sipping chicken soup / with rice." He proposed substituting large chains of glucose (starches) or amino acids (proteins) for simple glucose in the standard ORT mixture.

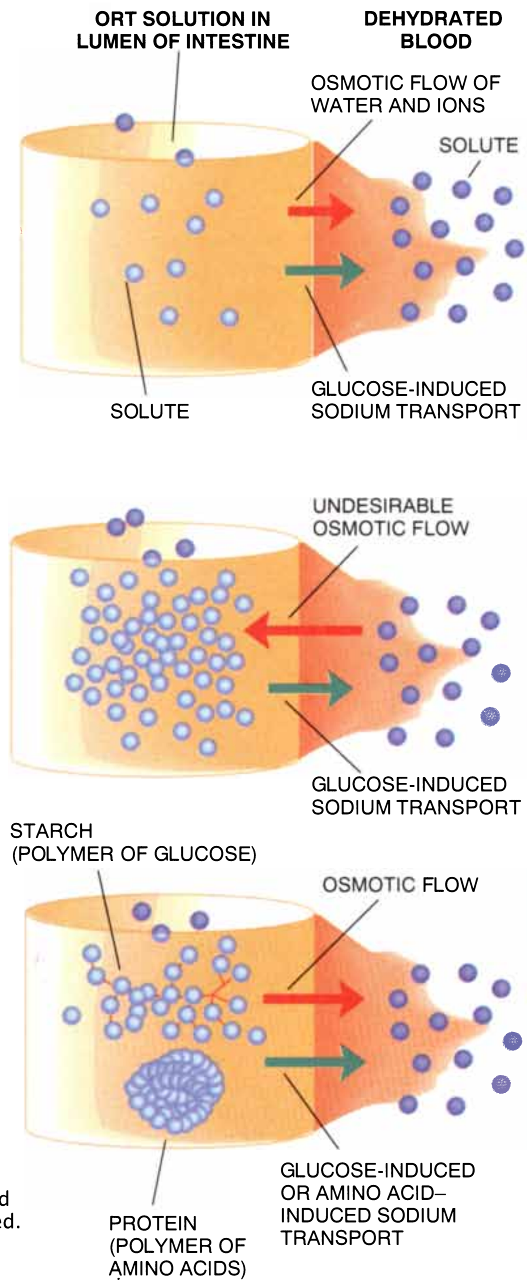
His recommendation stemmed from the understanding that any such polymer, potentially consisting of hundreds of smaller molecules, has the same osmotic effect as a single small molecule. True, starches and proteins are broken down into glucose or amino acids in the intestine, but the change occurs at the surface of the villus cells, whereupon the end products are instantly taken up by co-transport carriers. Hence, the resulting small molecules would not accumulate in the intestinal cavity to increase osmolarity. On the other hand, the large number of molecules entering the villus cells with sodium should markedly accelerate the movement of water from the lumen into the blood, thereby reducing the fluid lost in stool.

Most research on this second alternative for treating diarrhea has focused on cereals (grains) and legumes (beans and peas), both of which contain starch and protein. In 1982 F. C. Patra and Mahalanabis in Calcutta and, separately, A. Majid Molla and one of us (Greenough) in Dhaka reported that when glucose in the ORT solution was replaced by rice powder, the mixture rehydrated cholera patients. Significantly, the mixture also reduced stool volume by as much as 50 percent. Other studies achieved reductions of 15 to 49 percent in other kinds of diarrhea and also reduced vomiting. Since 1982, replacement of glucose with other grains or with legumes or roots has proved similarly beneficial. The benefit can certainly be attributed in part to the starches in these foods, but the proteins probably make a contribution as well; their role has not yet been assessed.

Food-based ORT offers advantages beyond its effects on diarrhea. It can be made at home with foods that people both like and can obtain easily at low

How Osmosis Affects Performance of Rehydration Solutions

If two solutions are separated by a water-permeable membrane, water will flow by osmosis from the solution containing fewer dissolved molecules of solute to the one containing more molecules, thereby balancing the concentrations.



Standard ORT
(osmolarity—the concentration of solute molecules in a solution—equals the normal osmolarity of blood)

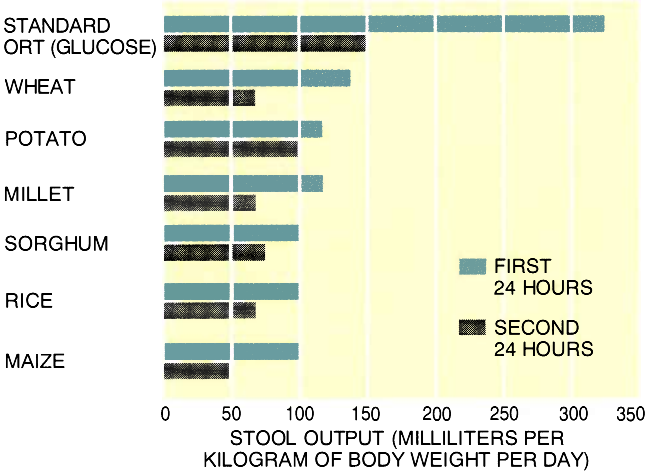
EFFECT: Co-transport of glucose and sodium induces a bloodward osmotic flow of water, which drags along additional ions. ORT exactly replaces water, sodium and other ions lost from the blood but does not reduce the extent or duration of diarrhea.

If extra glucose were added
(high osmolarity)

EFFECT: Solution is unacceptable because osmotic flow yields a net loss of water and ions from the blood—an osmotic penalty. Dehydration and risk of death increase.

Food-based ORT
(low osmolarity)

EFFECT: Each polymer has the same osmotic effect as a single glucose or amino acid molecule but markedly enhances nutrient-induced sodium transport when the polymer is broken apart at the villus cell surface. (Rapid uptake at the surface avoids an osmotic penalty.) Water and ions are returned to the blood quickly, and less of both are lost in the stool. The extent and duration of diarrhea are reduced.



JANUARY

In January
it's so nice
while slipping
on the sliding ice
to sip hot chicken soup
with rice.
Sipping once
sipping twice
sipping chicken soup
with rice.

6

COMPARISON of standard ORT and versions that replace its glucose with foods revealed that the alternative solutions reduce stool output in cholera patients (*left*). In other words, the food-based solutions accelerate absorption of fluid by the intestine. (The data are from the International Center for Diarrheal Disease Research, Bangladesh.) The last four lines of the verse at the right, from Maurice Sendak's *Chicken Soup with Rice*, could almost be a prescription for a food-based rehydration solution. Indeed, Michael Field, now at Columbia University, quoted the lines in the *New England Journal of Medicine* in 1977, when he suggested substituting starches and proteins for the glucose in standard ORT.

cost. It can also be cooked, which kills pathogens in the water (but costs fuel). For now, however, use of food-based ORT is limited, mainly because teaching millions of families how to prepare solutions at home is a major challenge for health education systems. Much research is therefore being devoted to creating a packageable formula based on precooked rice powder instead of glucose.

Of course, any packaged ORT mixture must be stable in many climates and inexpensive. No formulation is yet equal in these respects to the current WHO standard. Nevertheless, cereal-based products are being introduced commercially. For instance, a bottled and premixed solution called Ricelyte, which is based on a form of processed rice (rice-syrup solids), has recently entered the U.S. market.

Ultimately one would like to produce an ORT solution that not only rehydrates patients and reduces diarrhea with maximum speed (by exploiting all possible co-transport carriers) but also hastens the replacement of damaged villus cells. The cells can be hurt both by diarrhea-causing agents and fasting. Usually they are replaced every three to five days, but the process can be markedly extended in someone who has had diarrhea.

As malabsorption persists beyond the infection itself, so does the wastage of nutrients, which promotes continued weight loss and prolongs malnutrition. By repairing the intestinal lining and improving absorption rapidly, an optimal ORT solution would not only save lives, it would also improve the subsequent health of the survivors.

Much evidence in animals suggests that certain amino acids enhance the production of cells in the intestinal lining and that carbohydrates provide energy for those cells. These are important leads, but more research is needed to determine the ideal mixture of proteins and starch.

Other findings suggest that whether standard or food-based ORT is administered, patients should be encouraged to eat during their illness so that they receive crucial nutrients and enough calories (energy). They should be fed starting as soon as they can be coaxed to cooperate. Fortunately, ORT tends to restore the appetite quickly.

Many studies have evaluated the benefit of combining standard ORT with feeding and have found that children treated in that way have reduced stool output and a shorter duration of diarrhea. Moreover, more than a decade ago a series of studies from the Philippines, Iran, Turkey and the U.S. demonstrated that standard ORT supplemented by continued feeding of calorie-dense foods protects children with diarrhea from weight loss.

The studies indicate that the food protects the integrity of the intestine. They also imply that the popular advice to "rest the gut" by withholding food during diarrhea is almost certainly wrong. The foods chosen should contain starch and protein, which are digested easily, along with the fat needed to ensure intake of enough calories. (Fat contains more calories per gram than either protein or starch.) In infants, breast milk is always best.

Of course, the overarching goal is finding ways to prevent diarrhea in the first place. Better sanitation, cleaner water supplies and improved personal hygiene are crucial, especially hand washing with soap and water. At the medical level, vaccines against specific diarrhea-causing organisms are being developed. Yet before families will have confidence in these indirect measures, they need to see that their children, who are sick now, can be made well again. Oral rehydration therapy does this—sometimes spectacularly.

FURTHER READING

THE TREATMENT OF ACUTE DIARRHEA IN CHILDREN: AN HISTORICAL AND PHYSIOLOGICAL PERSPECTIVE. Norbert Hirschhorn in *American Journal of Clinical Nutrition*, Vol. 33, No. 3, pages 637–663; March 1980.

EFFECT OF DIARRHOEAL DISEASE CONTROL ON INFANT AND CHILDHOOD MORTALITY IN EGYPT. M. El-Rafie, W. A. Hassouna, N. Hirschhorn, S. Loza, P. Miller, A. Nagaty, S. Nasser, S. Riyad in *Lancet*, Vol. 335, No. 8685, pages 334–338; February 10, 1990.

ORAL REHYDRATION THERAPY. William B. Greenough III and Khin-Maung-U in *Diarrheal Diseases*. Edited by Michael Field. Elsevier Science Publishing, 1991.

MANAGEMENT OF ACUTE DIARRHEAL DISEASE. Edited by Fima Lifschitz. *Journal of Pediatrics*, Vol. 118, Supplement; April 1991.

CEREAL BASED ORAL REHYDRATION THERAPY FOR DIARRHOEA: REPORT OF AN INTERNATIONAL SYMPOSIUM, NOVEMBER 12–14, 1989. Edited by K. Elliot, K. Attawell, R. Wilson, N. Hirschhorn, W. B. Greenough III and Khin-Maung-U. Geneva, Aga Khan Foundation, and Columbia, Md., International Child Health Foundation, February 1990.

Photo: Reuters/Bettmann

Nuclear-generated electricity saves more oil each day than we used to import from Iraq and Kuwait.

Generating electricity with nuclear energy instead of imported oil helps reduce America's dangerous dependence on unstable energy sources.

With 112 operating plants in this country, nuclear electricity already cuts U.S. oil imports by 740,000 barrels every day. That's more than the oil we imported from Iraq and Kuwait before hostilities broke out last August.

But America still imports about half the oil it consumes—the equivalent of four huge supertankers of foreign oil every day.

We need more nuclear plants—to meet America's growing electricity demand and to bolster our independence from dangerously unstable energy sources.

For more information and two free booklets, write to the U.S. Council for Energy Awareness, P.O. Box 66080, Dept. ST10, Washington, D.C. 20035.

Nuclear energy means more energy independence.

Anyons

In quantum mechanics, indistinguishable particles are subject to special interactions. Recent work has revealed the existence of a rich new class of possibilities: anyons

by Frank Wilczek

One of the most surprising, profound and beautiful consequences of quantum mechanics is that it brings sharp, new meaning to the rules governing particles or objects that are exactly identical. Truly indistinguishable particles are subject to powerful special interactions, which do not exist for particles that are merely very similar. These special interactions are strongly attractive for one class of particles, bosons. They are strongly repulsive for another class, fermions.

Both bosons and fermions have been known since the infancy of quantum mechanics. Until recently, it was believed that fermions and bosons were the only possible kinds of identical particles. Indeed several textbooks of quantum mechanics present "proofs" of this factoid.

A critical examination of the foundations of the quantum mechanics of indistinguishable particles reveals instead that there are other consistent possibilities. In fact there is a continuous range of possibilities, containing fermions and bosons as two special cases. In the general case one calls the particles "anyons."

At first anyons were mostly thought of, by those few physicists who thought about them at all, as mathematical curiosities. Further investigation revealed the presence of anyons in theoretical models that quite closely resembled ones used to describe actual systems and materials. Detailed studies also demonstrated the depth and inner consistency of the concept of anyons. Still, it came as a stunning surprise to me—one of the few—when some quite tangible realizations of anyons were discovered in 1983. Anyons are the basic excitations in the spectacular states of matter known as fractional quantized Hall states [see "The Quantized Hall Effect," by Bertrand I. Halperin; SCIENTIFIC AMERICAN, April 1986]. Appreciation of this fact has led to a more profound understanding of several aspects of the quantized Hall effect itself and has stimulated considerable excitement and activity in the physics community.

Many of us now believe that this example, impressive as it is, is just the beginning. Anyons provide a new paradigm for the behavior of matter in quantum mechanics, and if there is any justice in the world, many other realizations await discovery. It is particularly intriguing that there is an extremely natural and powerful mechanism of superconductivity connected with anyons. A theory of high-temperature superconductivity has been constructed that exploits this feature of anyons. It may well apply to the copper oxide superconductors of recent fame.

In this article I shall review the logic that leads one to the concept of anyons, the situations in which they are at present known to occur and the mechanism of superconductivity they suggest.

FRANK WILCZEK is professor in the School of Natural Science at the Institute for Advanced Study in Princeton, N.J. He is the author of numerous scientific articles on theoretical physics and a popular book, *Longing for the Harmonies.* He is known, among other things, for the discovery of asymptotic freedom, the development of quantum chromodynamics, the invention of axions and the discovery and exploitation of new forms of quantum statistics (anyons). He has received a MacArthur Fellowship and the Sakurai Prize. He is a member of the National Academy of Sciences.

It is impossible to understand anyons without understanding bosons and fermions. To this end, I will now present a brief and somewhat unorthodox account of the essence of bosons and fermions. I will not begin to explain how the behavior of bosons plays a leading role in systems ranging from superfluidity to laser action. Nor will I review how the nature of fermions is central to understanding the periodic table and the stability of stars known as white dwarfs. Instead the focus here will be to examine critically the basic concepts in the simplest possible circumstances, to set the stage for anyons.

Let us start with some thought experiments, which are only slight idealizations of experiments that have actually been performed. The thought experiments are concerned with the behavior of two types of helium atoms, denoted ^{3}He and ^{4}He. These two types differ only in the nature of their nuclei. The ^{3}He nucleus contains two protons and one neutron, whereas the ^{4}He nucleus contains two protons and two neutrons. Because these two nuclei contain the same number of protons, they have the same electric charge. Therefore the electrical properties of these two nuclei are nearly identical. Since electrical forces are by far the most important factor in determining the interactions of the nucleus with the surrounding electrons and thus the chemical properties of the atom, ^{3}He and ^{4}He atoms have almost exactly the same chemical properties.

One can check this in a very concrete way, as follows. Consider a particle that is neither ^{3}He nor ^{4}He; call it *X*. *X* might be an iron atom, a water molecule or for that matter a strand of DNA. To see whether *X* interacts the same way with ^{3}He as it does with ^{4}He, one can collide it with each in turn and compare the responses. Such investigations, known as scattering experiments, are commonly used as a way to learn about the properties of particles.

So first imagine shooting ^{3}He atoms and *X* particles toward one another. As long as these particles are not moving too fast, they will survive the collision intact, deflected at some angle. After observing many collisions, one can calculate the probabilities to find particles emerging from a collision at different angles (that is, undergoing different amounts of deflection). This distribution of probabilities encodes a great deal of information about the nature of the interaction between the atoms.

Having performed (in our imagination) the scattering experiment with ^{3}He, let us try it again with ^{4}He. One finds that no matter what *X* is, there is very little difference in the outcome between its collisions with ^{3}He and its collisions with ^{4}He. The probability of deflection through a given angle is the same in both cases. These results are just what one might expect, since the most important forces are the electrical forces, and the electrical properties of ^{3}He and ^{4}He atoms are nearly identical. Yet something different—something strange and wonderful—happens if *X* itself is a ^{3}He or a ^{4}He atom. There are of course three possible cases: one may collide ^{3}He with ^{3}He, ^{3}He with ^{4}He, or ^{4}He with ^{4}He. Based on the previous experiments, one might expect the same result in all three cases. What actually happens, though, is quite different.

The simplest and most striking results are for particles deflected through a 90-degree angle. To observe the results, consider a setup in which two particle guns and two detectors rest on a table [*see illustration on page 61*]. Each gun can produce either ^{3}He or ^{4}He; one gun shoots particles due east, the other due west toward a central point. The detectors are positioned due north and south of the collision region, to catch the particles that have been deflected by 90 degrees. What are the results?

One finds that the probability of scattering through a 90-degree angle is exactly *twice* as great for collisions between two ^{4}He particles as for collisions between ^{3}He and ^{4}He, whereas for collisions between two ^{3}He particles, probability of scattering through 90 degrees simply *vanishes.* One could hardly ask for a more dramatic demonstration that there is something special about the quantum physics of identical particles.

The ^{4}He atoms are typical bosons, whereas the ^{3}He atoms are typical fermions. To understand their behavior properly, we must go back to the basic principles of quantum mechanics.

The fundamental ingredients used to describe quantum mechanical processes are complex numbers known as amplitudes. (Those readers unfamiliar with complex numbers will find a brief and, I hope, painless account of them in the box on the next page.) Like all complex numbers, an amplitude has a magnitude and a phase. To find the probability of the process, one must square the magnitude of its amplitude.

The most fundamental principle of quantum mechanics—the principle of superposition—can now be stated: to compute the total *amplitude* for some overall process that may have occurred in different ways, you must add the *amplitudes* for the different ways. Ordinari-

WINDING STAIRCASE is an apt metaphor for several aspects of anyon physics. The central distinction between anyons and bosons (or between anyons and fermions) is that one must keep track not only of particles that interchange position but also of those that wind around one another. The staircase also calls to mind the arrangement of currents in a solenoid and the motion of charged particles in magnetic fields—physical configurations that play key roles in the theory of anyons. The staircase, designed by Baccio Pontelli, is part of the Sistine Chapel in Vatican City in Rome.

ly, the *probability* that some overall result is achieved is the sum of the *probabilities* that it is achieved in different ways. (For instance, the probability that the die comes out 5 or 6 is the sum of the probabilities for 5 or 6 separately.) In quantum mechanics, we instead add amplitudes. This is not the place for more words on the principle of superposition, although it is endlessly fascinating—for present purposes we will simply assume it and consider the consequences.

Let us reexamine our thought experiments, using the concept of amplitudes. First, consider collisions between westerly ^{4}He atoms and easterly ^{3}He atoms. To calculate the total amplitude for the case in which the ^{4}He particle travels from the gun to the north detector, one adds the amplitudes for all the different trajectories leading to the final result of a ^{4}He atom hitting the north detector. The probability is then the square of this total amplitude. The same series of calculations can be applied to the case involving ^{3}He. Because of the symmetry of the situation, and because the important forces are the same, the probabilities for the arrival of ^{4}He and ^{3}He are equal. Thus the probability that *something* arrives in the north detector is exactly twice the probability for the arrival of either ^{4}He or ^{3}He.

Now consider collisions where both the westerly and the easterly atoms are ^{4}He. We now have amplitudes for two processes—the easterly atom winds up in the north detector, and the westerly particle winds up in the north detector—that lead to the *same* final result, namely, "^{4}He hitting the north detector." Thus the principle of superposition, which instructs us to add amplitudes for all ways of achieving a particular result, comes into play. The fact that at this point we must add amplitudes, not probabilities, is the crucial difference between processes involving indistinguishable particles and those involving distinguishable particles.

From the symmetry of the situation, the amplitudes for northern arrival from easterly and westerly departures are the same. Following the principle of superposition, we add these amplitudes to get the total amplitude of the collision process, which is therefore twice the amplitude of either set of trajectories. The probability that a ^{4}He particle will hit the detector is, then, the square of the magnitude of the total amplitude, equaling four times the probability for either set of trajectories. This result is twice what we found for the total rate of northerly arrival in the case of collisions between ^{3}He and ^{4}He atoms, and it agrees with experimental results.

What about the most striking result of all, namely, that the probability vanishes at 90 degrees for ^{3}He on ^{3}He? How can the argument of the previous paragraph, which leads to enhanced (rather than zero) probability for indistinguishable particles, be circumvented? The rule for combining the easterly and westerly amplitudes must be modified. The required modification is simple but may seem weird and ad hoc on first hearing. The modified rule is that for fermions, of which ^{3}He is an example, one must *subtract* the amplitudes for trajectories when the particles have been switched. Since the amplitudes for northern arrivals from easterly and westerly departure were equal, this rule will yield zero total amplitude and therefore zero probability for a northerly arrival.

In a fundamental sense, that is all there is to the traditional quantum statistics of bosons and fermions. Everything else follows from these basic rules. Thus to summarize: There are two possible classes of identical particles, bosons and fermions. For bosons (like ^{4}He), you simply add the amplitudes for all ways of getting to the final result; for fermions (like ^{3}He), you subtract if particles are switched.

The rule that we must subtract fermion amplitudes is startling and begs for a deeper explanation. Let us strengthen our understanding by seeking the "rules of the rules." Are there other possibilities for the rules, besides those that apply to bosons and fermions? What general requirements must a prescription for combining identical particle amplitudes satisfy?

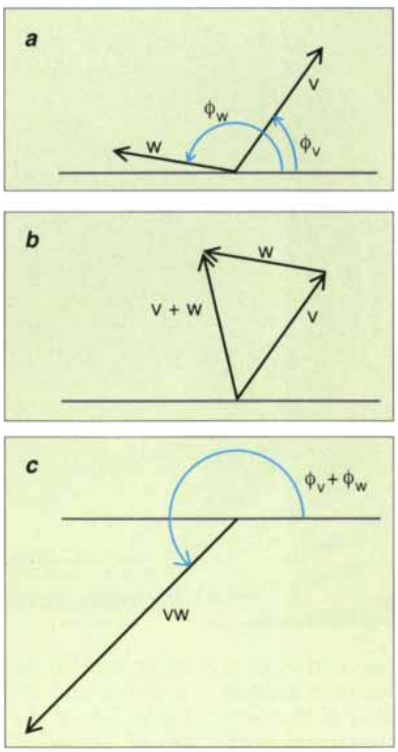
The most fundamental condition comes from a principle of quantum mechanics closely related to the superposition principle. In ordinary probability, one has the rule that the probability for a succession of two events is the probability for one multiplied by the probability for the other (given the first). For example, the probability that I will hit a fastball for a home run is the probability that a fastball will be thrown times the probability that if a fastball is thrown I will hit it for a home run. We mentioned in our discussion of the principle of superposition that the ordinary rules for adding probabilities are replaced, in quantum mechanics, by rules for adding amplitudes. So it should not seem altogether surprising that the rule for multiplying probabilities of successive events is replaced in quantum mechanics by the rule that to cal-

Complex Numbers

Just as real numbers can be represented as displacements along a line, complex numbers can be represented as displacements in a plane. The arrows labeled **v** and **w** in illustration ***a*** represent two complex numbers.

Each complex number has magnitude, the length of the arrow, and each also has phase, the direction of the arrow. Using this representation, one can readily visualize the addition and multiplication of complex numbers. To add the two complex numbers **v** and **w,** one displaces the arrows as depicted in illustration ***b***.

Multiplication is somewhat more complicated, as shown in illustration ***c***. The magnitude, or length, of the product of **v** and **w** is simply the length of **v** times the length of **w**. The direction, or phase, of the product of **v** and **w** is defined by the angle that is the sum of the direction angle for **v,** ϕ_V, and the direction angle for **w,** ϕ_W.



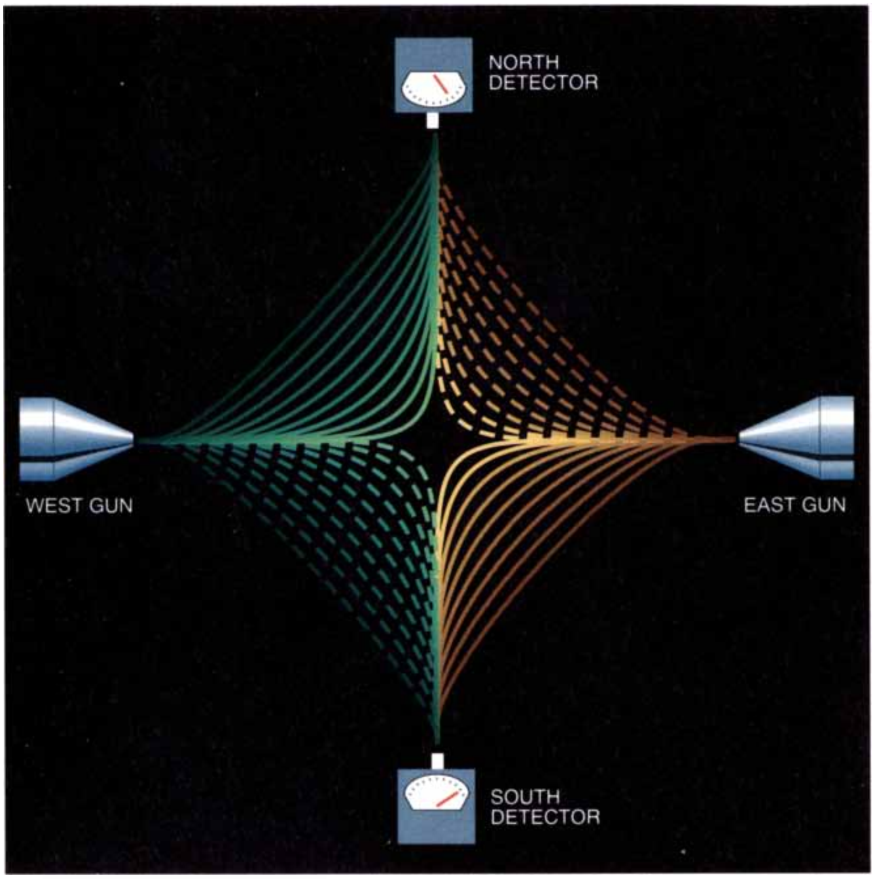
culate the total *amplitude* for a succession of events, one must multiply the partial amplitudes for every step in the succession.

This rule is extremely powerful. It allows one to evaluate the amplitude of a process with a complicated history in terms of much simpler ones. First, one breaks down the process into many events occurring over short intervals of time. Then one determines the amplitude for each shorter event. Finally, one multiplies the amplitudes for all the shorter events to get the amplitude for the complete history. Part of the power of this insight comes from the fact that for very short times the behavior of particles is essentially governed by classical mechanics. Thus by breaking down paths and reassembling them, it is possible to determine the quantum mechanical rule for amplitudes almost uniquely from the classical forces and interactions.

Some processes, however, cannot be broken down into a succession of small changes. In such cases one cannot use the rule for multiplying amplitudes, and one obtains no guidance from classical mechanics. An example of such a process occurs in the helium scattering experiment for indistinguishable atoms. In that experiment the trajectories leading to a northerly arrival fall into two classes, namely, easterly and westerly departures. Within each of those classes, the trajectories are all related by small changes. Between the easterly and westerly classes, however, the trajectories are related by a large change—the identity of the arriving particle is completely different. Therefore the relative amplitude for these trajectories cannot be determined from the amplitude for very small trajectories. In particular, it is not determined by the behavior of the particles in the classical limit.

Without the classical limit to guide us, it might seem that we are faced with an embarrassment of riches—any amplitude at all might seem to be possible. Fortunately, the possibilities are strongly constrained because the amplitudes for topologically distinct trajectories (that is, trajectories that cannot be continuously deformed into one another) must respect the general principles of quantum mechanics. In particular, the amplitudes must be consistent with the principle that the amplitudes for successive events are the products of the amplitudes for the successive subevents.

To see the power of this requirement, consider how it applies to a succession of two switching processes. If two particles exchange position and then exchange position again, clearly the net effect is no exchange at all. Thus for consistency, when we multiply by some switching factor and then multiply by the factor again, the overall result must be the same as multiplying by 1. In other words, the square of the switching factor must equal 1. We are left with just two mathematical possibilities: the factor itself must be 1 or –1. Consequently, either we simply add amplitudes or we subtract them when particles are switched. These are just the cases of bosons and fermions. Both cases are observed in nature, as we have seen.



SCATTERING EXPERIMENT can reveal the quantum mechanical behavior of particles. Two guns shoot particles toward one another. The particles that are deflected by 90 degrees are recorded in the two detectors. The solid lines show one set of possible trajectories for the deflected particles. The dotted lines show another set. The particles are more likely to take the trajectories close to the center (*brighter colors*) rather than the ones farther away (*darker colors*).

But is that the end of the story? If it were true that the only two disconnected classes of trajectories were those corresponding to particles exchanging their identity, then it would be. But it is far from obvious whether further discontinuous distinctions among the trajectories exist.

For example, consider trajectories that wind around one another a different number of times [*see illustration on next page*]. Because the number of windings is always a whole number, to change at all it must change by a whole number, that is, by a discrete jump. Thus continuous changes in the trajectory, which do not introduce any jumps, cannot change the number of windings at all. Does the possibility of windings change the classification of paths, leading to new possibilities for quantum statistics?

In three spatial dimensions the answer is no. The best view to see this is from the vantage point of one of the particles looking toward the other. In the course of the motion the relative position of the other particle traces out a curve in space. One can imagine shrinking this curve down to a point in such a way that the curve never passes our vantage point. In a similar manner, any apparent winding, no matter how complicated, can be undone. The original, complicated trajectory is continuously related to a trivial one.

In two spatial dimensions the situation is quite different. The trajectory of a particle can lasso our vantage point. If it does, then it cannot be shrunk without passing through our vantage point. Hence the trajectory can be related to a trivial one only by a discontinuous change, namely, a discrete jump over the particle that serves as our vantage point. Indeed the trajectory may wind around our vantage point repeatedly, either clockwise or counterclockwise. Trajectories with different net numbers of windings cannot be continuously related to one another.

Armed with this insight, let us reexamine the argument that led to the conclusion that bosons and fermions are the only consistent forms of rules for the interaction of identical particles. That argument was based on the fact that if one followed an exchange of identity with a second such exchange, the result could be continuously deformed into a trivial untangled trajectory. In two dimensions, however, even some trajectories that do not involve exchanges of identity still cannot be untangled completely. For instance, one cannot untangle a two-dimensional trajectory in which the particles wind around one another. If there are windings, we must be more careful about how we formulate the "rules of the rules."

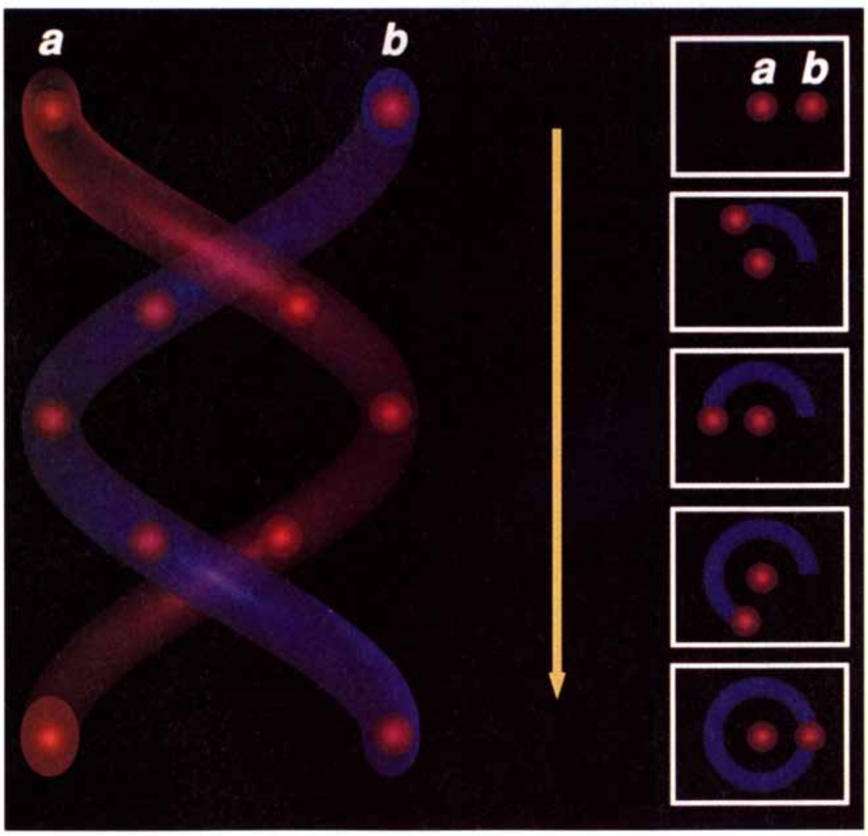
What is the relationship between exchange and winding? In a certain sense, an exchange can be considered as half a winding. To see this, let us consider one concrete way of making the exchange. Let two particles circle clockwise around their midpoint until they assume each other's original positions. This results in an exchange of identity. Do it again. Now the particles are back at their starting points, having wound around each other once. Thus a succession of two exchanges results in a trajectory with exactly one winding, which cannot be deformed to a trivial untangled trajectory—at least not in two dimensions. And so the requirement that the square of the factor for an exchange must be simply 1 no longer holds. In fact this factor can be any complex number. For convenience, I will say that β is the complex number multiplying the amplitude for an exchange in the clockwise direction. Then the amplitude for a complete winding in the clockwise direction will be β^2. The amplitude for two windings is β^4, and so forth. There is no requirement that β be either 1 (as for bosons) or else –1 (as for fermions). Bosons and fermions are merely two extreme cases; there is a continuous range of possibilities between them. The term "anyons" expresses this freedom to choose any complex number.

How does one handle a quantum mechanical process involving more than two anyons? The rules for the amplitudes of processes involving many identical particles can be derived from those involving just two particles. For bosons, we must simply add the amplitudes. For fermions, we count the total number of exchanges of identity, then multiply by –1 for those trajectories with an odd number of exchanges and finally add the amplitudes. For the general case of anyons, we must add up the total number of windings over all pairs, then multiply by β to the appropriate power and, once again, add the amplitudes.

We can summarize the results of our reexamination of the foundations of quantum mechanics for identical particles as follows. In three dimensions, we have attained a deep understanding of why bosons and fermions are the only possibilities. In two dimensions, they are not the only possibilities. There is a continuum of new possibilities—anyons—in between.

After such an arduous intellectual journey, the destination where we have arrived may appear disappointing. After all, the real world is three-dimensional, and the existence of exotic possibilities for identical particles in Flatland might seem rather academic. And yet anyons are quite relevant to the real world and indeed to the description of some of the most interesting states of matter. How can this be? The main point is that under certain circumstances matter does behave as if it were two-dimensional.

One such circumstance is when we are describing surface layers only one or a few atoms thick. Another arises in materials like graphite or the copper oxide high-temperature superconductors that are composed of planes of atoms stacked on top of one another. Yet another, relevant to the quantized Hall effect, arises when we consider electrons confined in a plane by electric fields. In all these cases, the states of motion in the transverse direction are quantized—meaning it takes a finite amount of energy to excite them.



INDISTINGUISHABLE PARTICLES exchange position once and then once again as they travel through time and space. The overall result is a process with no exchange. Looking from the particle *a* toward the particle *b*, the relative positions sweep out a closed curve in space (*panels 1 through 5*).

At sufficiently low temperatures the energy to excite them will not be available, and a two-dimensional description becomes absolutely appropriate.

Of course fundamental particles such as electrons or photons ultimately are capable of escaping from any Flatland, and at the most basic level (at high energies and in vacuum) they are certainly required to be either bosons or fermions. Yet the most direct and appropriate description of the low-energy behavior of a material is generally not in terms of these elementary particles. For example, an electron in a material exerts forces on the other constituents of the material and creates a little pocket of disturbance in its neighborhood, like a movie star moving through an admiring crowd. The basic excitations in a material may not behave at all like electrons or other elementary particles in vacuum. For this reason, they are known as quasiparticles. One can hope that quasiparticles in effectively two-dimensional materials are sometimes anyons.

For anyons to arise, these materials must also satisfy another subtle but important requirement. The definition of anyons includes the factor β, which is associated with a clockwise winding. For consistency, we must associate a factor β^{-1} with a counterclockwise winding. But the clockwise and counterclockwise windings are related by mirror reflections. The counterclockwise windings are equivalent to clockwise windings viewed in a mirror, a relationship known technically as a parity transformation. Similarly, if counterclockwise windings are reversed in time, they come to look like clockwise windings. For most practical purposes, the form of the fundamental laws of physics is unchanged by parity or time reversal. The world you see in a reflected mirror or when you run a movie backward in time obeys the same fundamental laws of physics (to a good approximation) as the real world.

If we assume that symmetry under parity and time-reversal transformations are maintained in a material, then we must conclude that quasiparticles winding clockwise around one another should have the same amplitude as particles winding counterclockwise and therefore that β equals β^{-1}. This equation forces β to be either 1 or –1, landing us back at bosons and fermions.

Fortunately, this requirement proves less restrictive than it appears. Although the basic laws of physics are for the most part unchanged by parity and time-reversal transformations, concrete situations and materials need not be. For example, the presence of a magnetic field, or rotation, destroys both these symmetries.

Thus the most obvious obstacles to the existence of anyons in reality are not insurmountable. Thinking about these obstacles does, however, focus our search: anyons must be quasiparticles in effectively two-dimensional materials that violate the symmetries of parity and time reversal.

But how would we know an anyon if we saw one? Particles—or quasiparticles—do not come conveniently labeled "boson," "fermion" or "something new: anyon here." To have a chance of recognizing anyons, we must know how anyons behave.

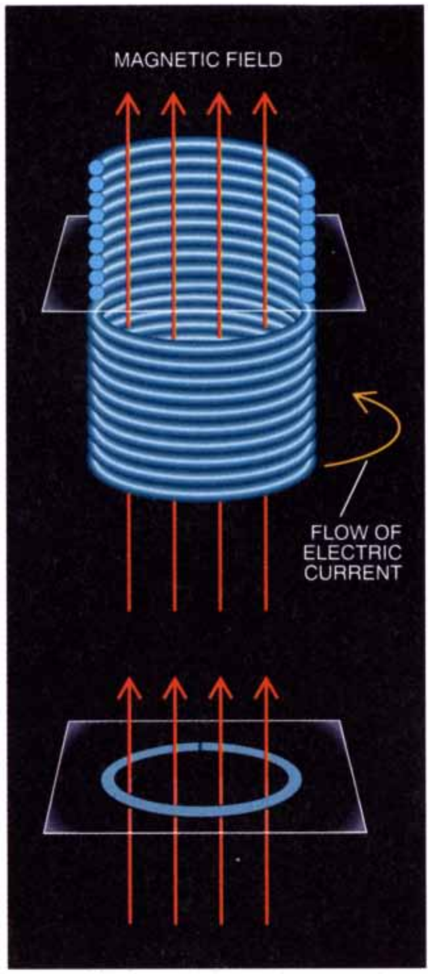
In attempting to understand a genuinely new concept like anyons, it is good intellectual strategy to relate it to more familiar things. Fortunately, there is a very powerful method for relating general anyons to the more familiar cases of bosons and fermions. A special magic available in two dimensions allows us to metamorphize anyons into bosons or fermions—or vice versa—by imaginative manipulations of fantasy magnetic fields. This alchemy is called statistical transmutation, and it is very important in the theory and practice of anyon physics.

Here is how statistical transmutation works. You may be familiar with the behavior of a long solenoid, which is no more and no less than a long winding of wire [*see illustration at right*]. When a current flows through a solenoid, there is a constant magnetic field (proportional to the current) inside the solenoid but no magnetic field outside. To generate the two-dimensional version of a solenoid, simply take a planar section of this setup, perpendicular to the axis of the solenoid. In this way we find that in two dimensions a magnetic field can be fully contained in a small region. This configuration is called a flux point and is easily generated under a variety of conditions. (Flux is a quantitative measure of the overall power of a magnetic field distribution. It is equal to the product of the strength of the magnetic field and the area it occupies.)

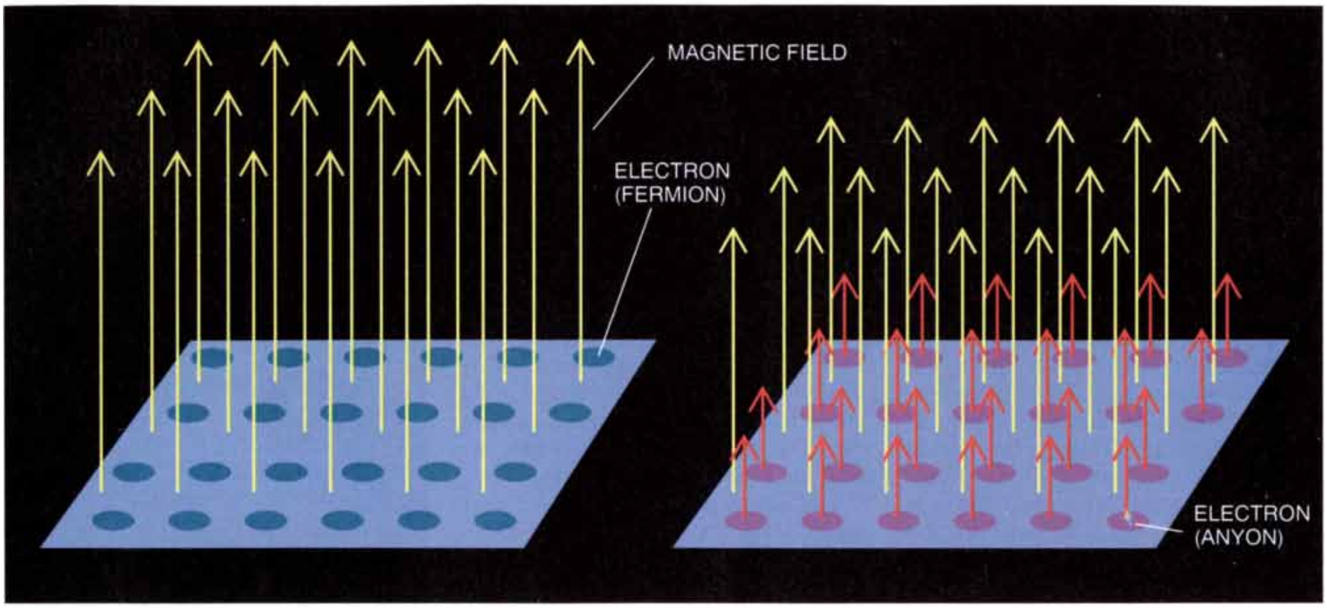
From the point of view of classical physics, a flux point is a rather negligible thing. A charged particle will feel no magnetic field and experience no force unless it happens to pass bang on through the middle—which is terribly unlikely. In quantum mechanics the situation is entirely different. The amplitude for a charged particle winding around a flux point is multiplied by a factor β that is precisely similar to the factor we discussed in connection with quantum statistics. For a process in which the charge winds n times around the flux point, the amplitude is multiplied by a factor of β to the nth power. The phase of β is proportional both to the charge and to the flux. Apart from this, the flux points have no effect on other particles.

Thus attaching a fixed amount of charge and flux to every member of a set of identical particles produces the same effect as altering the quantum statistics of those particles. This idea is the essence of statistical transmutation. It allows us to represent one type of anyon as another type of anyon with fictitious flux and charge attached.

If flux and charge are both present, one may expect statistical transmutation to occur. A common case arises in what are known as type II superconductors. When a thin layer of this kind of superconductor is exposed to a magnetic field, it will allow the field to pen-



SOLENOID produces a constant magnetic field inside the coils but none outside. A cross section through the solenoid yields a two-dimensional region that contains a certain amount of the magnetic field or, more accurately, flux.



STATISTICAL TRANSMUTATION can occur when electrons in two dimensions encounter a magnetic field. Each electron can "absorb" a certain amount of field (*red*), and in the process the electrons can transform from fermions to anyons.

etrate in localized regions. For conceptual purposes, we may idealize each of these regions as a flux point. Now suppose that an electron becomes attached to such a flux point. Then we have a charge attached to flux, and we may expect statistical transmutation to occur [*see illustration above*].

For a standard superconductor, it turns out that the amount of flux through a flux point is exactly enough to transmute the statistics of an electron from fermion to boson. Thus, whereas the free electron is a fermion, the electron attached to a flux point acts like a boson.

For various reasons, the statistical transmutation of flux points in ordinary type II superconductors may be difficult to observe. But the physics involved is closely related to one way (the best way, in my opinion) of understanding one of the most intriguing phenomena in modern physics, the fractional quantized Hall effect (FQHE). Unfortunately, a full discussion of the FQHE is beyond the scope of this article. Ruthlessly suppressing many fascinating aspects and technical ramifications of this effect, I will now briefly relate its essence.

Readers may remember that certain atoms that contain certain special numbers of electrons are particularly energetically favorable and stable (leading to the existence of inert elements, or noble gases). Likewise at certain special densities the two-dimensional electron gas in a magnetic field is especially stable. More precisely—and this is very important—it is the ratio of density to applied magnetic field that is the crucial factor. This ratio is called the filling factor. At the preferred filling factors the electrons, which usually behave like a gas, instead behave more like a liquid, resisting changes in density.

The existence of one of the preferred filling factors can be understood in a fairly straightforward way, similar to the way we understand closed shells in the atoms of an inert gas. This filling factor is usually defined to be 1, and in this case we refer to the integer quantized Hall effect. In 1983 it was discovered, to universal amazement, that other discrete fractional values of the filling factor—for example, exactly one third the obviously preferred value—are also especially preferred. At these values the electron gas becomes a liquid. That is the FQHE.

There is a lovely way to understand this effect, closely related to the physics of statistical transmutation. In the experimental arrangement used to study the quantized Hall effect, electrons are subject to an external magnetic field. The electrons being studied are completely different from the ones that generate the magnetic field. One can fantasize, however, that the electrons carry some of the magnetic field, that is, they have imaginary flux points attached. As a result, the electrons transmute into anyons. At the same time, a change occurs in the filling fraction, that is, the ratio of the electron density to the external magnetic field. (The change is the result of the electrons' "absorbing" some of the external field.) Therefore the original electrons (which are fermions) at a given filling fraction are related to anyons at a different filling fraction.

As each electron gathers more and more flux, the electrons change into anyons of various types. If the process goes far enough, we eventually come back to fermions. At that point our fantasy impinges on reality, since our imaginative swapping of magnetic field for flux has once more produced an acceptable state for physical (that is, fermionic) electrons. In the process, however, the value of the background field has been changed and with it the filling factor. Remarkably, this procedure succeeds in generating states at just the fractional filling factors that are observed to be most favorable.

The theory of anyons enables us to connect two special electron states that otherwise appear to be of a very different character—and a well-understood state to another that was somewhat mysterious. It also helps us to understand in a simple, satisfying way why the favorable filling factors are what they are.

This use of anyons as a conceptual tool is pleasing, but the best is yet to come. Anyons also appear directly in the FQHE. According to the construction I just outlined, electrons in the FQHE are actually in a way "superfermions." If one electron is interchanged with another, the phase of the amplitude is not only changed by π, which is always associated with fermions, but also (in the simplest case of filling factor one third) altered by an extra 2π, which is associated with the gathered-up flux. The total change in phase is therefore 3π. This fact strongly suggests that it ought to be possible to define another kind of pointlike excitation that is, in some sense, equivalent to one third of an electron.

In fact this hypothesis turns out to be true in a very strong sense. The quasiparticles associated with the filling factor one third FQHE carry one third of the charge of the electron—and one third of the electron's statistics! Mathematically, when one of these quasiparticles winds around another, the amplitude is multiplied by the cube root of –1. In the FQHE states that appear at other filling factors, quasiparticles emerge carrying other types of exotic anyon statistics.

Possibly the most exciting development in anyon physics recently has been the realization (mainly the work of Robert Laughlin of Stanford University) that an essentially new and very powerful mechanism of superfluidity and superconductivity is associated with the existence of anyons.

The essential attributes of a superconductor are that electric currents can flow in it and that their flow, once started, has no easy way of dissipating. What this second requirement means is best elucidated by considering how it can fail. For instance, in an ordinary (nonsuperconducting) metal, a current comes about as many electrons move more or less independently in response to an electric field. If the electric field is taken away, the electrons will slow down as they collide with one another and with the nuclei of the metal. In this way the flow of electrons is dissipated.

So a key aspect of superconductivity is that it must not be possible for single particles to slow down gradually, giving up energy to other particles or to vibrations. Yet they must be able to move. The only way to reconcile these requirements is if there is exactly one isolated state having lower energy than any other, for each possible value of the total current. We may think of this state as a state of correlated motion of all the particles, such that it is energetically unfavorable for any single particle to get out of line. The energy penalty for getting out of line must be greater than the energy gain for slowing down a little.

The states of the quantized Hall effect have something of this character. At any of the preferred filling factors, one finds a unique isolated state of lowest energy, in which the motion of all the electrons is correlated. Unfortunately, however, the quantized Hall effect takes place in the presence of a background magnetic field, which means that charged particles cannot flow in a normal way. As a result, although in a certain weak sense they are superconducting, the quantized Hall states do not exhibit the most dramatic consequences of superconductivity. But by using the idea of statistical transmutation, one can relate these flawed superconductors to some true ones. For according to that idea, one can trade, conceptually, a magnetic field for particle statistics. Let us start with a favored fractional quantized Hall state and trade *all* of the real field for fictitious flux attached to the particles, thus changing them into anyons. The resulting state retains its dissipationless character, and yet (since there is no real field left) it can flow. It is a superconducting state. Such states are excellent candidates for describing the behavior of materials whose quasiparticles are anyons of the appropriate type.

More detailed investigations, both analytic and numerical, have shown that the states predicted to be superfluid according to this qualitative argument are in fact so. This mechanism of anyon superconductivity is very robust and works even if the other interactions of the quasiparticles are highly repulsive.

(By comparison, the classic mechanism of superconductivity through pairing is quite delicate. It requires a net attractive interaction of some kind. This is rather tricky to arrange for electrons, because in most circumstances the Coulomb repulsion between electrons is the dominant force. The way it works in ordinary superconductors is that one electron of a pair attracts the much slower-moving positively charged ions in its vicinity and is absent from the premises when the other electron of the pair comes around and gets attracted by the concentration of positive charge. But all this can work only if the crystal is not too noisy and the electrons in a pair are reliably well separated. These two requirements make it difficult for this mechanism to work at any but extremely low temperatures.)

Are the copper oxide high-temperature superconductors anyon superconductors? Several general signs certainly encourage one to think they may be. The main characteristic the copper oxide superconductors have in common is two-dimensional planes of copper and oxygen. These two-dimensional structures, which display pronounced magnetic ordering of a poorly understood nature, seem like a promising breeding ground for anyon quasiparticles. Above all, there is the towering fact that these materials *are* high-temperature superconductors. One feels that such a striking, apparently rare and qualitatively new phenomenon deserves a worthy explanation that relates to fundamental properties of the materials in question.

In science, however, the ultimate criterion is not the aesthetic pleasure we take in our theories but whether they help us to comprehend particular facets of reality. Ultimately, only experiments can tell us whether the anyon mechanism of high-temperature superconductivity is relevant to these particular materials.

What are the relevant experiments? If anyons do reside in copper oxide superconductors, then one would expect that the symmetries of parity and time reversal would be broken in these materials. As this article is written, the issue is unsettled. Some experiments seem to show evidence that these symmetries are broken, but others of comparable sensitivity do not.

After 60 years of fruitful development, the theory of bosons and fermions may be getting a bit stale. Anyons provide a new paradigm for the behavior of matter in two dimensions. The states of the fractional quantized Hall effect provide one rich realization of this paradigm. Experience has taught us that nature makes abundant use of every one of the few simple and consistent possibilities quantum mechanics offers for the description of matter. I am therefore confident that many further realizations of anyons await discovery.

FURTHER READING

THE RELATIONSHIP BETWEEN HIGH-TEMPERATURE SUPERCONDUCTIVITY AND THE FRACTIONAL QUANTUM HALL EFFECT. R. B. Laughlin in *Science*, Vol. 242, pages 525–533; October 28, 1988.

FRACTIONAL STATISTICS AND ANYON SUPERCONDUCTIVITY. Frank Wilczek. World Scientific, 1990.

FRACTIONAL STATISTICS: QUANTUM POSSIBILITIES IN TWO DIMENSIONS. G. S. Canright and S. M. Girvin in *Science*, Vol. 247, pages 1197–1205; March 9, 1990.

THE HUNT FOR ANYONS IN OXIDE SUPERCONDUCTORS IS INCONCLUSIVE. Barbara Goss Levi in *Physics Today*, Vol. 44, No. 2, pages 17–20; February 1991.

René Jules Dubos

This quintessential environmentalist's quest for mechanisms of disease evolved into a philosopher's search for health

by Carol L. Moberg and Zanvil A. Cohn

During the course of his life, René Dubos progressed from ground-breaking studies of tuberculosis and pneumonia to investigations of the overall pattern of disease and ultimately the health of the earth. One of the most influential ecological thinkers of the 20th century, he combined the stern requirements and careful observations of a biologist with the daring views and infinite perspectives of a humanist. He believed that a living organism—microbe, person, society or planet—could be understood only in the context of the relationships it forms with everything else.

This view was shaped both by Dubos's distinguished career as a bacteriologist and by his personal experiences of health and disease. His ecological approach to discovering new drugs led to gramicidin, the first clinically useful antibiotic. That same approach, applied on a broader scale, led Dubos to a conviction that chemical tools for destroying disease might eventually prove a dead end in maintaining health.

Dubos's vision of organisms that influence their environments—and in turn are influenced by them—caused him to describe himself as a "despairing optimist" in the face of human depredations of the global ecosystem. Although people's actions could clearly change the face of the world for the worse, Dubos believed that they could also restore it and even create surroundings that would ultimately improve the human condition.

CAROL L. MOBERG and ZANVIL A. COHN work in the Laboratory of Cellular Physiology and Immunology at the Rockefeller University. Moberg, a research associate, began working with René Dubos on environmental issues in 1965. She received a Ph.D. in comparative literature from Columbia University in 1978. Cohn is co-head of the laboratory and a senior physician and professor at Rockefeller. He received his M.D. from Harvard Medical School in 1953 and joined Dubos's laboratory at Rockefeller in 1958. The two recently edited a book about the early work of Dubos and others, *Launching the Antibiotic Era*.

When Dubos was eight years old, severe myopia and an incapacitating attack of rheumatic fever forced him to make the first of many adjustments in his life. He abandoned his dreams of becoming a bicycle racer or tennis champion and channeled his now limited athletic abilities into an intellectual restlessness that dominated the rest of his years. Meditative walks in the French countryside nurtured his solitary spirit. He read avidly and found childhood heroes in translations of Buffalo Bill Westerns and Nick Carter detective stories. Young Dubos found a different kind of health, despite physical handicaps.

At age 14, Dubos was introduced to the molding force of the environment: an essay by Hippolyte Taine described how the countryside of the Île-de-France, the province surrounding Paris, affected the fables of La Fontaine. Dubos became sensitive to the way these tales depended on the region's landscapes, where nature took on human dimensions and also served as a measure of human life. At 80 years, Dubos still recited from memory his favorite La Fontaine fable, "The old man and the three youths." And he mused on his love of planting trees: "I [tell] myself, as did La Fontaine's old man, that someone will enjoy the shade the trees will cast after I am gone."

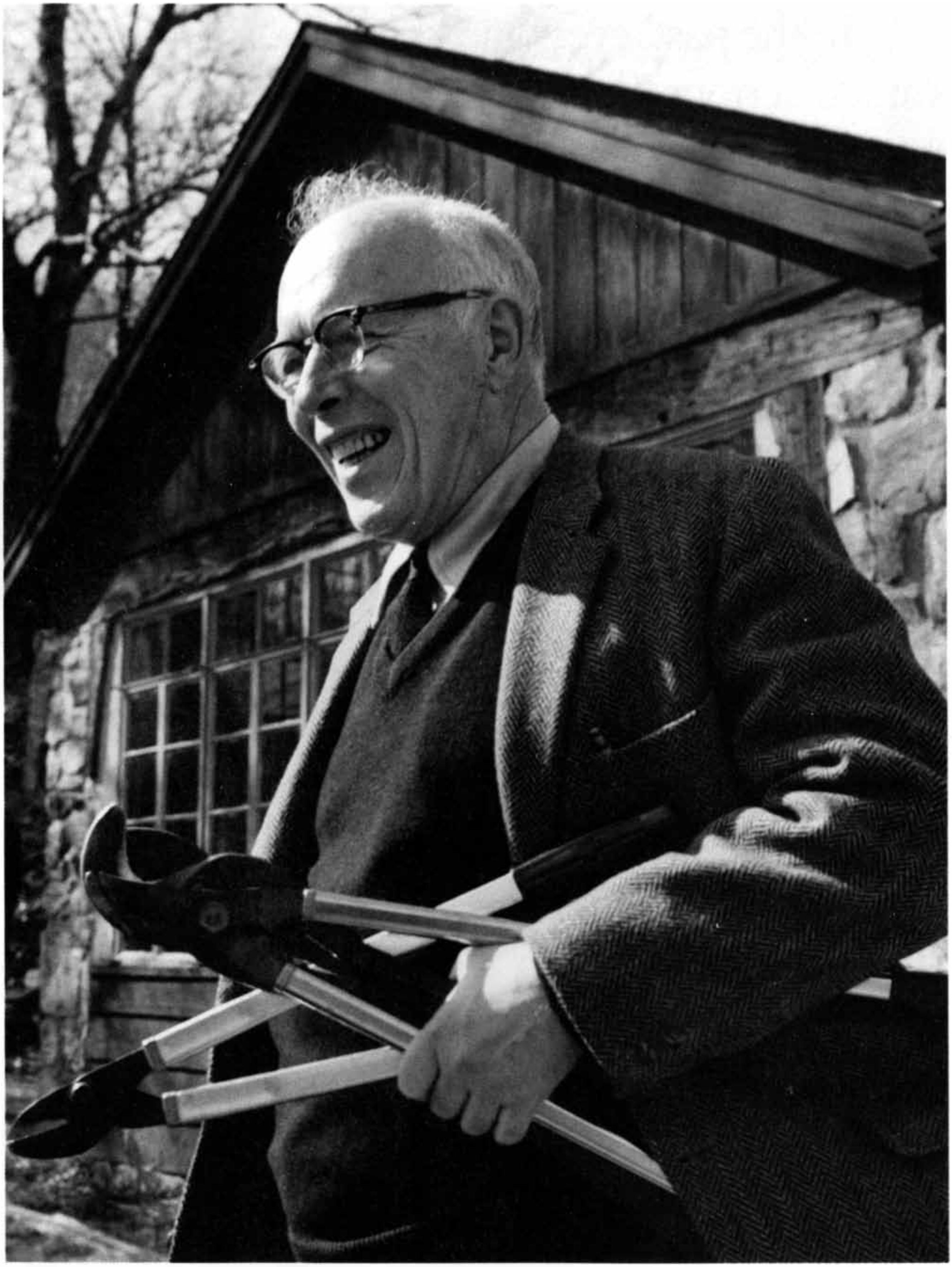
A series of chance events—and his responses to them—shaped the beginnings of Dubos's career. On recovering from a second bout of rheumatic fever, he found only one school, the Nouvel Institut National Agronomique in Paris, still open for enrollment that year. He excelled in all courses except microbiology. It was "intensely boring," he later recalled, because it dealt solely with taxonomy. He also disliked chemistry and told his mother he would never again enter a laboratory. On graduation Dubos won a scholarship to study agriculture and technology in Indochina but was then disqualified because of his rheumatic heart.

Undecided about a career, Dubos found a job in Rome with a branch of the League of Nations, writing abstracts for an agricultural journal. While sitting in the Palatine Gardens on a warm day in May, instead of reading about fertilizers, he turned to an article by the Russian soil microbiologist Sergei Winogradsky. Winogradsky wrote that microorganisms should be studied not in artificial laboratory cultures but in their natural environments in competition with other bacteria. Dubos later said his scholarly life began with this idea. He embraced this ecological approach to science and decided to study microbiology after all.

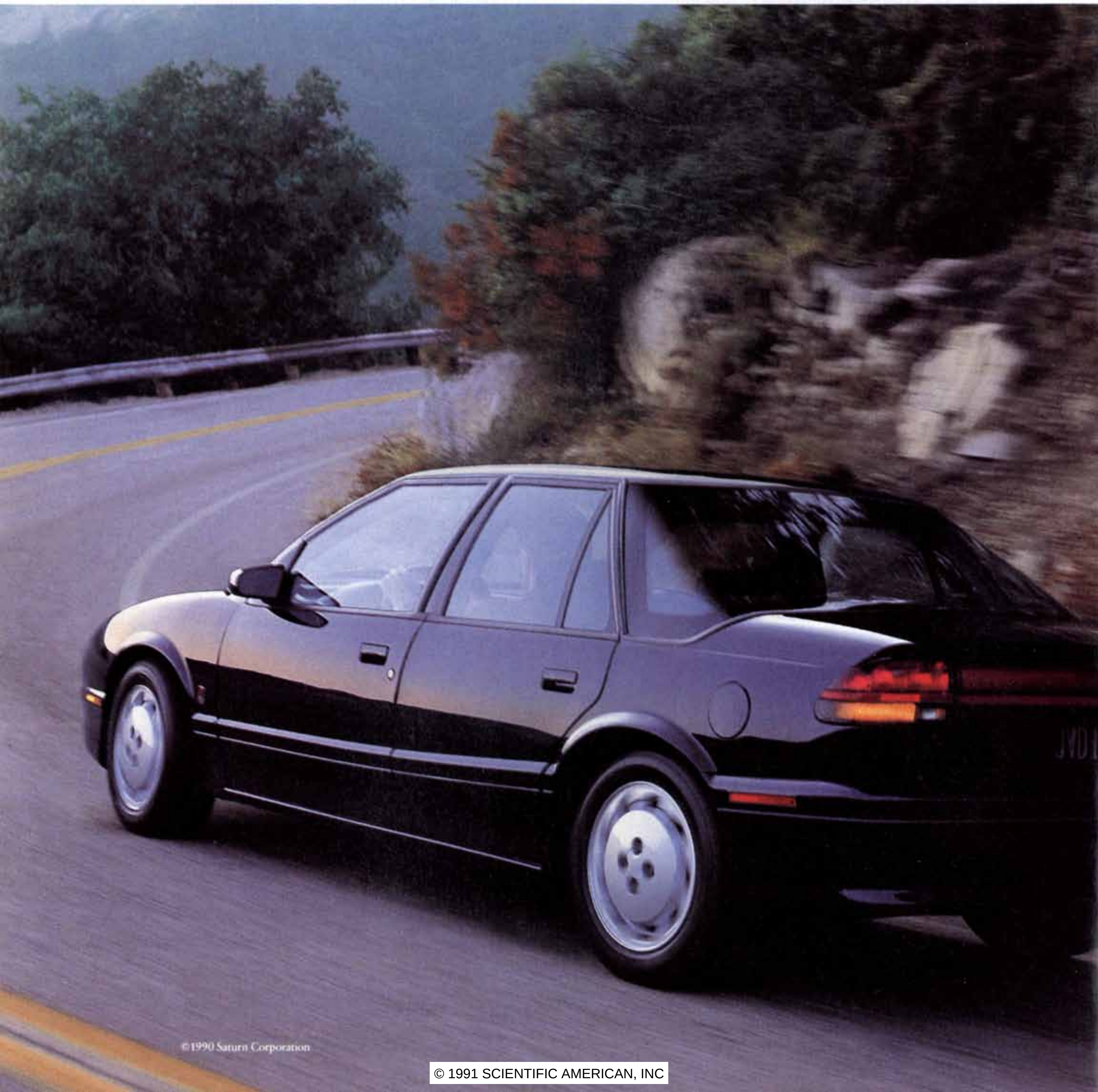
Dubos wanted to visit the U.S. before resuming his studies; instead circumstances were to dictate that he stay for more than 50 years. On the steamship *Rochambeau* he encountered Selman Waksman, a soil bacteriologist whom he had recently guided around Rome during an international congress. When Waksman learned Dubos had ambitions to study bacteriology but no definite plans, he offered the young Frenchman a fellowship to study at Rutgers University. Dubos arrived in New York and accompanied Waksman that same evening to the Rutgers campus in New Jersey.

Three years later Dubos earned his Ph.D. in soil microbiology. In the spirit of Winogradsky, his doctoral work demonstrated that the environmental characteristics of the soil, specifically pH, moisture and aeration, determine which microbes are activated to decompose cellulose.

As a student of the soil, Dubos became attracted to what he called the "philosophic necessity" of three concepts from which his scientific life took shape. The first was that countless mi-



RENÉ JULES DUBOS (1901–1982) was born in France and emigrated to the U.S. at age 23. He spent almost his entire career at the Rockefeller Institute for Medical Research (now the Rockefeller University), where he developed the first clinically useful antibiotic. He spent leisure time planting and pruning hemlock trees around his house in Garrison, N.Y.



crobes perform limited, well-defined tasks to recycle organic matter so that it does not accumulate in nature. The second was that microbes do not digest foods indiscriminately; rather they are fussy eaters that can be isolated and trained to perform highly specialized tasks. Finally, he examined how microbes function in complex mixtures: not only did they affect one another's lives and growth, but the environment also influenced their activities. He later enlarged on the idea of ecological interaction as he developed theories linking microbes with humans and then humans with the earth.

BACTERIAL POLYSACCHARIDE was mentor Oswald Avery's challenge to Dubos. The complex sugar coat protected a particularly virulent strain of pneumonia from the body's defenses; Dubos isolated an enzyme that could dissolve the protective coat.

Another chance encounter brought Dubos to the Rockefeller Institute for Medical Research in 1927. The National Research Council rejected Dubos's application for a fellowship because he was not a citizen but suggested that he consult with fellow Frenchman Alexis Carrel at Rockefeller. Carrel, a cardiologist, took Dubos to lunch in the institute dining room and sat him next to Oswald Avery, a physician whose research on the pneumococcus would lead him to discover DNA as the material of heredity in 1944.

When Dubos spoke with Avery, however, Avery's hospital laboratory was trying to produce a serum to cure the deadly disease lobar pneumonia. Avery had been thwarted by efforts to decompose safely the polysaccharide capsule surrounding the virulent type III pneumococcus. He knew that whoever could find a way to destroy the capsule without side effects would discover a cure for pneumonia.

Dubos was well prepared for this chance meeting. The two spent the afternoon discussing Dubos's experience with recovering microbes that digest cellulose, also a polysaccharide, and Avery's preoccupations with the pneumococcus capsule. Dubos asserted, rather brashly for a new Ph.D., "I think I can find a germ that can decompose that [capsule], and from that germ I can extract an enzyme." Avery, no doubt excited by such a promise, arranged a fellowship. Dubos was always grateful to Avery and Rockefeller for having "taken a person like me, knowing nothing at all about medicine, and coming from an agricultural experiment station, and given him a chance to work in a hospital."

Plunged into a hospital setting with sick and dying pneumonia patients, Dubos fed polysaccharide coatings of the type III pneumococci to a soil sample from a New Jersey cranberry bog. Most microbes, unaccustomed to polysaccharide diets, lapsed into dormancy, but one bacterium changed its behavior to digest the introduced food. From this bacterium, Dubos isolated an enzyme responsible for digesting the polysaccharide capsules, which he called SIII. Less than three years after Dubos's rash promise to Avery, triumph came: the enzyme cured infected animals by allowing the naked pneumococci to fall prey to the body's phagocytic mechanisms. Further development of the enzyme, however, was eclipsed by sulfa drugs, which were just becoming available.

A peculiarity of the cranberry bog bacterium was that it produced the SIII enzyme only when the capsular polysaccharide was its sole source of food. Significantly, the enzyme production did not involve any change in the genetic constitution of the microbe. The enzyme was produced as an adaptive response to a compelling nutritional force in the local environment. This simple fact heralded a fundamental principle: any living organism possesses multiple potentialities; the ones it expresses depend on external influences.

Such adaptability, Dubos believed, holds for people as well as microbes. "Each one of us," he said, "is born, so to speak, with the potentiality to become several different persons, but what we actually become depends upon the conditions under which we develop. These conditions, furthermore, are often largely of our own choosing."

Dubos had more successes with his soil-culture techniques. He isolated other bacterial enzymes: one, for example, that destroys creatinine, and another that converts creatine into creatinine. The latter is still used in assays to measure creatinine in blood and urine and to determine renal efficiency. He also found an enzyme that turned chocolate into a liquid form that dissolved easily in milk. Sold as Bosco, it remains a treat for children.

In 1938 Dubos recovered and partially purified an enzyme that selectively degraded nucleic acid in yeast cells and pneumococci. He named it ribonuclease. It later proved to be a valuable research tool for other Rockefeller scientists. Stanford Moore and William H. Stein used highly purified ribonuclease in their work on amino acid analysis of proteins, and Bruce Merrifield selected the small enzyme as the first test for the solid-phase method of protein synthesis—discoveries for which the three won Nobel Prizes in Chemistry.

Dubos's enzyme discoveries laid the groundwork for the development of antibiotics. He set out to find an enzyme that would destroy an entire bacterium rather than a specific cell structure. Gram-positive organisms such as pneumococci, staphylococci or streptococci served as the only food source for his soil samples. In 1939 his search culminated in the isolation of *Bacillus brevis,* from which he extracted an antibacterial agent that he named tyrothricin. Tyrothricin proved not to be a single enzyme; instead it contained two polypeptides, tyrocidine and gramicidin.

Tyrocidine is toxic to all living cells; gramicidin, which is active both in the test tube and in animals against gram-positive bacteria, is, however, limited to external use because it destroys red blood cells. (Elsie, the famous Borden cow, who came down with mastitis at the 1939 World's Fair in New York City, was one of the first invalids to respond successfully to gramicidin.) A powerful healer, gramicidin was the first antibiotic to be produced commercially and employed clinically. It is still in use.

Before gramicidin, chemotherapeutic agents were based on such chemical poisons as arsenic, mercury or dyes, which Paul Ehrlich predicted at the turn of the century would act as "magic bullets." None had. The specificity of gramicidin made it what Dubos's collaborator Rollin D. Hotchkiss called a "kinder, gentler drug." Exploiting ecological intuition, Dubos had used soil samples to seek out and exploit naturally occurring bacterial antagonisms. His systematic search provided a rational approach to chemotherapy.

The remarkable contribution of this work, like so many of Dubos's discoveries, was not the isolation of a specific substance. Rather it was his synthesis of major ideas that led to new areas of medical science. Motivated by a philosophical urge, Dubos continually sought to find broad principles at work in nature. He always followed Avery's admonition to be "bold in formulating hypotheses" and "humble in the presence of facts."

Dubos was excited by conceptual underpinnings of problems and by creating hypotheses, but he was notably diffident about his discoveries once they had been made. Just as he did not search for more antibiotics, he was given to abandoning research projects. Some were put aside because techniques and knowledge were not yet available to carry the work further. Other projects were no longer intellectually challenging once he had opened a pathway. Still others became more suitable for commercial laboratories. Throughout his life, Dubos generated new concepts with the conviction that others would develop his initial work. In formulating critical questions, he felt justified that he was "contributing something more effective than doing another experiment."

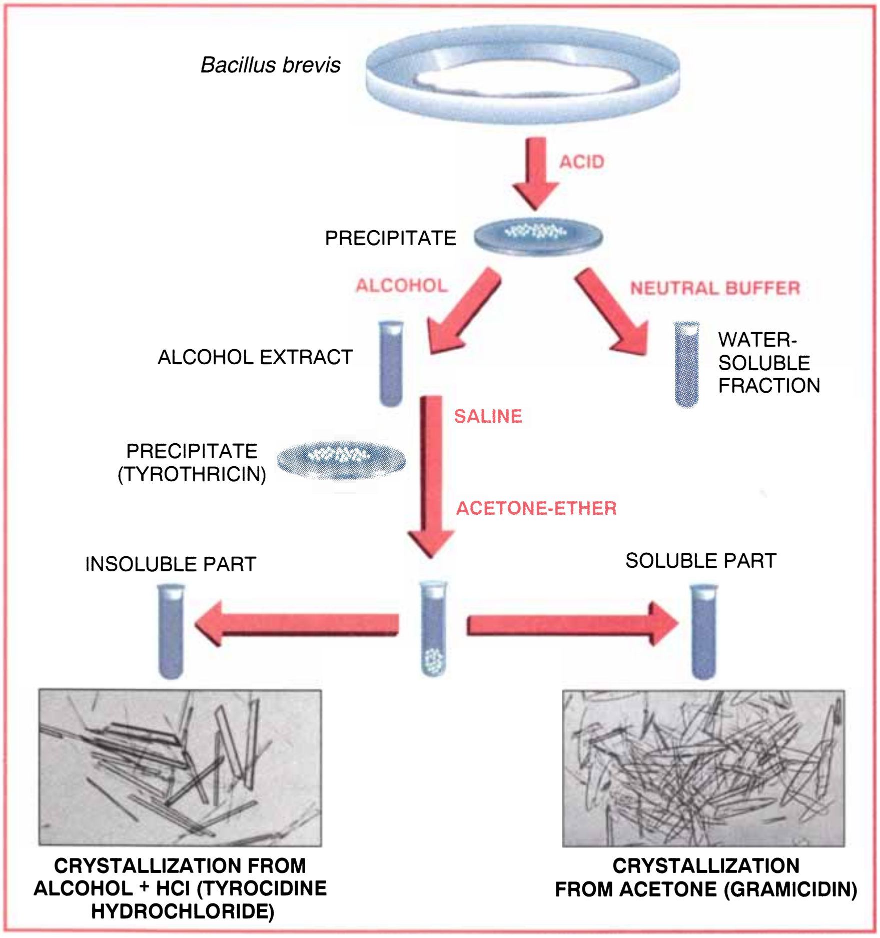
With the gramicidin discovery, medical science acquired a new set of tools. Dubos's work encouraged English scientists Howard Florey and Ernst Chain to revive the dormant research on penicillin, which Alexander Fleming had found accidentally in 1928. Their first report on penicillin's use as a drug appeared in 1940, a year after Dubos's articles on gramicidin. Other scientists began probing the soil for microbes that would produce more antibiotics. Waksman himself undertook numerous searches; one led to streptomycin and a Nobel Prize.

Not surprisingly, Dubos did not use the word "antibiotic" (against life), coined by Waksman in 1942. Dubos referred to such agents as antibacterial or antimicrobial, thereby emphasizing their selective actions that work to renew life and health.

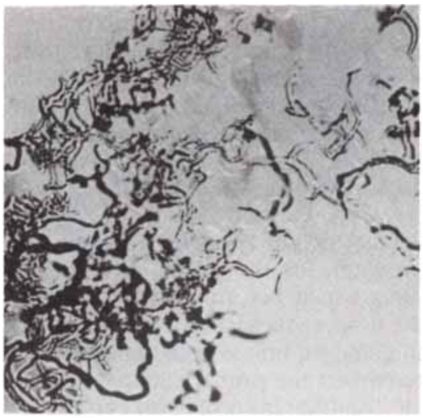
Even before other antibiotics became available, Dubos predicted that bacteria would adapt themselves to these drugs and produce resistant strains. Although he recognized great victories in the battle against fatal infections, he grew critical of chemotherapy. He argued that a drug's effect is determined not only by its action on the parasite but also by the conditions prevailing in the body of the host.

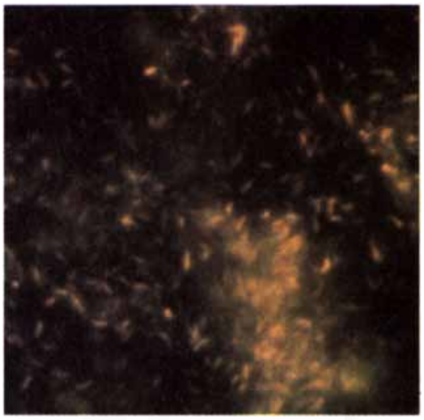
Dubos compared the "conquer mentality" of drugs to the cowboy philosophy in his childhood Western novels: "In the crime-ridden frontier town the hero, singlehanded, blasts out the desperadoes who were running rampant through the settlement," he wrote. "The story ends on a happy note because it appears that peace has been restored. But...the death of the villains does not solve the fundamental problem, for the rotten social conditions which had opened the town to the desperadoes will soon allow others to come in, unless something is done to correct the primary source of trouble." Shifting his ecological perspective from microbes to humans, Dubos decided the problem was not to control disease but to promote health.

A tragic event in his personal life focused Dubos's interest on the human condition in disease: his first wife died of tuberculosis in 1942, just after he had accepted a professorship at Harvard Medical School. Noting that she had suffered from tuberculosis as a child in France, Dubos believed it was reactivated by her anguish over



ISOLATION OF GRAMICIDIN, the first antibiotic, resulted from Dubos's ecological approach. He fed pneumococci, staphylococci and streptococci to soil organisms until he found a microbe—*Bacillus brevis*—that thrived on the pathogenic diet.





TUBERCLE BACILLI grown in ordinary culture media form clumps that cannot be studied easily (*left*). Dubos and his colleague Bernard D. Davis developed a medium containing nontoxic detergents that removed the bacilli's waxy capsule and so allowed rapid, well-dispersed growth (*right*). This advance permitted the first accurate quantitative studies of tubercle bacilli and their pathogenic properties.

family problems resulting from the war. Her disease alerted him to the environment's effects on the balance between human health and bacteria.

Under Avery, Dubos had approached medical microbiology with the 19th-century doctrine of specific etiology, which held that each infectious disease is caused by a single microbial agent. This "germ theory of disease" was a powerful force in the development of medicine because it meant diseases could be prevented or treated by attacking their microbes. Later, Dubos questioned this orthodox approach, observing that "microbial disease is the exception rather than the rule. Why do pathogens so often fail to cause disease after they have become established in the tissues?"

His research took a different direction when he returned in 1944 to Rockefeller and established a laboratory devoted to tuberculosis. He started with a technical advance: the introduction of nontoxic wetting agents into the culture medium to prevent the bacilli from forming clumps as they grew. This innovation permitted Dubos to pursue an approach similar to the one Avery used with the pneumococcus: he looked for morphological and immunologic distinctions between strains of bacilli that caused disease and those that did not. (He also pioneered methods for worldwide standardization of the bacillus Calmette-Guérin [BCG] vaccine against tuberculosis, thereby acting on his belief that prevention is better than cure.)

The morphological investigations were inconclusive. Dubos turned to testing the responses of tubercle bacilli and their hosts to stimuli that disturb the equilibrium in which the host remains at peace with potential pathogens. In these experiments, conducted over a period of 20 years, he incorporated concepts of disease that had preceded the germ theory. He investigated both the possible failure in a host's defense system and the impact of external factors on internal processes. In one phase, he identified naturally occurring tissue substances that affected the growth and viability of tubercle bacilli. And he observed that the complex microenvironment surrounding an inflammation can determine the course of the infectious process. Dubos also found that such environmental influences as diet, toxins, climate, crowding and pesticides affected susceptibility to infection and to disease.

The final phase of this work, with experimental animals, demonstrated that early environmental influences had a lasting effect on growth, development, nutritional requirements and resistance to various stresses. He labeled this phenomenon "biological Freudianism." The overall impact of Dubos's research established disease not as the deterministic result of the presence of pathological organisms but rather as an ecosystem encompassing multiple events. "Its solution," he predicted, "transcends treatment of symptoms in the individual patient and might require social reforms reaching even into the field of ethics."

Working and thinking ecologically, Dubos reformulated the theory of disease causation by implicating the total environment. He showed that a microbe is necessary but not sufficient to cause disease. He found that an infection with pathogenic organisms, such as tubercle bacilli, is not inherently destructive and can persist in a dormant state in the body for long periods. Under suitable conditions, even ubiquitous nonpathogenic microbes can cause diseases.

The important element in disease, he maintained, is not infection but rather any stress—external or internal—that alters resistance, provokes the onslaught of illness and then determines the outcome of the disease. Accepting disease as part of the total ecosystem, Dubos developed a new theme: if we want to improve our physical and spiritual well-being, we must first understand and then control our impact on our surroundings.

Unlike Avery and Louis Pasteur, two of his personal heroes, Dubos did not dramatize his work with a single "protocol experiment" to establish a principle. Instead he organized a campaign to get other scientists interested in an idea and to publicize its significance. His flair for portraying ideas took him from informal conversations to polished lectures. Those he infected with his hypotheses would think about them and come back with objections, new ideas or experimental results.

After a workday dedicated to experiments, colleagues would gather in Dubos's office. There, with feet on his desk and hands folded behind his head or pulling on wisps of hair, Dubos would speculate on what the day's results suggested. As the virologist Frank Fenner recalls, "Any exciting lead formed the base of an inverted pyramid of heady speculation, which often as not collapsed the next day. But we were all, most of all Dubos, enthusiastic about the work, and we were stimulated vastly by the imaginative leaps that René made."

Dubos was also fond of recounting favorite hypotheses and stories in the legendary Rockefeller lunchroom. Holding forth on such topics as growing truffles in the laboratory, infections that produce variegated tulips or why water does not freeze in fire hydrants in winter, he enlivened every meal with his curiosity and imagination. At the bimonthly Hospital Journal Club, Dubos's critiques caused much anguish among junior scientists—it was character building to see one's construction tumbled into a pile of bricks and then reassembled by Dubos into an architectonic masterpiece.

Dubos's comments sharpened his associates' minds and influenced their careers. While Rockefeller colleagues provided an ideal forum for him to rehearse emerging ideas, his penetrating insights and memorable presentations served to concentrate their attention on basic dilemmas. Moore, Stein and Merrifield are only three of dozens whose later fame can be traced to one of Dubos's sparks. (French Nobelist André Lwoff once wrote Dubos that he wished Nobel Prizes were given for "ideas rather than molecules.")

Dubos's ecologically based philosophy of disease is best embodied in *Mirage of Health,* published in 1959. He contended that people will never be free from disease, because they must continuously adapt to environments in flux. He predicted that increasingly crowded, uniform societies would bring forth new diseases. Health, he said, "is not necessarily a state [of] vigor and...well-being, not even...long life. To be healthy does not mean that you are free of all disease; it means that you can function, do what you want to do and become what you want to become." A shocking aspect of this definition, 30 years ago, was that it placed the burden of keeping healthy on the patient and not the physician and scientific medicine.

Outside the laboratory Dubos studied social and historical aspects of disease. Other personal experiences with illness convinced him how disease results from the dynamic processes of life. After his second wife, Jean, convalesced from tuberculosis, they collaborated on a study of environmental factors in tuberculosis in *The White Plague,* published in 1952. Then, ironically, as if providing further evidence in support of this thesis, Dubos fell victim to the pressures of social and professional obligations: he developed a massive gastric ulcer from which he nearly died. Once again he was forced to find a new kind of health: he adopted a plain way of life and an austere schedule to shelter an intense dedication to research, writing and lecturing.

Rarely did he interrupt his work. One exception was to plant and prune trees, particularly the hemlocks that grew luxuriantly on his Hudson Highlands property. He took sheer pleasure in clearing away brush and digging stones out of the soil to make room for his trees and to open vistas. Such physical exertion invigorated him for long hours of writing, frequently done out-of-doors, where trees, insects, birds and

The Shaper and the Shaped

The ability of René Dubos to think on many scales at the same time is most evident in his theories about the importance of the environment in the development of an organism. This principle was clear in the way that the same bacterium might express sharply different traits depending on the properties of its culture medium, in the way that external factors determine whether bacterial infection of a host animal leads to peaceful coexistence or disease and in the way that changing people's surroundings can alter their behavior.

Dubos illustrated this principle by citing the arguments of Winston Churchill in the debate over rebuilding the British House of Commons after it was bombed during World War II. Many people argued that the building should be redesigned to be more efficient and convenient, but Churchill urged that it should be rebuilt precisely as it had been, because to do otherwise would change the style of English parliamentary debate and ultimately the democratic foundations of English society itself. "We shape our buildings," he said, "and afterwards they shape us."

Another example, somewhat closer to home for Dubos, was Jamaica Bay in New York City. By the early 1960s urban development had transformed the bay from a home for oyster beds and a spawning ground for fish, surrounded by marshes that supported waterfowl, to a nearly lifeless body of water whose shoreline was made up almost entirely of landfill. More than 1,500 outflow pipes directed raw sewage into the water.

Residents of the area began fighting for a cleanup of Jamaica Bay, and a city parks department employee, Herbert Johnson, worked to find grasses, shrubs and trees that would colonize the fill. A turning point for the bay came in 1970, when Dubos and others mobilized public opinion to block expansion of Kennedy International Airport into the bay. Dubos ridiculed the idea that a two-year study would be required to determine whether the expansion would damage the Jamaica Bay ecosystem. Common sense, he said, left no doubt as to the adverse impact of dredging the bay and dumping into it several million cubic yards of fill.

Further improvements came with sewage treatment plants that helped to clean the water and with the conversion of roughly half of the bay's perimeter to parkland. Over the past 20 years more than 250 species of birds have returned to Jamaica Bay, as have mollusks and fish. The bay is now a wildlife refuge in the Gateway National Recreation Area, and a peninsula across from Kennedy Airport has been renamed Dubos Point Wetland Park. Dubos often cited Jamaica Bay to show how a few motivated men and women can have a positive effect on an environment, which in turn has a positive effect on many more people. "If I were Billy Graham," he once said, "I would preach to people that the best way to save their souls is to save the environment of cities like New York."

Dubos Point Wetland Park

flowers provided his office furnishings.

René Dubos's transition to human ecologist began in the 1950s. His well-publicized views on the links between human health and environmental forces made him a representative for those disturbed about the earth's health. Although he sounded some of the earliest warnings of ecological disasters, he grew critical of doomsayers. Instead he turned his knowledge of ecological principles toward formulating problems of great complexity in environmental issues. Just as he had disturbed orthodoxy in science and medicine, he restated ecology with bold, new hypotheses. He introduced human contexts in every issue and demanded fundamental changes in the ways people think and live.

Dubos singled out humanity itself as a balancing factor in the earth's health. "There is no 'natural' ecology," he argued. "Man has changed everything in nature." From this unique stance, he addressed environmental crises only as aspects of a larger, more significant problem—the disruption of relationships linking individuals to nature. Although his views on abuses of the earth were well developed, he believed that ecological crises threatened to destroy the quality of life rather than the human population itself. He focused on preserving humanness and set out to stimulate curiosity, adventure and excitement about life's potentialities.

People adapt so unconsciously to their surroundings, he worried, that they would no longer mind the stench of automobile exhausts, ugly urban sprawl, "starless skies, treeless avenues, shapeless buildings, tasteless bread, joyless celebrations." He predicted that loss of sensual perceptions would be compensated by stimulations from loud noises, bright lights and drugs. Just as he had observed microbes in the soil, he described adaptations taking place between humanity and the earth in which one continuously shaped the other. "We do not live *on* the planet earth but *with* the life it harbors and *within* the environment that life creates." Recognizing great dangers from passive adaptation, especially boredom and disenchantment of young people, Dubos even more strongly advocated the human capacity and need to respond creatively to one's surroundings.

At the height of his lecturing fame in the 1970s, Dubos deliberately startled his audiences with seemingly unorthodox questions. "Do you realize," he would begin, "that lawns are unnatural and their upkeep an ecological crime?" "Do you really need medicine every time you get sick?" "Why do you admire that ecological disaster and want to visit that celebrated example of soil erosion, the Grand Canyon?"

He toppled fashionable notions in ecology. "Nature does not know best," and he would describe how inefficient, wasteful and destructive undisturbed nature could be. "There are no resources, only human resourcefulness," and he would describe how aluminum became a resource only after people isolated it. "Humanized nature" is not anticonservationist, he argued, and he reminded listeners that wherever people settle, they transform nature and adapt it to biological needs that have not changed for thousands of years.

A visible scientist with a wise and hopeful message, Dubos stood apart from the environmental movement of the 1970s that viewed humanity as aggressor and nature as victim. Instead he focused on the interactions between people and their surroundings in which each is continuously modified by the other. He recognized that the quality of life was deteriorating but still had great faith in the human potential for renewal, creation and self-transformation. As an elder statesman of the environmental movement, he redefined ecology as a humanistic science.

Dubos dared his audiences to deal with problems at their source: challenge not toxic wastes and dump sites but the need for disposable containers and the resulting wasted resources. Challenge not wetland disruption but the need for more airports or parking lots. Put fundamental needs of life before claims of profit, prestige or power. And he called on individuals to initiate actions. He strongly believed that solutions would not come from "the official proclamations made in great universities, policy statements from governments nor recommendations from expert panels. Rather it is all the motivated individuals of the world who can save it."

A charismatic speaker, Dubos gave striking performances with his charming accent and avuncular manner. Tall, vigorous and rosy-cheeked, with durable white wisps on a balding head, he radiated an inexhaustible joie de vivre. One was immediately drawn in by his attentive blue eyes filtered through thick glasses, a shy yet broad smile and beautiful, large hands that punctuated every sentence. He could sketch intentionally extravagant hypotheses or speak in simple parables. He coined numerous mottoes to simplify his vast messages, among them, "Think globally, act locally," which continues to inspire environmentalists. He used personal anecdotes to illustrate his wide experience of human nature. With an astonishing flow of knowledge and wit amid eloquence, he moved audiences to laughter, anger, tears and standing ovations.

Building on his lectures, he wrote several books in which he developed a science of human ecology. He described the influences of natural and man-made environments on physical and mental well-being. And he emphasized the human ability to form distinct relationships with places, persons and cultures, thus giving rise to rich and diverse habitats and ways of life.

In his last works, Dubos expanded his theories on how individuals can improve on nature and even remove environmental degradation, and he encouraged the responsible use of science and technology. He also amplified his dictum that "wherever human beings are concerned, trend is not destiny." As biologist he deepened the understanding of ourselves, while as humanist he encouraged our capacity to create benign and beautiful surroundings.

By the end of his life, Dubos saw the environmental movement, whose philosophical underpinnings he had helped shape, grow from a small fringe element to a major political and cultural force. He saw the first Earth Day in April 1970 as more than a counterculture celebration. He predicted the public would come "to regard the right to a healthy and pleasant environment as one of its *natural rights*" and to believe that "they are entitled to environmental quality." And on Earth Day in 1980 he set forth what remains our most difficult and pressing environmental problem: defining individuals' rights and duties in the management of this planet's health.

FURTHER READING

THE TORCH OF LIFE. René Dubos. Simon & Schuster, 1962.

SO HUMAN AN ANIMAL. René Dubos. Charles Scribner's Sons, 1968.

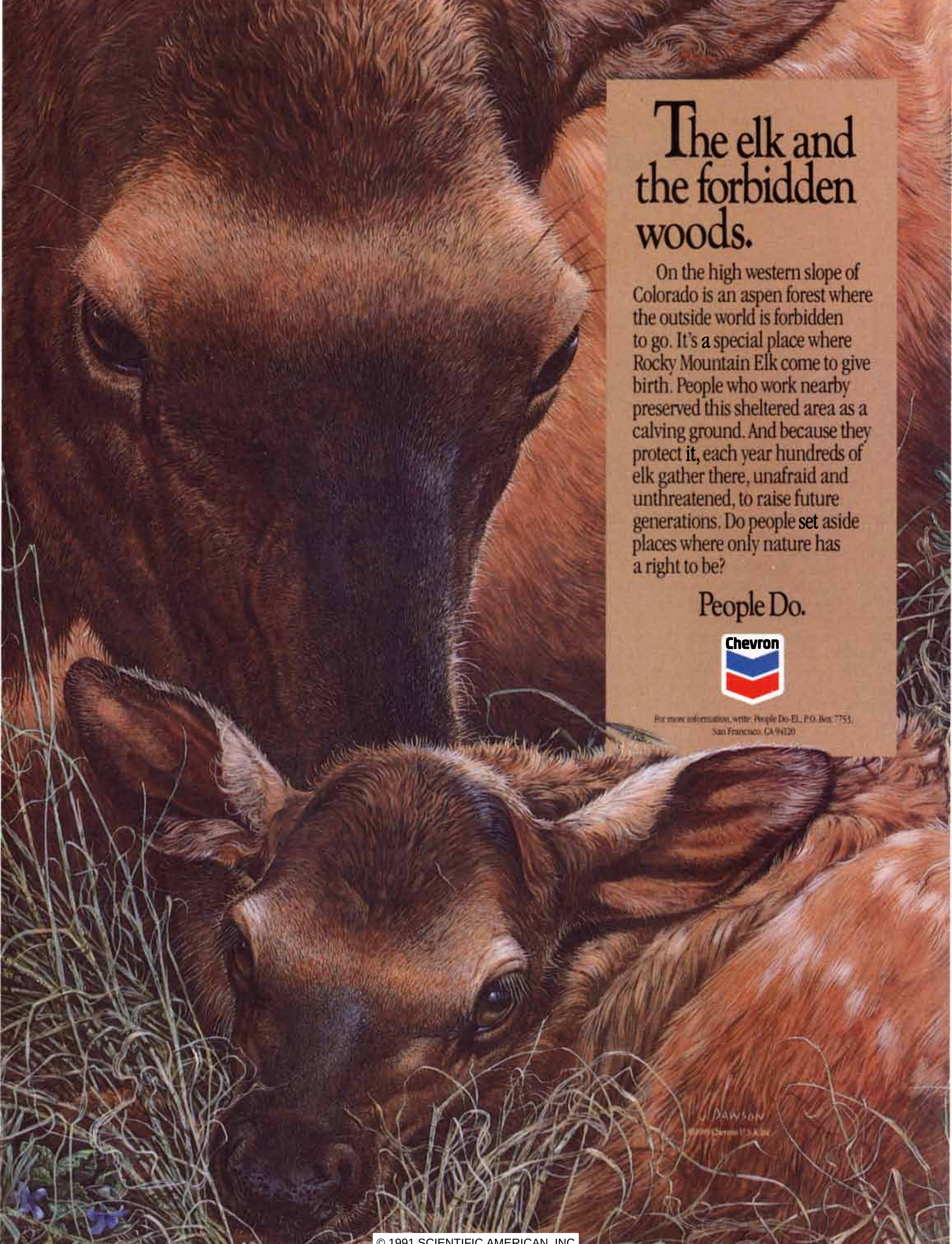
LOUIS PASTEUR: FREE LANCE OF SCIENCE. René Dubos. Charles Scribner's Sons, 1976.

THE WOOING OF EARTH. René Dubos. Charles Scribner's Sons, 1980.

MIRAGE OF HEALTH: UTOPIA, PROGRESS AND BIOLOGICAL CHANGE. René Dubos. Rutgers University Press, 1987.

LAUNCHING THE ANTIBIOTIC ERA: PERSONAL ACCOUNTS OF THE DISCOVERY AND USE OF THE FIRST ANTIBIOTICS. Edited by Carol L. Moberg and Zanvil A. Cohn. Rockefeller University Press, 1990.

THE WORLD OF RENÉ DUBOS: A COLLECTION FROM HIS WRITINGS. Osborn Segerberg, Jr. Edited by Gerard Piel. Henry Holt & Company, 1990.


The elk and the forbidden woods.
On the high western slope of Colorado is an aspen forest where the outside world is forbidden to go. It's a special place where Rocky Mountain Elk come to give birth. People who work nearby preserved this sheltered area as a calving ground. And because they protect it, each year hundreds of elk gather there, unafraid and unthreatened, to raise future generations. Do people set aside places where only nature has a right to be?
People Do.
Chevron
For more information, write: People Do-EL, P.O. Box 7753, San Francisco, CA 94120.
J. DAWSON

The Silicon Retina

A chip based on the neural architecture of the eye proves a new, more powerful way of doing computations

by Misha A. Mahowald and Carver Mead

The eye is the window through which the mind perceives the world around it. It is also a window through which to discern the workings of the brain. The retina, a thin sheet of tissue that lines the orb of the eye, converts raw light into the nerve signals that the brain interprets as visual images. This tiny outpost of the central nervous system must extract all the essential features of the visual scene rapidly and reliably under lighting conditions that range from the dark of a moonless night to the stark glare of the noontime sun.

The retina's ability to perform these tasks outstrips that of the most powerful supercomputers. Yet individual neurons in the retina are about a million times slower than electronic devices and consume one ten-millionth as much power. They also operate with far less precision than do digital computers. Understanding how the retina manages this feat will undoubtedly yield profound insights into the computational principles of other, less accessible regions of the brain.

Clearly, biological computation must be very different from its digital counterpart. To elucidate this difference, we decided to build a silicon chip inspired by the neural architecture and function of the retina. Our artificial retina generates, in real time, outputs that mimic signals observed in real retinas. Our success persuades us that this approach not only clarifies the nature of biological computation but also demonstrates that the principles of neural information processing offer a powerful new engineering paradigm.

Conventional electronic image-processing systems bear little resemblance to the human retina. Typically they consist of a photosensitive array that delivers signals corresponding to the absolute value of the illumination at each point in an image, backed by a formidable computer that attempts to extract geometric features from the resulting digital data.

The retina, in contrast, contains five layers of cells, through which information flows both vertically (from one layer to the next) and horizontally (among neighboring cells in the same layer). The sensing of photons and the processing of the information they contain are inextricably combined. We believe that this architecture is crucial to the formation of visual images.

The top three layers of the retina—photoreceptors, horizontal cells and bipolar cells—are the best understood. These outer layers are the ones whose organization we have chosen to emulate in the silicon retina.

The first layer consists of rod and cone cells that convert incoming light to electrical signals. Horizontal cells—the second layer—make connections to both photoreceptors and bipolar cells through the triad synapse. Each horizontal cell is also connected to its neighbors by gap junctions through which ions diffuse. The potential of any given horizontal cell is thus determined by the spatially weighted average of the potentials of cells around it. Nearby cells make the strongest contribution; distant ones, relatively less.

Each bipolar cell receives inputs from a photoreceptor and a horizontal cell and then produces a signal proportional to the difference between the two. Information from the bipolar cell passes through the amacrine cell layer to the ganglion cells and thence toward the optic nerve.

The most crucial function of these first three layers is adaptation. The photoreceptors, horizontal cells and bipolar cells take widely varying amounts of incoming light and adapt their response to produce a signal with a much narrower dynamic range that nonetheless captures the important information in a scene. Adaptation is necessary if the system is to respond sensitively to small local changes in the image against a background whose intensity may vary by a factor of a million from midnight to high noon.

The retina copes with this tremendous input range in several stages. The first biological trick is to use two different kinds of receptors: rods are sensitive to low light levels and cones to higher ones. Furthermore, the cones themselves can alter the range of light intensities to which they respond, depending on the average long-term brightness in a scene. (These adaptive mechanisms explain why people stepping into bright sunlight from semidarkness experience the scene as washed out and overexposed.)

The bipolar cells have a narrower dynamic range than either the rods or the cones. The crucial element in enhancing their response to the important elements in an image is the triad synapse. The triad synapse mediates feedback between the horizontal cells and the cones. As a result, the bipolar cell does not have to respond to the absolute brightness of the scene; it responds only to the difference between the photoreceptor signal and the local average signal as computed by the horizontal cell network.

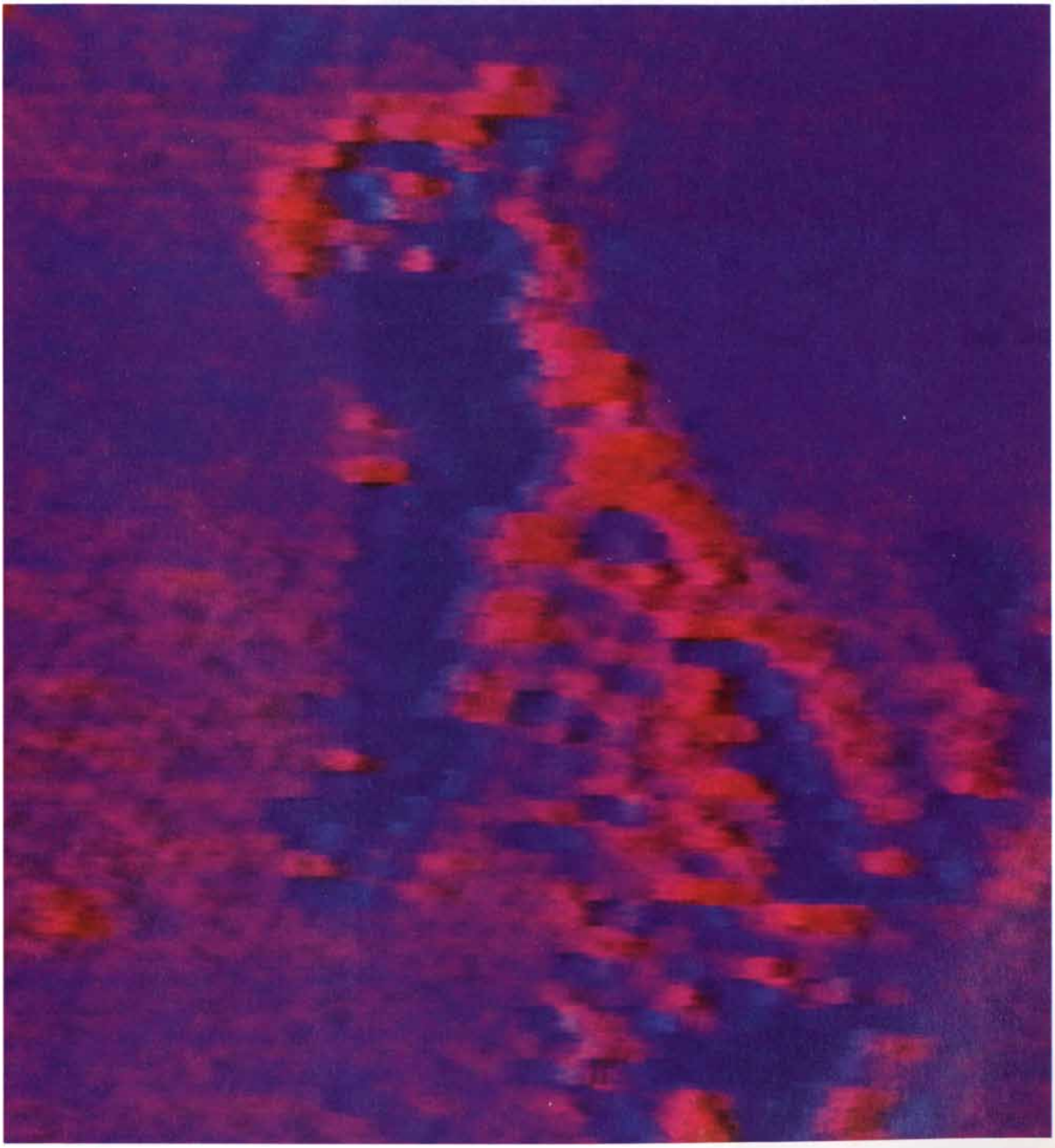
In addition, both the photoreceptors and the horizontal cells produce logarithmic signals, so that the output of the bipolar cell—the difference between the two—actually corresponds to the ratio of local light intensity to back-

MISHA A. MAHOWALD and CARVER MEAD work on analog very large scale integrated circuits at the California Institute of Technology. Mahowald, a doctoral student, designs neuronally inspired vision systems. She received her B.Sc. in biology from Caltech in 1985. Mead is Gordon and Betty Moore Professor of Computer Science at Caltech, where he has taught for more than 30 years. He played a major role in the development of design methods for digital VLSI and is co-author of the standard textbook in the field. He is now working to model in silicon biological structures such as the cochlea and the retina.

ground intensity, irrespective of the absolute light level. Performing further visual processing in terms of the intensity ratio enables the retina to see detail in shaded and bright areas within the same scene.

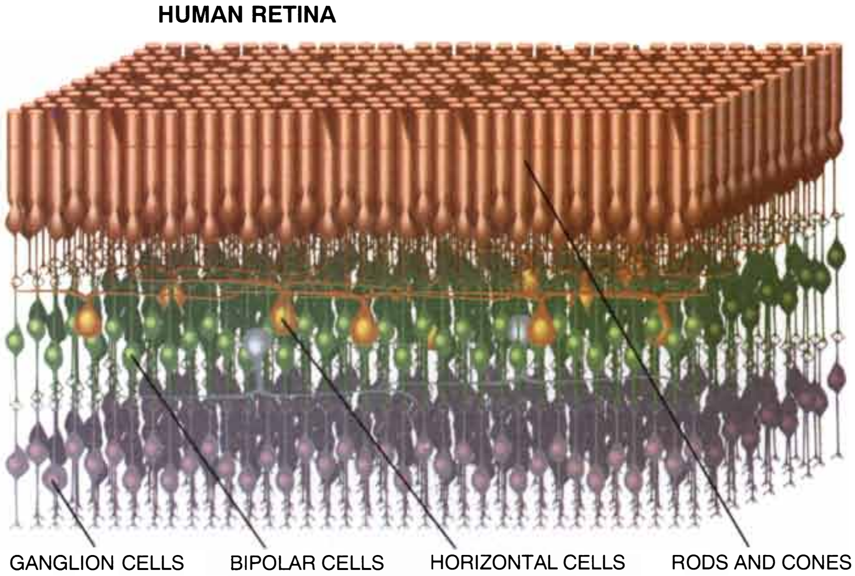
This local adaptation does not just ensure reliable signaling of small changes in image brightness. It also suppresses features of images that are not of interest while enhancing those that are. Large, uniform areas produce only weak visual signals because the impulses from any single photoreceptor are essentially canceled by the spatial average signal from the horizontal cell network. Edges, in contrast, produce strong signals because receptors on both sides of the edge sense light levels significantly different from the local average.

The relatively slow temporal response of the horizontal cell network also enhances the visual system's response to moving images. Photoreceptors produce signals from the image of a moving object while the horizontal cell signal against which they are compared is still reporting the previous intensity level. Unlike a camera, which produces a single snapshot of an image, the ret-

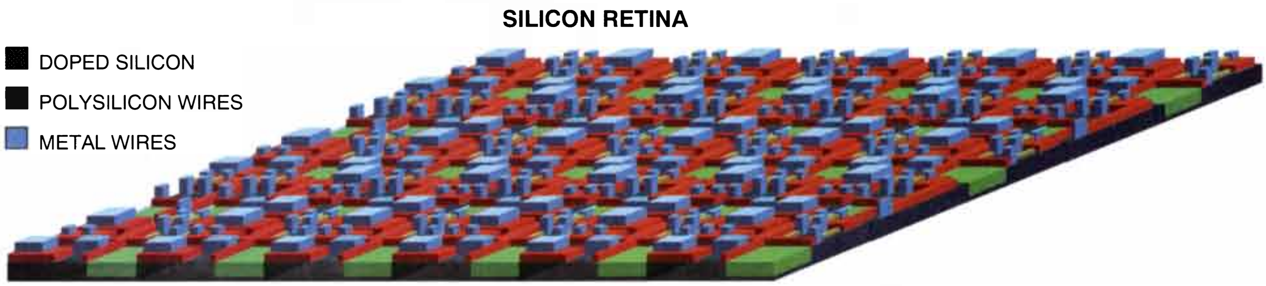


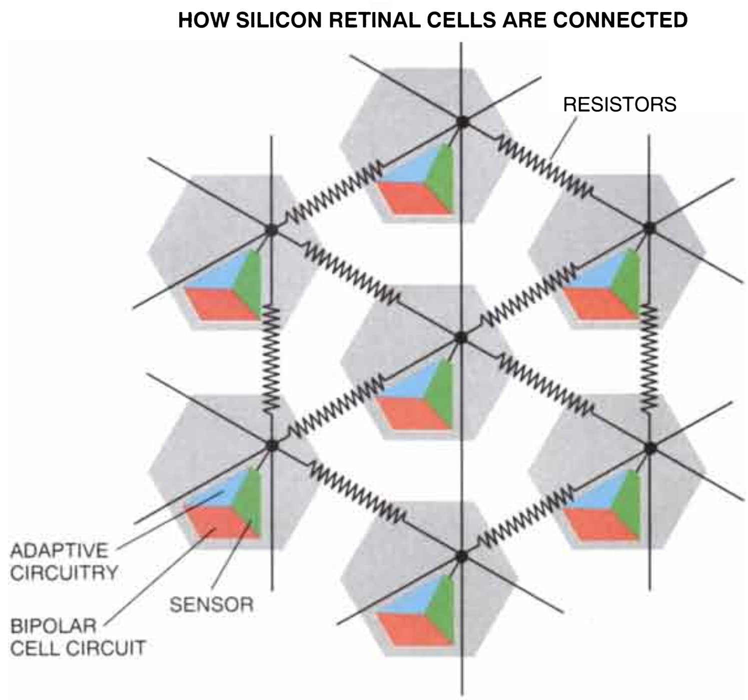
MOVING CAT as seen by silicon retina shows initial stages of biological image processing. (Areas of the image that are darker than their surroundings appear blue; those that are lighter appear red.) The retina responds most strongly to moving images: the cat's head and forelegs appear in sharp relief while stationary parts of its body fade into the background.

Modeling Neural Structures in Silicon



The human retina consists of cells that conduct neural signals both within layers and from one layer to another. The silicon retina models the functions of the outermost three layers—photoreceptors (rods and cones), horizontal cells and bipolar cells. The rods and cones transform light into electrical signals; the horizontal cells, meanwhile, respond to the average light intensity in their neighborhood. Bipolar cells transmit a signal corresponding to the ratio of the signals from rods and horizontal cells through the ganglion cells, where it is further processed before being delivered to the brain.





Each silicon photoreceptor mimics a cone cell. It contains both a photosensor and adaptive circuitry that adjusts its response to cope with changing light levels. A network of variable resistors mimics the horizontal cell layer, supplying feedback based on the average amount of light striking nearby photoreceptors. And bipolar cell circuitry amplifies the difference between the signal from the photoreceptor and the local average. The physical layout of the chip (*above*) contains circuitry in staggered blocks. Silicon areas doped with impurities (*green*) are the basis for transistors and photosensors, polysilicon (*red*) forms wires and resistors, and metal lines (*blue*) act as low-resistance wires. The functional diagram at the left shows the arrangement of receptor circuitry and the hexagonal grid of variable resistors that makes up the horizontal cell network. The response of the retinal circuit closely approximates the behavior of the human retina.

ina devotes itself largely to reporting changes.

By the mid-1980s neuroscientists had learned enough about the operation of nerves and synapses to know there is no mystery to what they do. In no single instance is there a function done by a neural element that cannot, from the point of view of a systems designer, be duplicated by electronic devices. Our goal in building a silicon retina was not to reproduce the biology to the last detail but rather to create a simplified version that contains the minimum structure needed to mimic the biological function.

Each pixel of our model retina consists of three parts: a photoreceptor, horizontal cell connections and a bipolar cell. The photoreceptor includes both a photosensitive element and a feedback loop that mimics the slow adaptive mechanism of cones in the biological retina. The photosensor, a bipolar transistor, produces a current proportional to the number of photons it absorbs. The feedback loop amplifies the difference between the instantaneous photocurrent and its long-term average level. The output voltage of this circuit is proportional to the logarithm of the light intensity.

At its utmost sensitivity, the photoreceptor can form images from light fluxes of about 100,000 photons per second—about the intensity of light from a moonlit scene focused on the chip through a standard camera lens. (That is also near the low end of the operating range of vertebrate retina cones.) Large changes in intensity saturate the photoreceptor response until it has adapted to the new light level.

To imitate the horizontal cells, we built a simple hexagonal network of resistors and capacitors. Each node in the network is linked to a single photoreceptor and, through identical variable resistors, to its six neighboring nodes. The capacitors correspond to the charge storage capacity of horizontal cell membranes, whose fine branchings present a large surface for storing ionic charge from the extracellular fluid. The resistors, meanwhile, model the gap junctions that couple adjacent horizontal cells in the vertebrate retina.

The voltage at each node in the horizontal cell network therefore presents a spatially weighted average of the photoreceptor inputs to the network. By varying the value of the resistor, we can modulate the effective area over which signals are averaged—the greater the resistance, the smaller the area over which the signals can spread. The horizontal cells also feed back to the photoreceptors and reduce their response to areas of uniform intensity.

The final output of each pixel in the silicon retina comes from an amplifier that senses the voltage difference between the output of a photoreceptor unit and the corresponding node in the horizontal cell network. The behavior of this amplifier resembles that of the vertebrate bipolar cell.

The result is a semiconductor chip containing roughly 2,500 pixels—photoreceptors and their associated image-processing circuitry—in a 50-by-50 array. The retina chip also incorporates wiring and amplifier circuits that enable us either to study the output of each pixel individually or to scan the outputs of all the pixels and feed them to a television monitor, which displays the image processed by the entire array. (The retina has gone through about 20 iterations, each requiring a few months for the chip's design and fabrication. It continues to evolve and to generate new, special-purpose designs to test particular hypotheses about image formation.)

The behavior of the adaptive retina is remarkably similar to that of biological systems. We first examined how the output of a single pixel responds to changes in light intensity when the surrounding cells are at a fixed background illumination. The shape of the response curve is similar to that of bipolar cells in the vertebrate retina. In addition, changes in the background illumination alter the potential of the horizontal cell network so that the response curve of the silicon retina shifts in the same manner as in biological retinas.

The silicon retina also has a temporal response that closely resembles that of bipolar cells. When the intensity of light is suddenly increased, there is a large jump in output voltage, equal to the difference between the new input and the previous average voltage stored in the resistive network. The response then settles down to a plateau as the

LINCOLN PORTRAIT (*top*) eventually disappears as the silicon retina adapts itself to an immobile picture. Once the retina has "adapted the image away," substitution of a blank sheet of paper yields a negative afterimage—just as the human visual system perceives afterimages when the eye looks away from bright objects. The bright band around Lincoln's head in the first image arises because the retina enhances the contrast of borders between light and dark areas.

network computes a new average voltage. When the light is suddenly decreased to its original intensity, the output voltage plunges below its original value because the network now has a larger average potential than it had originally. Finally, as the network returns to the original average value, the output also returns to its former state. In a biological retina the slow response of the horizontal cells ensures that rapid full-field changes in intensity—which might correspond to the shadow of a predator passing over an animal—pass through the bipolar cells without attenuation.

In subsequent tests, we found our silicon retina to be subject to many of the same optical illusions that humans perceive. The most obvious illusion is that of simultaneous contrast: a gray square appears darker when placed against a white background than when placed against a black background. Other illusions include the Mach bands (apparent bright and dark bands adjacent to transitions from dark to light) and the Herring grid, in which gray spots appear at the intersection of a grid of white lines [*see box on opposite page*].

Such optical illusions provide important insight into the biological retina's role in reducing the bandwidth of visual information and extracting only the essential features of the image. The illusions are created because the retina selectively encodes visual information. That our retinal model also sometimes generates an illusory output gives us additional confidence in our interpretation of the principles by which the biological retina operates.

The behavior of the artificial retina demonstrates the remarkable power of the analog computing paradigm embodied in neural circuits. The digital paradigm dominating computation today assumes that information must be digitized to guard against noise and degradation. In a digital device, voltages within a certain range are translated into bits having a value of, say, one, whereas voltages within a different range are translated into zeros. Each device along the computational pathway restores the voltages to their proper range. Digitization imposes precision on an inherently imprecise physical system.

A neuron, in contrast, is an analog device: its computations are based on smoothly varying ion currents rather than on bits representing discrete ones and zeros. Yet neural systems are superbly efficient information processors. One reason is that neural systems work with basic physics rather than trying constantly to work against it.

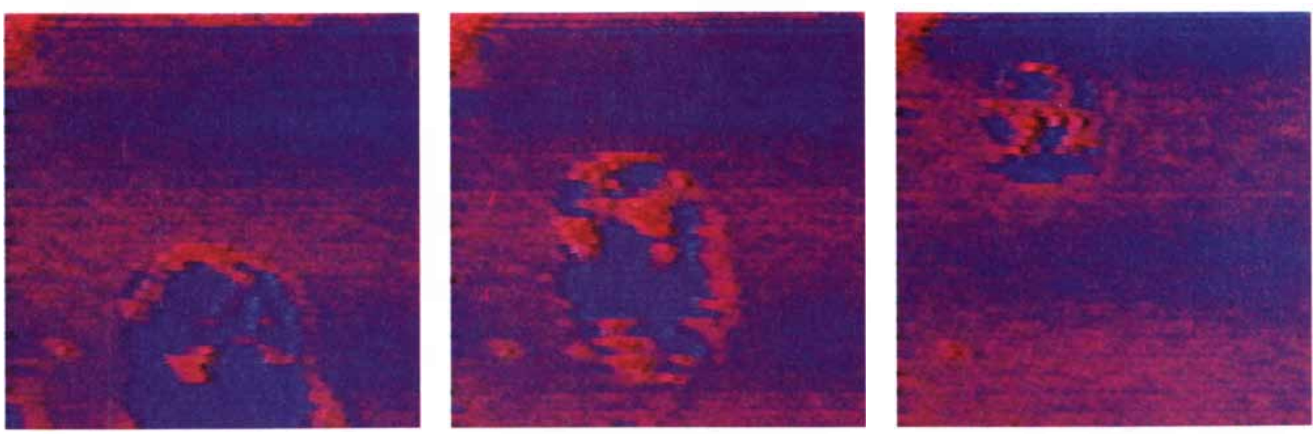
Although nature knows nothing of bits, Boolean algebra or linear systems theory, a vast array of physical phenomena implement important mathematical functions. The conservation of charge, for example, dictates that electric currents will add and subtract. Thermodynamic properties of ions cause the current flowing into a cell to be an exponential function of the voltage across the membrane.

Working with physics helps to explain why the most efficient digital integrated circuits envisioned will consume about 10^{-9} joule per operation, whereas neurons expend only 10^{-16} joule. In digital systems, data and computational operations must be converted into binary code, a process that requires about 10,000 digital voltage changes per operation. Analog devices carry out the same operation in one step and so decrease the power consumption of silicon circuits by a factor of about 10,000.

Even more important, however, the capacity of analog neural circuits to operate in unpredictable environments depends on their ability to represent information in context. They respond to differences in signal amplitude rather than to absolute signal levels, thus largely eliminating the need for precise calibration. The context for a neural signal may be the local average light intensity—as it is when a photoreceptor signal is balanced against the signal from the horizontal cell network at a triad synapse. Or it may be the previous behavior of a neural circuit itself, as in the long-term adaptation of a photoreceptor to changing light levels. The context of a signal may also be some more complex collection of neural patterns, including those that constitute learning.

The interplay of context and adaptation is a fundamental principle of the neural paradigm. It also imposes some interesting constraints on neurally inspired circuits. Because only changes and differences convey information, constant change is a necessity for neural systems—rather than a source of difficulty, as it is for digital systems. When showing an image to the digital retina, for example, we must constantly keep it in motion, or the retina will adapt and no longer perceive it. This requirement for change firmly situates a neural circuit in the world that it observes, in contrast to digital circuits, whose design implicitly assumes separation between the system and the outside world.

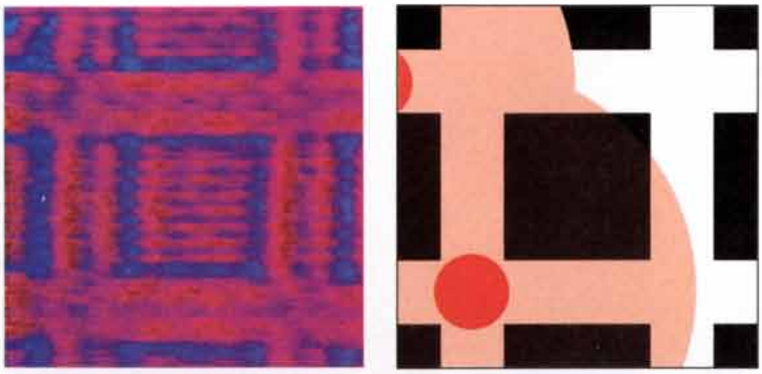
We have taken the first step in simulating the computations done by the brain to process a visual image. How readily can this strategy be extended to other types of brain computations? It may seem that the essentially two-dimensional nature of today's integrated circuits would severely limit efforts to model neural



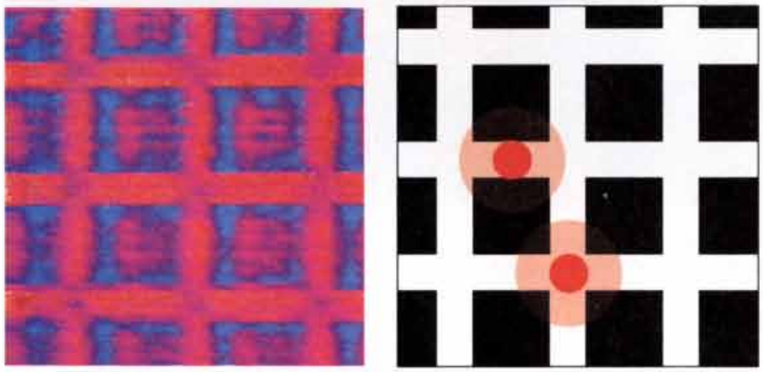
SOCCERBALL in motion shows how the delayed response of the horizontal cell network affects the retina's perception. The ball leaves behind a trail of excitation: bright where the dark spots have just passed; dark where bright parts have been.

Optical Illusions and the Silicon Retina

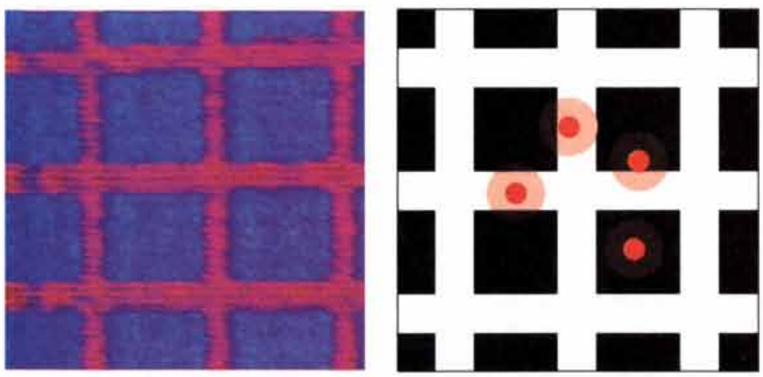
That the silicon retina is subject to some of the same misperceptions as is the human visual system suggests it has captured some essential biological principles. The Herring grid is one well-studied illusion: gray patches appear at the intersections of a grid of black squares on a white background. These patches occur because the retina's response at a given point in the visual field depends on the light intensity at nearby points. (This is the so-called center-surround effect.) The neighborhood of the intersections contains more white space and so reduces the apparent brightness of the intersection itself. A simpler example of the same effect is the illusion of simultaneous contrast (*bottom*), in which a gray square appears darker or lighter depending on the brightness of its background.



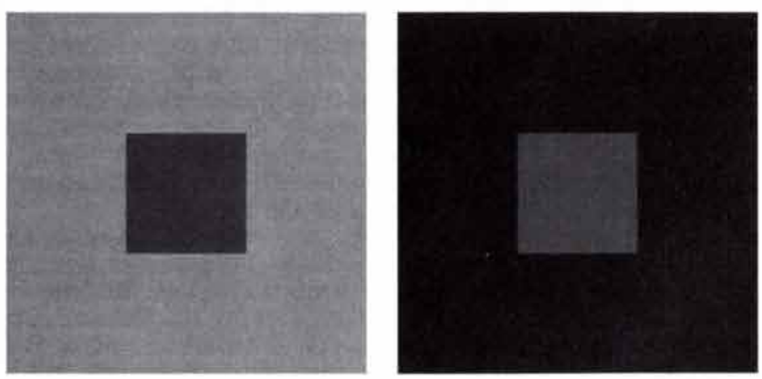
Close-up of the grid reveals no illusory brightness change because both the center and the surround of the receptive field are smaller than the space between the squares.



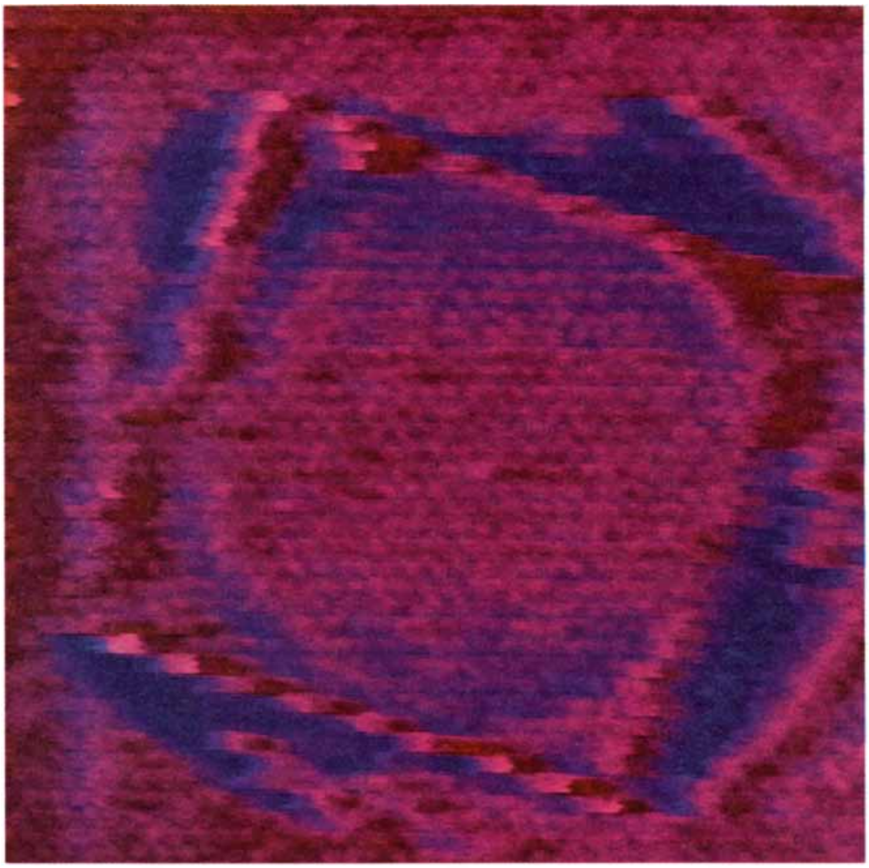
When the size of the center receptive field is comparable to the space between the squares, the illusion appears.



The illusion disappears again when the grid is viewed from a distance, because the average intensity registered by the surround is roughly the same everywhere.



The small squares in both of these images are the same shade of gray. The retina, however, perceives brightness in relation to an object's background, and so the small square on the right appears lighter.



ROTATING SQUARE appears to leave a trail of dark (*blue*) pixels as it spins. The effect results from the slow decay of the voltage in the silicon retina's horizontal cell network: the bright square increases the potential of the network so that background pixels appear dark in comparison. Meanwhile the circular area in the center of the square appears in the background color; its intensity does not change over time, and so the retina adapts it away.

tissue. But many parts of the central nervous system are in fact thin sheets that carry two-dimensional representations of computationally relevant information. The retina is merely the most obvious example. Furthermore, in both neural and silicon systems, the active devices—be they synapses or transistors—occupy no more than 1 or 2 percent of the space; "wire" occupies the remaining area. One can be sure, therefore, that the limitation of connectivity has forced the design of many parts of the brain into a highly specific form.

Specialized wiring patterns are one clear adaptation to situations in which the number of processing elements is limited by the total amount of wire needed to accomplish a computation. The brain's wiring, for instance, ensures that closely related information is mapped onto neighboring groups of neurons. As an example, the cortical areas that perform the early processing of visual information preserve the spatial relations of the image. This maplike organization of the cortex allows most of the brain's wiring to be short and highly shared. Similarly, we designed the silicon retina so that the resistors of the horizontal cell network implement computations for the entire circuit, not just the immediately adjacent cells.

The future development of the silicon retina and similar neurally inspired chips leads along two potentially divergent paths. One is the development of improved machine vision. A single chip containing an array of relatively simple analog circuits, after all, can perform the same functions as a multiple-chip system containing an image sensor and many powerful microprocessors and large memory chips. Some work is already in progress toward binocular circuits—side-by-side silicon retinas that can determine the distance of objects in a scene.

Real vision (or something somewhat closer to it than what exists now) will probably require retina chips containing perhaps 100 times more pixels as well as additional circuits that mimic the movement-sensitive and edge-enhancing functions of the amacrine and ganglion cells. Ultimately such systems will also incorporate additional neural circuits to recognize the patterns that the retina generates.

Another path will take researchers toward a grander objective: understanding the brain. For years, biologists have tacitly assumed that when they have understood the operation of each molecule in a nerve membrane, they will understand the operation of the brain. But both the digital and the analog paradigms of computation make it clear that this assumption is wrong. After all, a computer is built from a completely known arrangement of devices whose operation is understood in minute detail. Yet it is often impossible to prove that even a simple computer program will calculate its desired result or, for that matter, whether the computation will even terminate.

No matter how well the brain's architecture is mapped out, such mapping alone will not lead to a global view of the principles and representations on which the nervous system is organized. The interactions of the computations are simply too complex. If, however, workers can build silicon systems according to a deliberate and well-defined biological metaphor, they may be able to test and advance researchers' understanding of the nervous system.

The success of this venture can create a bridge between neurobiology and the information sciences, and it will also greatly deepen the understanding of computation as a physical process. It will give rise to an entirely new view of information processing that harnesses the power of analog collective systems to solve problems that are intractable by conventional digital methods.

FURTHER READING

THE CONTROL OF SENSITIVITY IN THE RETINA. Frank S. Werblin in *Scientific American,* Vol. 228, No. 1, pages 71–79; January 1973.

THE RETINA: AN APPROACHABLE PART OF THE BRAIN. John E. Dowling. Belknap Press of Harvard University Press, 1987.

ADAPTIVE RETINA. Carver Mead in *Analog VLSI Implementation of Neural Systems.* Edited by Carver Mead and Mohammed Ismail. Kluwer Academic Publishers, 1989.

AN ELECTRONIC PHOTORECEPTOR SENSITIVE TO SMALL CHANGES IN INTENSITY. T. Delbrück and C. A. Mead in *Advances in Neural Information Processing Systems I.* Edited by David Touretzky. Morgan Kaufmann Publishers, 1989.

SILICON RETINA. M. A. Mahowald and Carver Mead in *Analog VLSI and Neural Systems.* Edited by Carver Mead. Addison-Wesley, 1989.

The Genesis of Ores

A precise series of geologic events takes place before metals become minable. Water, magma and other fluids as well as weather and wind play essential roles in the formation of ores

by George Brimhall

When some 80,000 people rushed to California in the mid-19th century to find gold, many of them knew where to look. The now famous forty-niners worked the gravel in modern riverbeds and washed ancient river deposits exposed in nearby cliffs into sluices with jets of water. When these enormous deposits were exhausted, they traced the gold upstream to its source: the Mother Lode, a system of white quartz veins containing gold that is 150 miles long and at points more than one mile deep.

What the forty-niners did not know was how the gold had gotten into the veins, why it was found alongside quartz or why the Mother Lode existed at all. Only recently have scientists begun to understand the processes by which gold and other ores—that is, minerals or rocks containing useful elements such as silver, iron, copper and tin—are formed. They have also discovered where in the earth metals originate. Today it is clear that ores are an integral part of the earth's formation, of its dynamic evolution and of the activities that shape its surface.

Just as the California streams and rivers transported gold nuggets and flakes from veins into streambeds, different fluids above and below the surface of the earth transport other ore constituents. These fluids and their pathways, along with the processes that shape the earth's crust, provide the key to understanding how metallic ores are formed. By tracing the different fluid and tectonic systems that have changed the earth over time, researchers have come to understand the genesis of these valuable deposits.

The creation of ores and their placement close to the earth's surface are the result of much more than simple geologic chance. Only an exact series of physical and chemical events, occurring in the right environment and sequence and followed by certain climatic conditions, can give rise to a high concentration of these compounds so crucial to the development of civilization and technology.

Tracking metals by deciphering changes in the earth's surface is not an easy task. Fortunately, nature has simplified one aspect of this pursuit: the total amount of metal in the earth has remained constant through the ages. Except for a few metals such as lead, which is formed in part by the radioactive decay of uranium and thorium, those metals found in ore deposits and rocks around the world have resided within the earth for its entire 4.6-billion-year history.

The earth's invariant stock of ore metals is found in two geospheres, or physicochemical systems. The outermost geosphere is a thin, exceedingly reactive shell near the surface of the earth. It includes the atmosphere, biosphere, hydrosphere, lithosphere and asthenosphere, on which the oceanic and continental plates move. Below this complex geosphere is another composed of the mantle and the core.

Although almost all metals are primordial—that is, they have been neither created nor destroyed since the beginning of the earth—they have moved from place to place. Complex chemical and thermal interactions between the two geospheres have caused the redistribution of elements, metallic and nonmetallic alike. The migration of certain metals began very early, even as the earth was formed, according to one theory, by the accretion of meteorites in the early solar system.

The protoearth started with approximately the composition of a primitive meteorite—an idea supported by Brian H. Mason of the U.S. National Museum of Natural History in Washington, D.C., and Alfred E. Ringwood of the Austra-

GEORGE BRIMHALL is professor of geology at the University of California at Berkeley. After completing graduate work there in 1972, he worked for the Anaconda Company doing mine development, exploration and geologic research. Brimhall then taught at Johns Hopkins University before returning to Berkeley. His research interests include the geology and geochemistry of mineral deposits in the crust. He is currently working with students and colleagues to integrate the mechanisms of metal transport with those of soil genesis, hydrology and geomorphology. When he is not tracking metals, Brimhall and his wife and two daughters enjoy hiking mountain trails.

LA ESCONDIDA COPPER MINE in the Atacama Desert of Chile provides geologists with a window into the earth. The region's slow erosion rate and aridity were crucial to both the enrichment and preservation of copper some 15 million years ago.

lian National University in Canberra. Meteorites initially accreted together at slow speeds, but the pace accelerated tremendously as the earth grew larger and its gravitational attraction increased. As a result, the primitive meteoritic mass was entirely restructured, and the earth differentiated into discrete concentric shells.

Each shell is composed of markedly different metals. Iron droplets separated from the mantle, sank and coalesced to form the metallic core as did other siderophilic, or metal-loving, elements, such as nickel, cobalt, gold and platinum. In contrast, the elements concentrated in the mantle are lithophilic, or rock-loving, ones, such as silicon, magnesium, aluminum and calcium. Each combines easily with oxygen, the most common element in rocks, to form minerals called silicates within the mantle and crust. Next in prevalence after the lithophilic elements, which constitute the bulk of the crust, are the ore-forming elements: largely chalcophilic, or sulfur-loving, metals, such as copper and zinc, along with volatile elements, such as chlorine and sulfur.

These different ore elements, trapped for billions of years deep within the interior of the earth, occasionally ascend toward the surface where their fate is determined. They rise in magma, which is buoyant and fluid relative to the mantle. And depending on the type of environment or the other substances encountered during the ascent, these metallic elements either become minable metals or remain useless rocks.

For example, encountering sulfur can determine a metal's form. In the presence of large amounts of sulfur, metallic elements precipitate as metal-rich sulfide minerals that are distinct from common rock-forming minerals. Sulfide minerals, because of their high metal content, have unusual physical properties, such as high density, so they can be industrially gathered from surrounding waste rock. Without ample sulfides, metals join with rock-forming minerals only in trace concentrations—parts per million or less. These amounts are too low to be recovered economically. In common rocks, metals occur in this dispersed and useless form.

Most knowledge about the pivotal processes—physical, chemical and mineralogical—that lead to the concentration of ores comes from mining. Mines, open pit and underground, are natural laboratories of priceless scientific value in which geochemical history can be interpreted. Additional windows on the earth are provided by thousands of drill holes that may penetrate the crust to depths of more than 12,000 feet. With information gathered by studying mines around the world, researchers, myself included, have come to understand the varied geologic and geochemical processes that must act in concert to create and preserve high concentrations of metals in the crust.

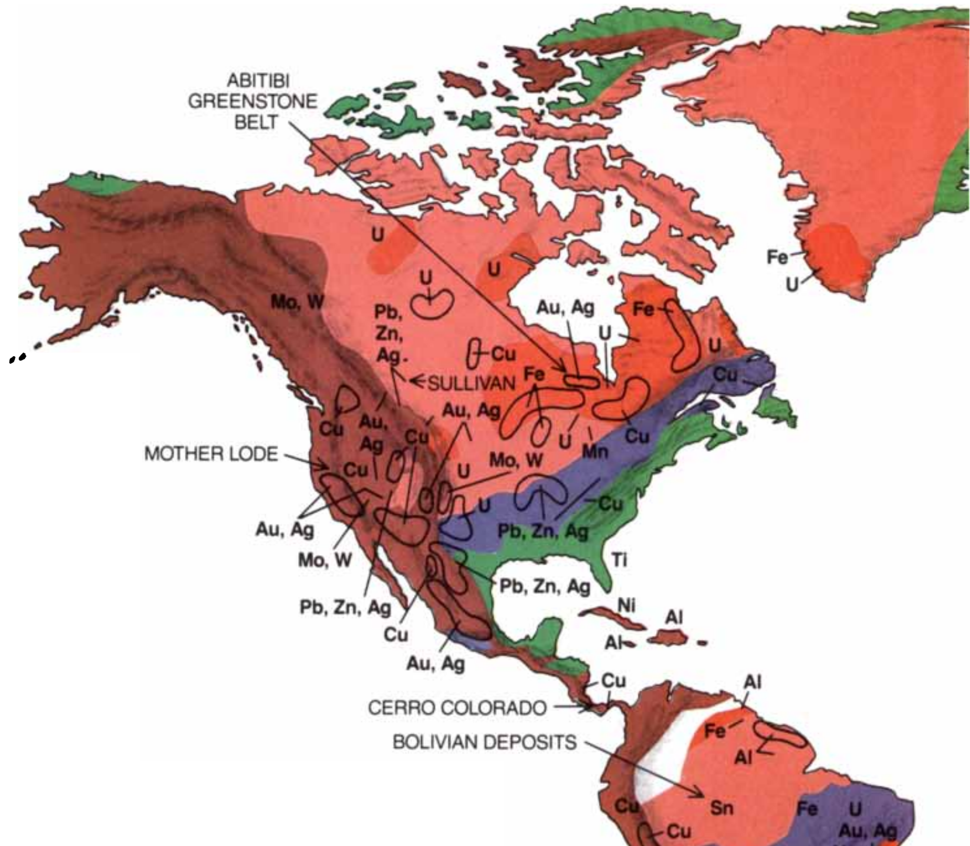
The introduction and redistribution of metals into the crust require enormous forces. The sheer magnitude of this activity can be seen, for instance, in a large copper deposit such as La Escondida in Chile, which was formed as magma rose to the surface. This deposit contains 1.8 billion tons of ore that yields 60 billion pounds of minable copper, according to James A. Bratt of BHP-Utah International. Such a concentration is 300 times greater than the normal crustal abundance of 55 parts per million and is worth more than $60 billion.

My work has shown that the process of enrichment began very early in the earth's history with the evolution of the crust and culminated with a series of events bringing together ore metals with sulfides and oxides. To be preserved, these metal-rich rocks formed near the earth's surface had to be protected from erosion. Accordingly, I have explored the importance of the environment in ore preservation. The surroundings had to be stable for significant periods of geologic time while uplift and erosion had minimal effect.

Indeed, such preservation occurs only if ores are incorporated into long-lived crust. Almost all the known ore deposits, even the very old ones, are found in continental crust, either in rock sequences that reflect the proximity of continental crust at the time of mineralization or in fragments of oceanic

crust that have been incorporated into continental crust [see "The Continental Crust," by B. Clark Burchfiel; SCIENTIFIC AMERICAN, September 1983].

Rarely does oceanic crust survive, however, because the mantle is in a state of constant circulation, or convection: magma rising from beneath spreading seafloor ridges erupts and is conveyed laterally until it sinks beneath the continents in areas called subduction zones. The latest seafloor spreading began about 200 million years ago and produced the present placement of continents and oceans. Less than 0.001 percent of the oceanic crust created over the course of that spreading has been preserved. This estimate by Robert G. Coleman of Stanford University suggests that the rest of the oceanic crust disappeared into the mantle through subduction.



In addition to knowing the locations of concentrated ore deposits, geologists know a lot about the rocks that host them. Usually metal concentrations form in sedimentary and volcanic rocks created under water, on land or at least very near the surface. They are therefore referred to as supracrustal rocks. Only supracrustal environments provide all the factors necessary for ore deposition, including the geologic preservation of ores. It is fortunate for human purposes that ores are usually found at minable depths.

It is not all luck, however; the location near the surface is unique for several reasons. First, the environment is characterized by sharp gradients in physical and chemical conditions because the earth's surface is the interface between the solid earth, atmosphere and hydrosphere. In particular, a steep thermal gradient there causes ore minerals to precipitate during the cooling of magmas and hydrothermal solutions.

Perhaps the most significant aspect of the surface that enhances ore deposition is the unusual physical properties of near-surface rocks compared with those in the deep crust and mantle. At great depths, where pressures are high, open fractures in rock are relatively uncommon, but at the surface fractured rocks predominate. The low temperatures and pressures found near the surface cause rocks, which are brittle, to crack suddenly under tectonic and local stresses. In contrast, higher temperatures and pressures deep in the earth cause rocks to accommodate stresses more gradually—they give way by squeezing or by plastic flow and are consequently less permeable.

Cracks allow the passage and rapid circulation of magmatic and aqueous fluids, which are in turn critical to the process of ore transport. Silicate magma is one of several fluids that transport metals. Once elements are near the surface, a number of liquids take over, including water escaping from magmas, steam, seawater and groundwater. These fluids are highly mobile in comparison with viscous magmas. They are also very reactive chemically, making them excellent solvents for salts, acids and bases, all of which solubilize metals efficiently.

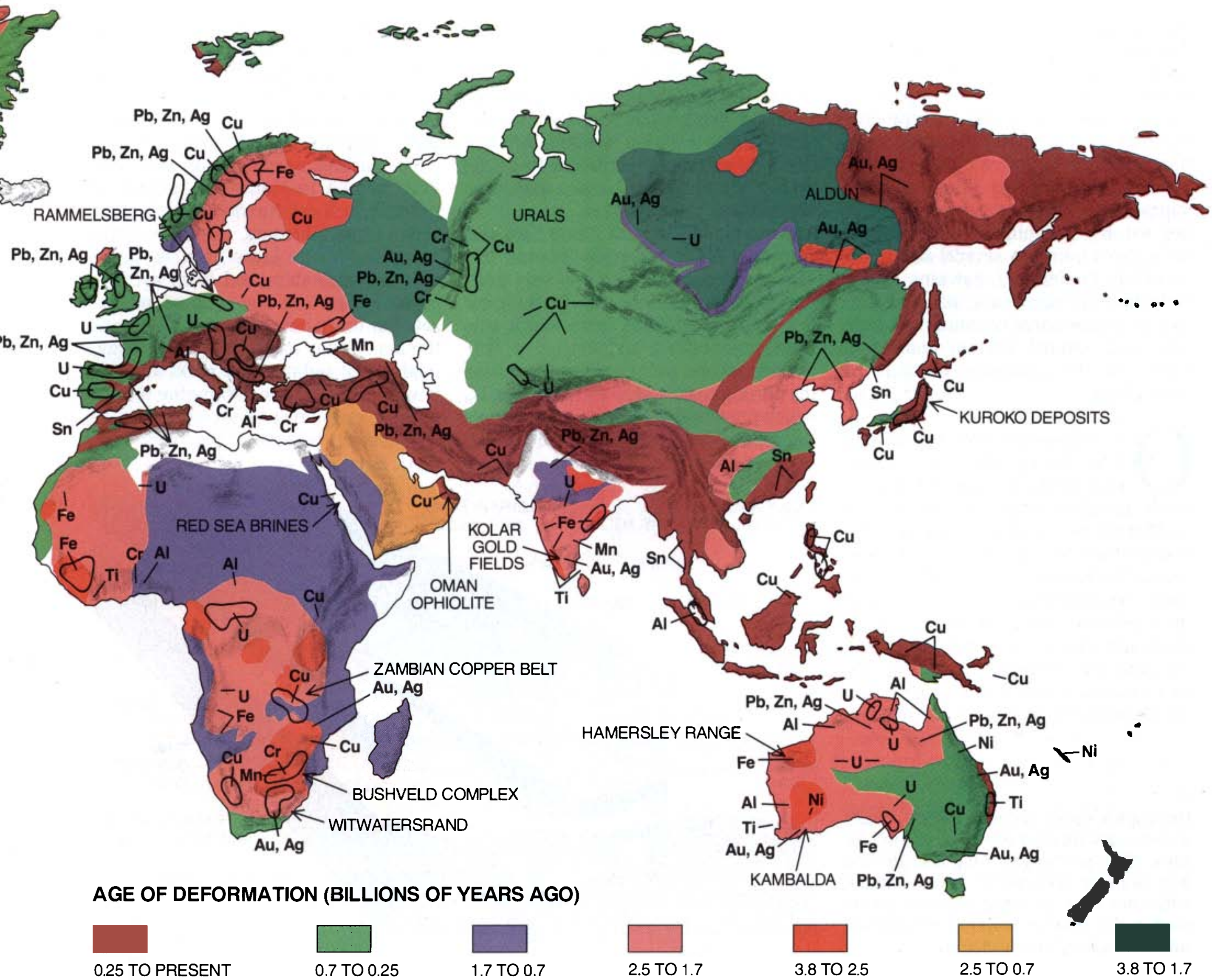
Through fracture networks, fluids such as groundwater or seawater have easy access to the rocks formed in the subsurface. Vigorous chemical reactions take place between these fluids and the minerals exposed on rock walls, affecting the composition of both the rocks and the fluids by mutual interaction.

When the composition of the ore-bearing fluid is changed, its capacity to transport metals diminishes. For example, acid-oxidizing liquids become neutralized and reduced as they pass through rocks and alter minerals. Oxidation-reduction and hydrolysis reactions are common causes of ore-metal sulfide precipitation.

In addition to reacting chemically, fluids can become fractionated, or distilled into various phases. Metals can preferentially enter one phase over the other and separate from the parent fluid. These later phases include steam and dense brines.

Over the course of geologic history, the interactions between fluids and rocks have changed. The evolution of different tectonic environments, as well as that of the biosphere and atmosphere, modified the composition of the rocks, minerals and fluids involved. Many researchers have shown how such changes are reflected in ore deposition, including Richard W. Hutchinson of the Colorado School of Mines, William S. Fyfe of the University of Western Ontario, Heinrich D. Holland of Harvard University and the late Charles Meyer of the University of California at Berkeley.

The distribution of major mines can reveal much about the evolving nature of the crust and of the fluid-transport mechanisms. These mines describe the relation of the earth's ore-forming "engines" to the



SOURCE: B. Clark Burchfiel

DISTINCT ERAS of crustal change correspond to large ore deposits. As the early continents grew by the accretion of crust, varying kinds of ores were deposited. For instance, iron was concentrated primarily in early Proterozoic rocks (*pink*) that surrounded older Archean rocks (*red*). Ore deposits also correspond to processes such as orogenesis, or mountain building.

ELEMENTS

Cu	COPPER	Ni	NICKEL
Fe	IRON	Mn	MANGANESE
Au, Ag	GOLD AND SILVER	Al	ALUMINUM
Pb, Zn, Ag	LEAD, ZINC AND SILVER	Ti	TITANIUM
Mo, W	MOLYBDENUM AND TUNGSTEN	Sn	TIN
Cr	CHROMIUM	U	URANIUM

forces shaping the face of the earth and to the underlying deformations that provide fluids from greater depths access to the surface.

To understand the complex global geochemical patterns of numerous types of ore deposits, researchers must categorize mines by the dominant metal, the character of the environment where they were deposited and the age of the deposit. Such classification is necessary because ore formation and related geologic events occur by spontaneous processes. These events dissipate energy, reflecting the slow but irreversible degradation of the earth's energy sources. This idea suggests that the present may not always provide a perfect key to the past, thus contradicting the underlying principle of geologic interpretation: uniformitarianism.

By considering the spatial and temporal occurrences of ore deposits, we can also understand other aspects of the transport engines. As one can see on the map, the most obvious correlation is between mineralization and orogenesis, or mountain building. Low-grade copper deposition, caused by the circulation of copper-rich water from magmas, takes place almost exclusively in zones within the youngest orogenic belts: the Chilean Andes and the North American Cordillera.

Other types of deposits have a well-defined spatial and temporal distribution, but they seem completely unrelated to present tectonic patterns such as orogenesis. To make sense of the diverse types of ore deposits, I will organize ore-forming environments not only by the global distribution of mines described above but by the nature of the transport fluids and the supracrustal tectonic age regimes as defined by Hutchinson. This framework owes much to several scientists, including Frederick J. Sawkins of the University of Minnesota, Richard Sillitoe, an independent consultant in London, and Samuel Epstein and Hugh Taylor of the California Institute of Technology.

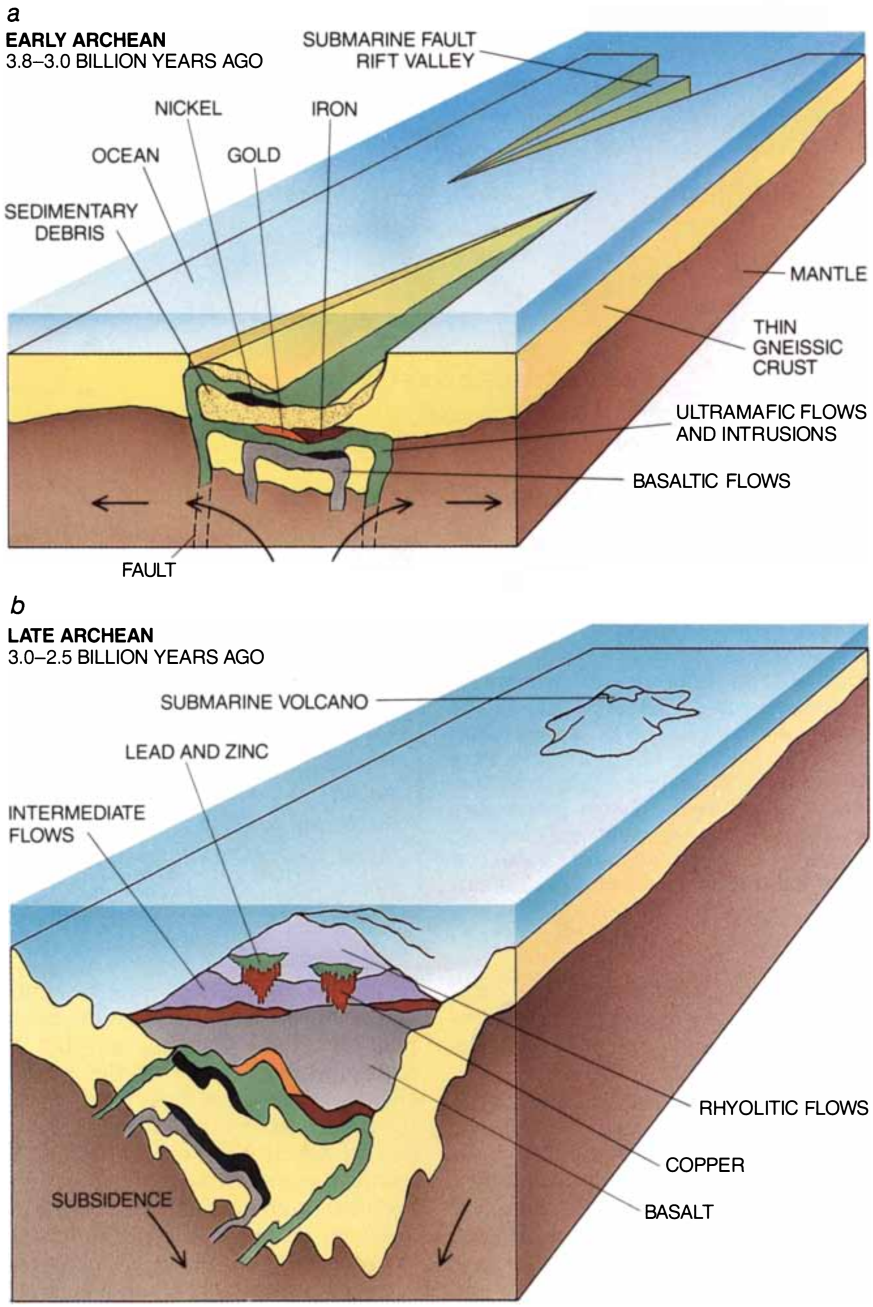
Ore deposition can be divided into five generalized tectonic eras, each characterized by a different transport engine for metals and a different group of ore deposits. (Of course, the older the process, the less certain the interpretation.) I will refer to these five periods as the early Archean (which lasted from 3.8 to 3.0 billion years ago), the late Archean (3.0 to 2.5), the early Proterozoic (2.5 to 1.7), the mid- to late Proterozoic (1.7 to 0.7) and the Phanerozoic (0.7 to the present).

Geologic evidence prior to the early Archean is generally lacking either because the earth was intensely bombarded by meteorites or because the crust was completely recycled into the mantle. But starting around 3.8 billion years ago, we can begin to document the enormous changes that have shaped the earth's surface.

In particular, we can study how the transport of ore changed. During the early Archean, most ore minerals were deposited on or near the seafloor by the eruption of lavas from which sulfides precipitated and settled downward in layers. Heated seawater coursing through lavas also leached metals, which precipitated and rained down on the seafloor near submarine hot-spring vent areas. More recently, meteoric, or fresh, water became involved in ore genesis. (It is unclear, however, whether the record of this change is real or simply caused by the destruction of older examples of such surficial ores.)

The earliest record shows that ores first formed in elongated submarine troughs. In these furrows, which often occur in parallel sets, primitive ultramafic and mafic lavas (that is, those rich in magnesium and iron) ascended to the surface and erupted on the seafloor during the rifting of protocontinents. Today chert, an often gray, quartz-rich sedimentary rock, and pillow lavas reflect the submarine nature of these early volcanic layers.

Extensive oceanic crust was also



a
During the early Archean period, crustal tectonics formed ore deposits in seafloor rifts, or troughs. Magnesium- and iron-rich, or ultramafic, lavas reacted with seawater, creating primary greenstones and their associated deposits of nickel, copper, iron and gold.

b
Changes in the late Archean caused the melting of older crusts and formed ore deposits in newer, or secondary, greenstones. Lava erupted on the seafloor and reacted with seawater, causing metals such as zinc and copper to leach out and sulfide deposits to precipitate.

c
During the early Proterozoic period, tectonic faulting exposed and eroded older ores, and rivers carried them away. Banded iron formations precipitated from seawater, and intrusions into the continental crust created deposits of chromium and platinum.

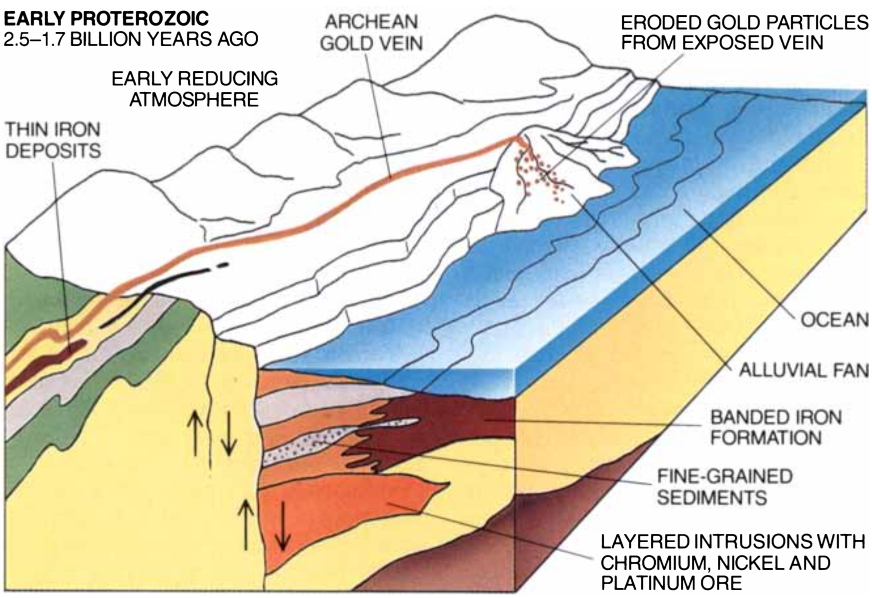
d
After the oxygenation of the atmosphere during the mid- to late Proterozoic, layered deposits of copper, uranium, zinc and lead were formed. Dense brines escaping from sedimentary basins reacted with limestone to form lead and zinc sulfide deposits.

made in this fashion. Its composition was modified as seawater circulated through fractured mafic lavas, producing greenish secondary minerals. Because of their color and because they are the earliest-known type of oceanic volcanic crust, Andrew Glikson of the Bureau of Mineral Resources in Canberra called these layers of minerals primary greenstones.

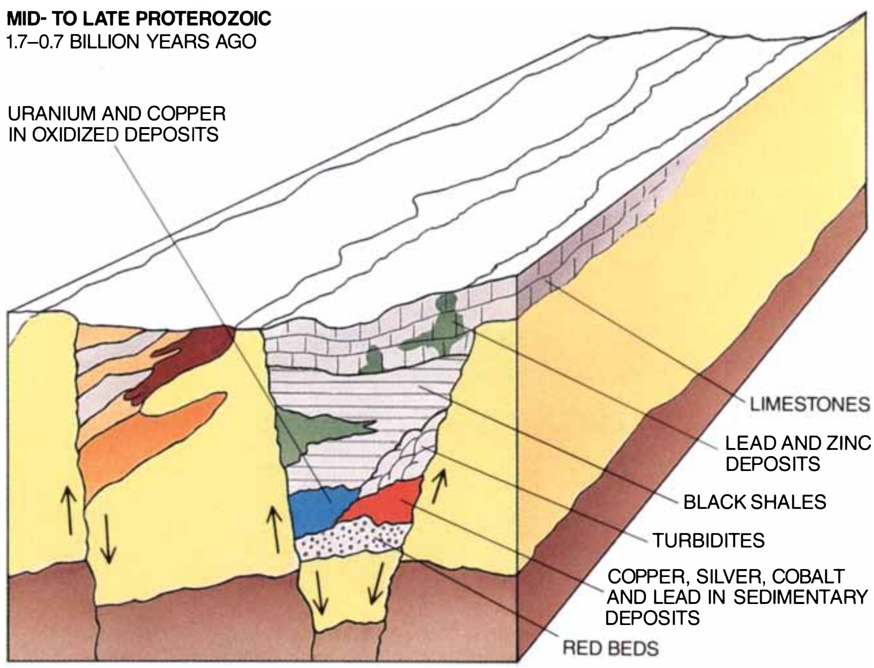
Primary greenstone belts are the host rocks for three major types of ore deposits. Nickel and copper sulfides, occurring as droplets in magmas, settled to the base of the ultramafic lavas, forming some of the rich ores of Kambalda in western Australia, first recognized by Roy Woodall of the Western Mining Corporation in Australia. Chemical reaction of seawater with lava on the seafloor also caused the precipitation of iron-rich sediments.

The third type of deposit found in primary greenstone belts consists of gold-rich sediments: chert, carbonate and iron sulfide formations. They are similar to the volcanic and sedimentary strata described above and were also formed by eruptions on the seafloor that involved seawater circulation. During later low-grade metamorphism, these layered protores (proto-ores) were leached of their metal content, which later became concentrated in the form of lodes, or veins, as in Quebec's Abitibi Greenstone Belt or the Kolar Gold Fields of India. Surprisingly, these ancient rocks have remained relatively unscathed.

c



d



Eventually, in the late Archean, the continued submarine eruption of dense ultramafic lavas caused the primary greenstone sequences to subside. As they reached greater depths, these layers were subjected to increasing temperatures, greater metamorphic effects and even melting. On melting, new magmas were created and ascended to the surface, often piled one on top of the other in the form of sequences of basalt and rhyolite, a silica-rich rock similar to granite. These are called secondary greenstones because their formation involved the fusion and the differentiation of more primitive lavas.

Also during the late Archean, massive zinc and copper sulfide ores were formed on the seafloor by dense, metal-rich brines leaching metals from hot volcanic piles below. Underneath these sulfides, which formed near seafloor vent areas, lay feeder zones interlaced with copper sulfide veins. Sulfides were usually produced after a rhyolitic phase of eruption and during the cooling of the underlying volcanic material before the next eruption.

The third supracrustal tectonic environment spanned the early Proterozoic. Ore deposits then formed in troughs that were near or in continental rift systems or aboveground. Evidence suggests that the rifts formed near continental margins with no major orogenic volcanic activity.

Older uplifted blocks of crust containing earlier ore deposits eroded, creating this era's sediment. Alluvial fans deposited sediments containing gold. (Streams could concentrate gold because it has a high density compared with common minerals.) Heavy minerals were concentrated near the bottom of channels. Accordingly, ancient river, or placer, deposits, such as the Witwatersrand in South Africa, harbor a major portion of the world's gold reserves.

At the same time, waste pyrite, a mineral composed of iron and sulfur, was being transported by streams. This migration indicates that the atmosphere was less oxidizing than it is today. Another indication of the state of the atmosphere in the early Proterozoic comes from the widespread distribution of layered deposits of iron oxides covering huge regions, such as the Hamersley Range of Western Australia and the Lake Superior region. According to Holland of Harvard, these banded iron formations were formed by chemical sedimentation. Cold, oxygen-deficient seawater rich in dissolved iron met with oxygenated surface waters. This encounter led to the precipitation of ferric oxides.

But the creation of banded iron formations ceased abruptly about 1.8 billion years ago when the atmosphere became oxygenated. The mobility of dissolved ferrous iron was limited as it became oxidized into a highly insoluble form. This metal, after it was freed by the chemical weathering of rocks, was immediately fixed at the outcrop as stable oxide minerals—either hematite or magnetite; it could not proceed into rivers and then to the ocean as before.

Simultaneously, another type of ore deposit began to form in cratonal regions, the growth centers of continents. For the first time in geologic history, mafic intrusions invaded a thickened continental crust and filled large magma chambers. After interaction with surrounding wall rocks, these magmas precipitated deposits of chromium and iron and titanium oxides as well as the major platinum deposits of the world: the Bushveld Complex of South Africa and the Stillwater Complex of Montana.

After the oxygenation of the atmosphere, the mobility of many metals—aside from ferrous iron—increased. This change marks the beginning of the mid-Proterozoic era, about 1.7 billion years ago. Continental rifts were again the site of metal deposition. Red, oxidized continental sediments encasing hematite formed. The ocean advanced over these red beds, leading to the deposition of organic and sulfide-rich sediments. This layered sequence facilitated chemical reactions between the two, and the interactions in turn provided the basis for high levels of metal transport.

Evaporation in arid regions at the edge of continents began to induce the upward flow of water from the oxidized red beds. This flow brought metals into contact with overlying black shales, which held abundant sulfides. As a result, cobalt and copper sulfides were deposited in shales, such as the Zambian Copper Belt of Africa.

Like copper, uranium becomes increasingly mobile at the surface. The weathering of granites containing uranium resulted in major redeposition events near unconformities and in continental rifts that had coarse sedimentary debris. (Unconformities are areas between sedimentary rock and the rocks below them, which indicate that erosion rather than deposition took place.)

At the margins of thick sedimentary basins, limestones were replaced by acid brines expelled during compaction of the basin. Lead and zinc deposits in the Mississippi Valley formed in this manner by neutralization, as described by Dimitri A. Sverjensky of Johns Hopkins University.

The Phanerozoic is the fifth and latest supracrustal era of ore deposition. Because it is the most recent era in the history of the earth, its complexities are quite well understood. Many processes of fluid transport still operate today and may be studied directly by observation rather than solely by deduction.

The present configuration of the continents and oceans results from protocontinental fragmentation that began about 200 million years ago with seafloor spreading. Large convection engines within the mantle divided the lithosphere into vast mobile plates. Most deposition of ores occurs at the margins of these plates, at the points where they are created or deformed and consumed.

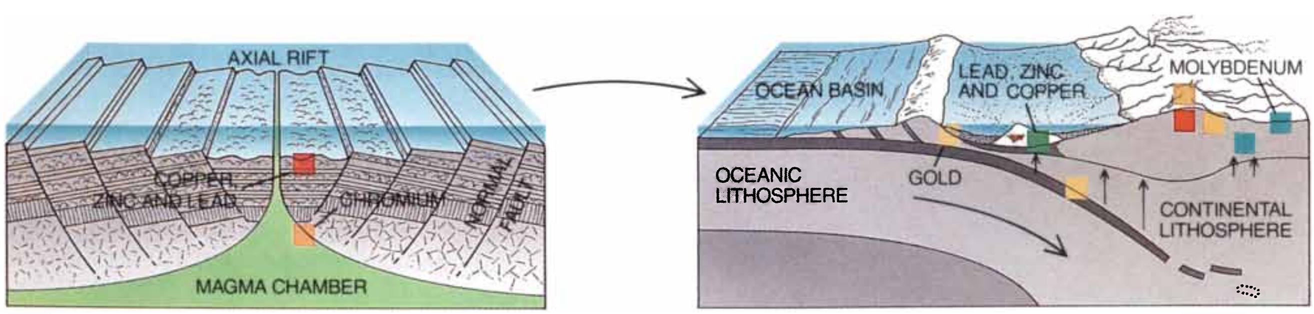
The ore-forming environments of the Phanerozoic are diverse. They include a wide variety of fluids: intrusive and extrusive silicate magmas, seawater and meteoric water. Consequently, many important metals were deposited in several types of ore deposits.

Particularly exciting observations have been made on the seafloor. Ophiolites, which are interpreted to be slices of ancient seafloor, often contain massive copper-rich sulfide deposits. After lateral transport on the asthenosphere, these ores are emplaced at continental margins. Since the development of plate tectonic theory, scientists have thought that these ores formed at seafloor-spreading centers, where seawater leached metals and reprecipitated them as sulfides. Indeed, deep-diving submarines, such as *Alvin*, have allowed scientists to observe active hydrothermal vent areas. Speculations were confirmed when researchers discovered deposits of metal sulfides, called black smokers, on the seafloor.

Subduction of oceanic crust in mountain-building belts seems to cause the partial melting of both the oceanic crust and the overlying continental crust and mantle. Magmas ascend and erupt to form the Andean-type volcanoes of North and South America. Occasionally these same magmas, if rich in water to start with, release an aqueous fluid along their ascent path and then freeze. Magmatic salts and copper are fractionated into the escaping water, which shatters surrounding rocks through its violent expansion. Aqueous fluids circulate within this permeable network, driven by the heat given off by the cooling intrusive magma. Fluids and exposed rocks vigorously interact, causing copper sulfides to precipitate; in so doing, they gave rise to three fourths of the known reserves of copper.

A new type of gold deposit related to subduction was recognized in an environment that was originally mined for mercury. (Mercury's affinity for gold is well known, as it is commonly used in amalgamation processes.) This different type of deposit, recognized by Don Gustafson, an independent consultant in Reno, Nev., occurs at the permeable fault contact of marine sediments scraped off the subducted Pacific Plate and overlying sandstones, where hot-spring activity occurred near volcanic centers.

Regional faults control other types of gold vein systems—as in the case of the Mother Lode of California—by guiding the circulation of large amounts of aqueous fluids through metamorphic wall rocks. The Mother Lode may actually mark a former subduction zone



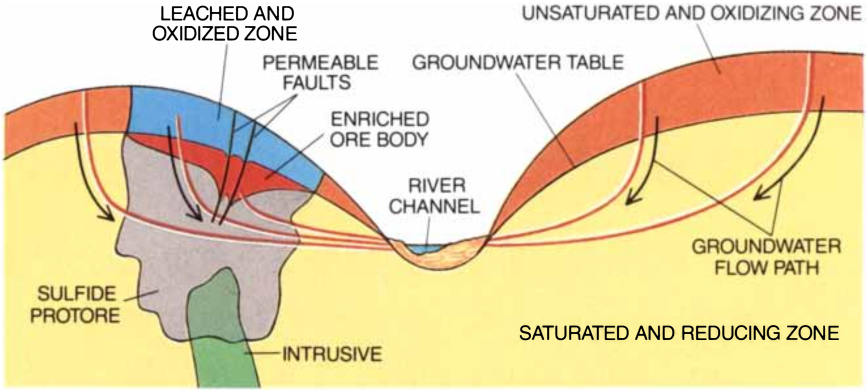
PHANEROZOIC crustal tectonics include seafloor-spreading centers as well as subduction and orogenic zones at the continental margins. Ore formation continues as magma reacts with seawater and intrudes into the continental plates.

along which seawater or metamorphic water circulated over great depths, according to John K.F.P. Bohlke, formerly at Argonne National Laboratory.

Gold eroded from primary ore deposits such as the Mother Lode accumulates as particles in streams, as a lucky few of the forty-niners found. Such secondary deposits are in many respects younger, smaller versions of the ancient gold and uranium conglomerates of the Witwatersrand. Many primary ore deposits formed in the Phanerozoic have been discovered by prospecting backward to the source of the later, secondary sites of surficial ore deposition.

In addition to deposits of gold, tin and other heavy minerals accumulated in riverbeds during the Phanerozoic by erosional transport from mountainous source areas. The buildup of other metals, such as aluminum, nickel or gold, took place in subdued, or flat, landscapes, where erosion was minimal. These subsurface deposits resulted from the chemical weathering of rocks under oxidizing conditions.

Bauxites and laterites—residual accumulations of aluminum and iron—occur in regions where tropical rainfall and high temperatures promoted intense chemical reactions between surface water and rocks. Leaching removed most elements but left behind a residue enriched in insoluble minerals containing aluminum, nickel or gold. We have shown that in addition to chemical weathering, transport and deposition of aluminous windblown dust can contribute to the enrichment of bauxite deposits. The wind itself then becomes an ore-forming fluid, especially along dust trajectories, where soils act as dust collectors.



WEATHERING IS PIVOTAL to ore concentration. Above the groundwater table, rock pores filled with air oxidize copper, causing the ore to leach out. As this copper reaches the water table—where there is less oxidation—it reprecipitates. Such so-called secondary enrichment makes copper ore minable.

The combination of air and water is particularly effective in causing local transport of metal underground. Today's oxygen-rich atmosphere is pivotal to this regional transport, called secondary enrichment of ores, which is so important to successful mining. Although many elements are leached away by groundwater as noted above, some, such as copper, are reprecipitated farther down.

Rocks above the water table have pores and fractures full of air. In the presence of oxygen, pyrite—the most common sulfide mineral—and most sulfides oxidize to form sulfuric acid and mobile metal ions. As acid water carrying these ions migrates downward, they meet the groundwater table. Here reduction replaces oxidation, and copper ions replace pyrite. This process enriches the ore below the groundwater table, sometimes by a factor of two or more.

Many primary ore deposits, exposed by uplift and erosion, underwent weathering and oxidation by descending groundwaters. The metals were leached from near the surface and reprecipitated at or near a local water table, thereby completing the sequence of processes leading to ore deposition. Without this secondary enrichment, most copper deposits of the Western Hemisphere would not be minable.

Weathering environments are perhaps the best-understood ore-forming systems because they involve the downward migration of fluids [*see illustration above*]. Water leaves the leached, oxidized protore relic exposed at the surface. Miners can then cut into the underlying zone of secondary enrichment, making these deposits very easy to study, unlike the primary part of the system. Such enrichment is unlikely before the oxygenation of the atmosphere, as indicated by the presence of pyrite in the Witwatersrand.

Since oxidative enrichment occurs near the earth's surface and is related to hydrologic factors like the position of the groundwater table, it is ultimately controlled by climatic changes influencing rainfall. Optimal conditions for secondary copper enrichment are attained during the transition from a wet to a dry climate because the lowering groundwater table exposes more primary sulfides to oxidative weathering. Charles Alpers, now at McGill University in Montreal, and I have shown that intense supergene, or downward, enrichment of copper deposits in the Atacama Desert of Chile occurred during such a major climatic transition 15 million years ago.

Furthermore, while climatic transitions provide the prerequisites for enrichment, they also support a feedback mechanism necessary for preservation of the enriched ores. As the rainfall rate dropped to its present hyperarid state, the erosion rate decreased as well, protecting the deposits from further erosion.

From their formation to their modification at the surface of the earth, ore deposits are geologically transitory and reflect dynamic processes within the earth as well as atmospheric and climatic influences on hydrologic systems. As highly reactive supracrustal systems, they then serve as geochemical sensors providing a powerful record and set of tracer elements for deducing the history, transport paths and forces operative in the crust.

As the understanding of the complex transport of ancient metals to their sites of deposition improves, the delicate balances necessary for the preservation of ores become more apparent. Further study will reveal the quantitative rates of various surficial processes while metals continue to respond vigorously to changing environments.

FURTHER READING

ATLAS OF ECONOMIC MINERAL DEPOSITS. Colin J. Dixon. Cornell University Press, 1979.

MINERAL DEPOSITS AS GUIDES TO SUPRACRUSTAL EVOLUTION. R. W. Hutchinson in *Evolution of the Earth*. Edited by R. J. O'Connell and W. S. Fyfe. American Geophysical Union, 1981.

METAL DEPOSITS IN RELATION TO PLATE TECTONICS. Frederick J. Sawkins. Springer-Verlag, 1984.

PRELIMINARY FRACTIONATION PATTERNS OF ORE METALS THROUGH EARTH HISTORY. George H. Brimhall in *Chemical Geology*, Vol. 64, No. 12, pages 1–16; August 25, 1987.

The Genetics of Thoroughbred Horses

Thoroughbreds are the most prized of domestic animals yet the least touched by modern genetics. New studies explore how inheritance affects their fertility and track performance

by Patrick Cunningham

The breeding and racing of Thoroughbred horses was once exclusively the sport of kings, but today it is also very much the sport of the ordinary citizen. Thoroughbred racing and all that goes with it constitute a huge leisure industry employing many thousands of people. Worldwide, $1.5 billion in racing prizes was paid out in 1984, more than half of it in the U.S. In that same year the total of officially recorded bets amounted to $33 billion, which helps to explain why gambling on horse races is an immense and closely regulated business in most countries. For those involved in the breeding and ownership of racehorses, it is variously a business, a sport, an art form, a tax strategy and a financial speculation.

In part because those interests are so diverse, much less organized study has been made of the genetics of the Thoroughbred than of the genetics of other domesticated species. The small size of the average horse-breeding venture has also discouraged such research. The breeding of pigs and poultry is primarily in the hands of companies that have the resources and economic incentives to harness modern population genetic theory to their breeding programs. Cattle breeding, too, is dominated by artificial insemination businesses, which implement large, scientifically planned breeding programs. No such concentrations of breeding power exist in the Thoroughbred world.

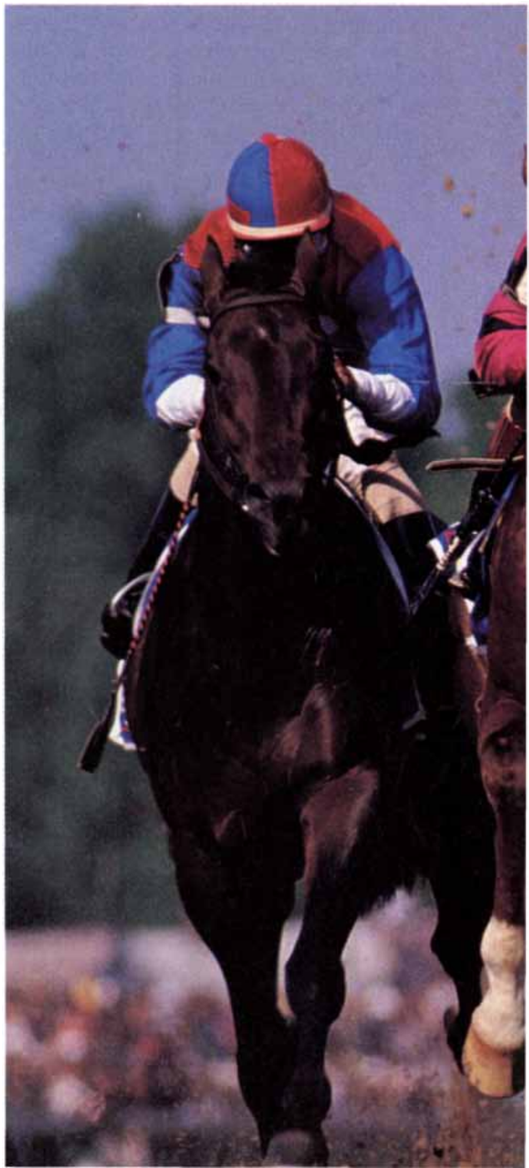
PATRICK CUNNINGHAM, professor of animal genetics at Trinity College in Dublin, has worked on genetic aspects of domestic livestock improvement programs for more than 20 years. He was head of animal breeding and genetics and deputy director of the Irish National Agricultural Research Institute until 1988. Today he works in Rome as the director of the Animal Production and Health Division of the Food and Agriculture Organization (FAO) of the United Nations.

The Thoroughbred, however, is in many ways an ideal animal to which to apply genetic theory. It is the best documented of all domestic animals, with meticulous pedigree records going back more than 20 generations. It has a single breeding objective—success on the track—and the performance of individual horses for this character is carefully and extensively documented.

Genetic studies of Thoroughbreds are particularly timely because of two problems facing the industry. First, Thoroughbreds display disturbingly low fertility: on a global average, only slightly more than 50 foals are produced by every 100 Thoroughbred mares. In the British and Irish population the annual reproduction rate is better—about 67 percent—but even this figure looks very poor compared with those in other single-offspring species, such as cattle, for which a rate of 85 percent is considered normal.

The other problem is that contrary to what one would expect from the efforts to breed and train winners, the racing performance of Thoroughbreds is not uniformly improving. The three English classic races are the St. Leger, a 1.75-mile race open to horses of both sexes, and the Oaks and the Derby, 1.5-mile races open to fillies and colts, respectively. Winning times for those races improved from the 1840s up to about 1910 but since then have been relatively static.

A possible genetic explanation for both the infertility and the static performances could be that these problems



THOROUGHBRED HORSES have been selectively bred for their racing prowess for more than three centuries. Although the pedigrees and track records of all these animals have been documented meticulously, formal genetic studies of Thoroughbreds have been almost nonexistent until recently.

arise from the fairly small gene pool of the Thoroughbreds. High levels of inbreeding might limit fecundity. Selection might also have achieved some optimal combination of racing traits in the Thoroughbred population. Without genetic studies, however, it was impossible to say whether such speculations might be right or wrong.

My colleagues and I at the National Agricultural Research Institute of Ireland became involved in these problems about 16 years ago, when I was summoned to the office of the minister for agriculture, who was also a noted racehorse owner and breeder. He told me he was appalled at the lack of scientific method in the horse-breeding industry. Could we improve the situation? The minister eventually went on to become prime minister, and we began a series of investigations to establish a scientific basis for the raising of Thoroughbreds.

One outcome of our research has been some new perspectives on the genetic origins of racehorses. The modern Thoroughbred population descended from a handful of stallions imported from North Africa and the Middle East into England in the 1600s. There is less surviving documentation about the mares in the founding generations, but many of them were imported from the same sources.

For about a century, the population remained relatively small, and Thoroughbred racing was the sport of a very limited royal coterie. The Tudor and early Stuart kings maintained studs, although these were dispersed by Oliver Cromwell in 1649. After the Stuart restoration, the patronage of King Charles II gave renewed impetus to Thoroughbred breeding and racing. The sport continued to develop strongly throughout the 18th century, and it was then that the three oldest classic races were established: the St. Leger in 1776, the Oaks in 1779 and the Derby in 1780.

In 1791 James Weatherby established his famous *Stud Book,* or horse-breeding record. In the fifth edition, published in 1891, the editors note: "The following list of the earliest known mares...combined with the imported Eastern stallions [are] the source, with hardly any exception, of the whole of the so-called thorough-bred stock now existant, as may be seen by the direct descent from them of almost all the principal stallions." That list comprised 80 foundation animals. From this narrow base, the Thoroughbred populution has expanded and now numbers approximately half a million worldwide.

Since its creation, Weatherby's *Stud Book* has continued to document the parentage of all Thoroughbreds born in Britain and Ireland. Annually it reports on about 18,000 mares. We have analyzed those pedigree records with the primary objective of measuring the level of inbreeding. The same calculations enabled us to quantify the contribution of the earliest ancestors to the genetic makeup of today's population.

A listing of the most important of these foundation animals confirms the

The Pillars of the *Stud Book*

Three prominent stallions are often called "the pillars of the *Stud Book*" because they appear in the bloodlines of an astonishingly large proportion of all modern Thoroughbreds. Those horses are, from left to right, the Godolphin Arabian (born about 1725), the Darley Arabian (born about 1688) and the Byerley Turk (born about 1690). The cat pictured with the Godolphin Arabian was the horse's inseparable companion, and all illustrations of the stallion show them together.

A fourth horse, the Curwen Bay Barb (*not pictured*), which was born about 1699, should be added to this select list of stallions because its genetic legacy to modern horses is even greater than that of the Byerley Turk. The significance of this horse is often overlooked, but one of its grandsons was highly prolific and passed along many of its genes.

These paintings of the horses are by Julie A. Wear. The Godolphin Arabian picture hangs in the Kentucky Derby Museum in Louisville; the other two are privately held. All three are based on earlier illustrations by the noted 18th-century painter John Wootton and on written records.

prominence of the three "pillars of the *Stud Book*": the Godolphin Arabian, the Darley Arabian and the Byerley Turk. Those three stallions appear in the pedigrees of a remarkably large proportion of the Thoroughbred population. It seems, however, that a fourth stallion—the Curwen Bay Barb—should have a place with this famous trio because his genetic contribution is slightly higher than that of the Byerley Turk. He is not usually recognized as a prominent stallion, because most of his contribution to the breed came through only one prolific descendant, his grandson Partner, which was foaled in 1718.

These four top stallions donated about one third of the genes in the current population, and the top 10 contributors are responsible for half of the genetic makeup. If we include all ancestors with a contribution of 1 percent or greater, the list includes 21 more horses (11 of them mares) and accounts for 80 percent of the makeup of the modern population. Those figures hinted that inbreeding, which seriously harms the fertility of most animal species, might be a problem for Thoroughbreds.

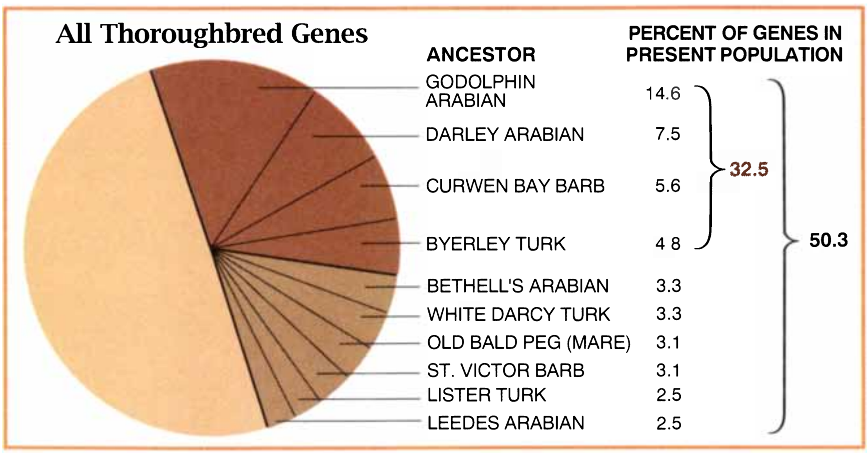
To begin a systematic search for the causes of the low rate of reproduction, my colleagues and I needed to determine the success rate per covering (mating attempt). We also had to learn how the age of mares, the timing of conception, levels of inbreeding and other factors affect that rate. We therefore analyzed the detailed breeding records for 2,466 coverings of 639 mares bred at a leading Irish stud farm between 1964 and 1976.

The results indicated that only about 40 percent of the coverings overall produced a live foal. Two percent ended in late-stage miscarriage, 26 percent ended in early embryonic loss and 32 percent were infertile matings. Distinguishing between an early loss and an infertile mating is difficult; our studies led us to adopt a rule of thumb that an estrus interval of less than 30 days signaled that an embryo had not been conceived.

The first coverings of the mares were generally more successful than subsequent ones: they produced live foals 41 percent of the time, compared with only 39 percent for second coverings and 37 percent for third coverings. Moreover, first services that took place after May 15 had a 44 percent foaling rate, which was much better than the 39 percent rate of first coverings carried out earlier in the season.

After foaling, many mares enter estrus again after an interval of only about 10 days. This estrus is known as the foaling heat. Breeders often mate a mare during this period in an attempt to advance her foaling date in the following season. About half of the first coverings in our data took place during the foaling heat, but they produced foals only about 29 percent of the time. That result is considerably less successful than the 49 percent foaling rate for first coverings during the next estrus cycle.

Observers have often noticed that in many mammal species a mother nursing its young is less likely to conceive than females that are not lactating. Nevertheless, lactation did not appear to affect the fertility of the Thoroughbreds in our data. There was no difference in the success rates of 829 mares



JUST TEN HORSES have contributed more than half of the genes found in the modern generation of Thoroughbreds. The top four horses, including "the pillars of the *Stud Book*," have collectively donated almost one third of the genes.

with foal at foot and of 489 mares that had not foaled in the previous year.

The age of a mare might also be expected to affect her success in producing foals. Age, however, had no noticeable effect on fertility until after a mare was 13 years old. Thereafter, a mare's fertility became considerably poorer.

We next checked the effects of inbreeding. Garry A. T. Mahon of Trinity College in Dublin and I considered all 10,569 mares in volume 35 of the *Stud Book*, which covers the period from 1961 to 1964. We followed the breeding history of each mare back through earlier volumes and forward through subsequent volumes to collect information on almost 100,000 mare-years of reproduction. For each mare, we calculated an index of her lifetime reproductive success. In parallel with this analysis, we calculated an inbreeding coefficient for each mare. The coefficient signifies the expected percentage of the genes in an offspring that would be identical because of relatedness in the parents.

Because of the scale of the computing task involved, we limited these inbreeding calculations to consideration of the previous five generations. For a smaller number of mares, we pursued the inbreeding analysis right back to the foundation animals of more than 20 generations ago.

Our analysis showed that over five generations, negligible inbreeding—only about 1 percent—had occurred. Only about two out of 100 mares had inbreeding coefficients higher than 4 percent. When we compared the foaling success rates with the levels of inbreeding, we found that fertility declined by about 7 percent with every 10 percent rise in the inbreeding coefficient—a reduction that was not statistically significant because of the small range of inbreeding levels.

It is therefore difficult to establish a strong statistical relation between the levels of Thoroughbred inbreeding and infertility. The literature on the effects of inbreeding on equine fertility is not extensive and offers variable results. A recent study in Norwegian Trotters showed a 4.3 percent reduction in the foaling rate per 10 percent rise in inbreeding of the offspring. The average level of inbreeding in the population that they studied, however, was 5.7 percent, which is much higher than we found among Thoroughbreds.

When we traced the pedigrees of some individuals back to the foundation animals, we arrived at an average inbreeding coefficient of 12.5 percent. Most of that inbreeding is attributable to the concentration of genes from a few prominent stallions in the early 1700s. Again, there was no evidence that inbreeding was having any significant effect on current fertility. During the 22 generations over which those horses were bred, natural selection, reinforced by breeders deliberately eliminating less fertile mares, could have counteracted any deleterious effects. Slow, gradual inbreeding, then, may be less damaging than fast inbreeding.

In our studies, Mahon and I believed it was important to estimate how much of Thoroughbred fertility was strictly under genetic control in any case. We calculated the degree to which fertility is inherited by looking at how much more similar for this trait the daughters of any one sire were to one another than they were to randomly selected members of the population. The heritability estimate for mare fertility was only 7.7 percent, which suggests that genetic differences in fertility are responsible for only a small part of the total variation in this trait. This conclusion agrees with many studies in other species and indicates that most variation in fertility is caused by environmental factors.

Further calculations based on this heritability estimate confirmed our belief that any deleterious effects of inbreeding during the 19th century could have been dissipated by selection for fertility. Breeders would need to reject only 8 percent of mares for infertility to counteract the inbreeding depression. Our conclusion is that the current level of inbreeding is not a problem and that the harm from long-term inbreeding has probably been offset by selection.

If inbreeding is not contributing to Thoroughbreds' infertility, then perhaps some aspect of seasonal breeding is at work. Like most nontropical animals, the horse is a seasonal breeder. In the Northern Hemisphere, most mares reach the peak of their natural fertility in May, June and July and then cease ovulating in January, February and March. In the Southern Hemisphere the peaks and lows are shifted by six months.

That pattern of changing seasonal fertility has been studied separately in Australia by Virginia E. Osborne of the University of Sydney and in Ireland by

Joe Jennings of the Agricultural Research Institute. They have independently confirmed that in the winter fewer than 20 percent of mares ovulate and that by midsummer 90 percent do. It should therefore be much easier to breed mares successfully during certain periods of the year.

That fact may seem obvious, but it runs contrary to routine breeding practice. In Britain, Ireland and the U.S. all horses born in one calendar year become officially one year old on the next January 1. In Australia this registration date is August 1. Because breeders want their horses to have the advantages of relative maturity and strength when competing with others of the same official age, breeders try to have their foals born early. Pregnancy in a mare lasts 11 months, so the breeding season runs from February 15 to July 15 in the Northern Hemisphere and from September 8 to December 31 in the Southern Hemisphere.

When these official breeding season dates are superimposed on the periods of ovarian activity, it is immediately apparent that the breeding season includes some of the mares' least fertile months and excludes some of the most fertile. The practice of starting the year's registrations on January 1 in the Northern Hemisphere therefore costs the industry something in fertility.

The magnitude of the loss is difficult to quantify because most mares enter the breeding season in foal (pregnant), which alters their subsequent ovulation pattern. Close to 40 percent of the mares, however, are maidens or were not in foal in the previous season. For this group, we calculated the net effect on fertility of pushing the five-month breeding season back from its present starting date of February 15.

We discovered that in the early part of the season, each week's delay adds about 1.5 percent in net fertility to the mares. Changing the registration date from January 1 to March 1 and the start of the breeding season from February 15 to April 15 should add about 14 percent to the fertility of the studied group and perhaps 10 percent to the mare population as a whole. Similar results apply to changes in the Australian dates. The potential benefits of changing the dates are significant enough to merit consideration by the horse-breeding community.

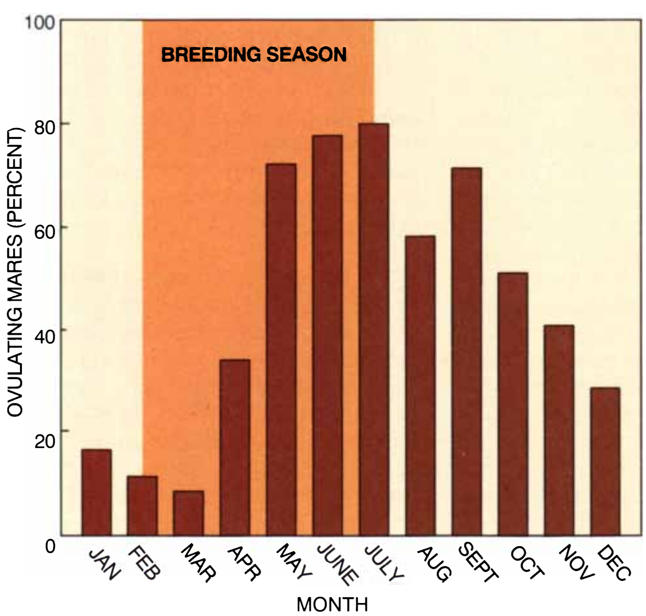
Ironically, we have found that the reason breeders want their foals born early in the year—to increase their value—may sometimes be less valid than most breeders believe. As a side study, John Ruane of Trinity and I did a small analysis of the effect of birth date of foals on their subsequent sale value and on their performance at two and three years.

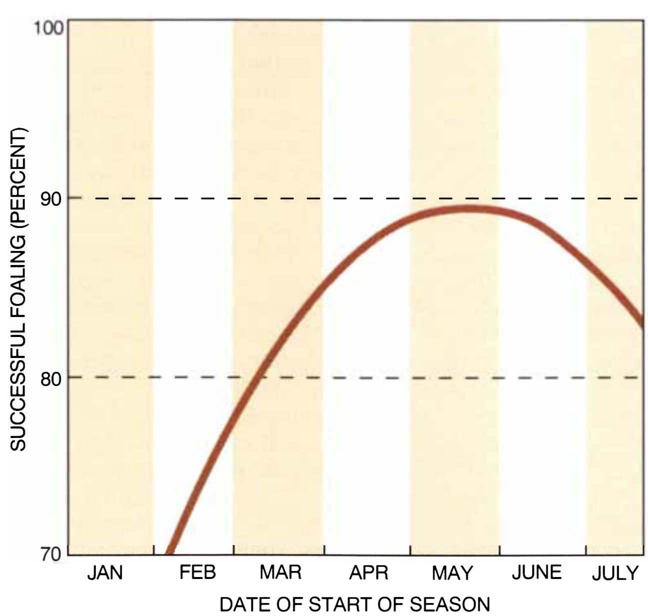
For the analysis of sale value, we considered 553 yearlings sold in October 1985 and calculated the average sale price commanded by the foals born in each month. Horses for sale are classified into three groups: Invitation, with an average sale price of $220,000; Premier, with an average of $36,000; and Open, with an average of $8,000. The month of birth, we found, had no effect on the sale prices for fillies nor for any colts except those in the Open category—that is, at the less expensive end of the market.

Our rationalization of this pattern is that for the more expensive and highly bred animals, pedigree and appearance together determine the sale price irrespective of birth date. At the low-priced end of the market, many fillies are bought for breeding, and the month of their birth is unimportant because they do not race. Pedigree distinction counts for little among the cheap colts, however: each animal is evaluated on its individual traits, not on its bloodline. Because age is considered to be an advantage, colts born late in the year are discounted.

Even some of those late colts may be undervalued, according to our performance analysis. To assess the racing ability of two- and three-year-old Thoroughbreds born during various months, we used their Timeform ratings, which we consider to be the best available quantification of performance for horses in Britain and Ireland. The rating is expressed in pounds and represents the handicap weight that the horse would be given in an open or free race. Timeform was established in 1948 by the late Phil Bull, a remarkable English mathematician turned punter (or gambler, for American readers). The service produces annual ratings of almost all the horses that compete.

We looked at the end-of-year Timeform ratings for all horses that were two-year-olds in 1981 and three-year-olds in 1982 and classified the animals





MARE FERTILITY fluctuates seasonally, as shown in data gathered in Ireland (*left*). The number of ovulating mares peaks in the summer and reaches a nadir in the winter and early spring. Currently the breeding season falls during some of the least fertile months. If the breeding season were postponed, the foaling rate could rise dramatically (*right*).

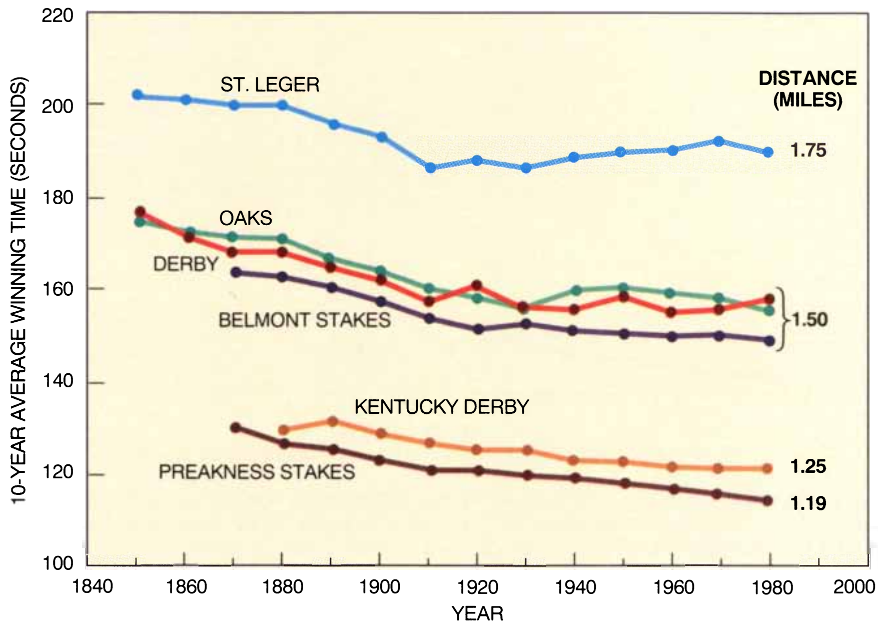
by their month of birth. No decline in the ratings was apparent for horses born in the first five months of the year, but there was some reduction for the few born in late May and June. Taken together, our results suggest that the influence of birth date on performance is commonly overrated.

The entire selection and breeding process in Thoroughbreds is founded on the belief that racing performance is inherited. Attempts to analyze the genetics of performance in a systematic way have involved some distinguished names, including Charles Darwin and Francis Galton. It is only in recent decades, however, that good estimates of the heritability of performance based on adequate data have been produced.

Toward that end, my colleagues and I have made several major analyses of Timeform data, the most recent of which included the end-of-year records for 31,263 three-year-olds that raced between 1961 and 1985. We have attempted to measure whether groups of half brothers or half sisters have ratings that are more alike than those of randomly assembled groups. Similarly, we have also looked at the extent to which the ratings of parents and their offspring resemble one another more than do those of random pairs of individuals selected from two generations.

Our best estimate says that track performance, as measured by the Timeform rating, is about 35 percent heritable. In other words, about 35 percent of all the variation that we observe in track performance is controlled by heritable factors and the remaining 65 percent by other influences, such as training and nutrition. If a mare and a stallion are each rated 10 percent higher than the average for the population, then we can expect that their offspring will have ratings that are on average about 3.5 percent higher. Bear in mind, however, that there is not a straightforward correlation between a horse's handicap rating and its actual speed.

With that performance heritability figure in mind, Barry Gaffney of Trinity and I sought to estimate how much the performance of Thoroughbreds should be improving over time, based on the idea that the horses with the best track records are favored for breeding. The average generation length in Thoroughbreds is about 11 years. Approximately 6 percent of colts and about 53 percent of fillies are selected for breeding. Putting this information together with the estimated heritability of performance, we calculated that, on average, genetic improvements in Thoroughbreds should raise the mean Timeform ratings by 0.92 unit each year.



WINNING TIMES are not steadily improving in all Thoroughbred races. In the long British classic races (the St. Leger, the Oaks and the Derby) and in the Belmont race in the U.S., the average winning times are not decreasing much, if at all. In contrast, the winning times for the shorter Kentucky and Preakness races are still dropping.

We then tried to verify that genetic changes in the Thoroughbred population were taking place at the predicted rate. Working with 11,328 Timeform ratings for three-year-old Thoroughbreds, we estimated the relative genetic merit of the stallions born in the years 1952 to 1977. Our analysis showed that although the average genetic value varied somewhat from year to year (as one would expect), it had a steady upward trend that averaged 0.94 Timeform unit per year. This figure was remarkably close to our prediction and confirmed our belief that selection is steadily improving the average racing performance in the population.

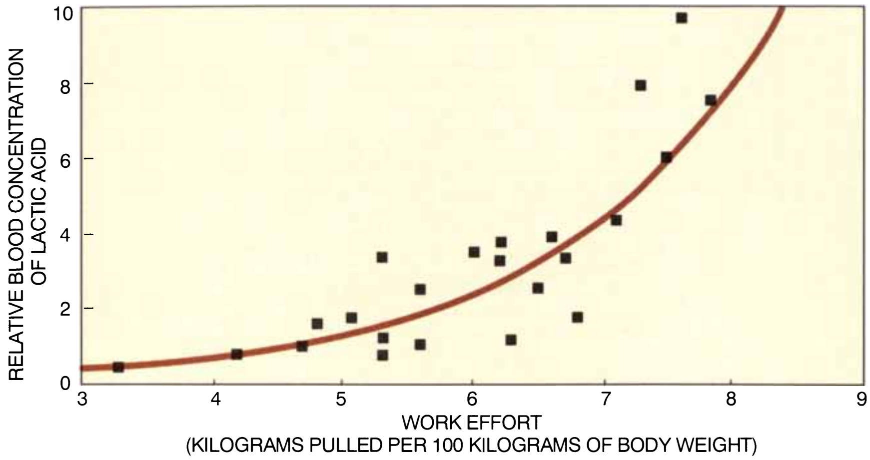
Yet, contrary to our conclusion, the irrefutable fact remains that modern Thoroughbreds are not bettering the times of their forebears in the English classic races. If steady genetic improvement is occurring, how can we explain static winning times?

One possibility is that a physiological ceiling to performance is preventing further improvement in the Thoroughbreds. There is some support for this idea: although the performance plateau is evident in the St. Leger and other long races, winning times are continuing to improve in the shorter, less demanding races common in the U.S., such as the Kentucky Derby.

Richard L. Willham and his colleagues at Iowa State University have found evidence that selective breeding is still boosting the performance of quarter horses. Quarter-horse races are run over relatively short distances, such as 320 meters (about 0.2 mile). After analyzing more than a million finishing time records, Willham's group estimated the genetic trends in the quarter-horse population between 1960 and 1983. They also found consistent average yearly improvements of 0.47, 0.43 and 0.16 percent in the finishing times for the 320-, 366- and 402-meter races, respectively. The genetic trend could have accounted for one third of the total improvement in the finishing times.

The discrepancy between the improvements in the quarter-horse races and the classic races would make sense if a physiological limit on performance came into play beyond a certain threshold of effort. If that hypothesis is true, then it should be possible to pinpoint the physiological elements that restrict performance.

The two most likely limiting factors both relate to blood circulation and muscle metabolism. The first one is the blood-borne supply of oxygen to the muscles for the regeneration of adenosine triphosphate (ATP), the molecular fuel that provides energy for muscle contraction. The second factor is the rate at which lactic acid is cleared from the muscles. Lactic acid is a waste product of the anaerobic (oxygen-free) breakdown of glycogen, an energy-rich carbohydrate that provides a second source of ATP. If lactic acid accumu-



LACTIC ACID, a waste product of anaerobic muscle contractions, accumulates faster as a horse exerts itself more. The inability to eliminate lactic acid during the long classic races may explain the plateau in their winning times.

lates, a muscle cannot derive as much energy anaerobically.

In every race, a running horse relies on both aerobic and anaerobic muscle contractions. Although the flow of blood to the muscles increases dramatically during exercise, a horse's maximum effort cannot be sustained by aerobic contraction alone. Similarly, the depletion of glycogen reserves in the muscles and liver limits activity based on anaerobic contraction. The balance of the aerobic and anaerobic contributions to muscular activity depends on many variables but particularly on the amount of activity involved—that is, the length of the race.

The 20-second sprinting races of American quarter horses are mainly aerobic. The 2.5 minutes required to run the English classics, conversely, shift the balance, so that most of the energy supply is anaerobic. That difference may partially explain why finishing times in longer races seem to have reached a plateau, whereas records continue to be broken in the sprints.

The physiology of blood circulation and exercise in horses has been studied by G. Frederick Fregin and D. Paul Thomas, then at the University of Pennsylvania. They found that as horses stepped up their efforts during exercise, their heart rate rose steadily from a resting figure of about 40 beats per minute to a maximal rate of close to 200. At the same time, the volume of blood pumped by the heart with each beat increased linearly. Running at full stretch, a 500-kilogram horse pumps 250 liters of blood through its system each minute, which is the equivalent of 10 times its total blood volume. Because blood circulation and oxygen delivery can rise linearly with exercise, a lack of oxygen is not likely to limit a horse's performance.

A quite different picture, however, is presented by the levels of lactic acid in the blood over a similar range of effort. Fregin and Thomas showed that the elimination of lactic acid does not keep pace with growing exertion: blood lactic acid levels rise much faster. Indeed, at a horse's maximum effort, the clearance of lactic acid from the body lags critically. That observation suggests that the limiting factor in racing performance may well be the accumulation of lactic acid in the muscles.

Even if one assumes that such a physiological ceiling exists, further progress should be possible. In human athletics, training regimens are now often targeted specifically on shifting the lactic acid clearance rate. Such methods could well have an application in the training of horses. Furthermore, it might well be possible to select deliberately for more efficient lactic acid metabolism. In horses, that project would be very long term, although it might be feasible to test its applicability in a selection experiment in mice.

Human society has domesticated a dozen or so of the 4,000 mammalian species now living. For many centuries, the horse was perhaps the most important of them, as the dominant work and transport animal outside the tropics. Those roles have now nearly disappeared. Nevertheless, the number of horses continues to grow in many developed countries as their uses in sports expand. In this context, the breeding and racing of Thoroughbreds could be regarded as the principal human involvement with horses today.

Yet compared with the other domesticated species, horses and Thoroughbreds in particular are practically untouched by science. An exception is in veterinary medicine, because the owners of expensive horses demand the scientifically based prevention, diagnosis and treatment of equine illnesses. In other respects, the horse industry has little incentive and some disincentive to adopt scientific methods.

The rates of genetic improvement in cattle and pig populations have been accelerated greatly by artificial insemination and embryo-transfer techniques, which permit valuable males and females to contribute more widely to the next generation. The authorities that regulate Thoroughbred horse breeding nonetheless forbid the use of these methods.

Nobody is much interested in improving the average racing times because advantages accrue only to the owners of the fastest horses; what does it matter if all horses race 10 percent faster? The introduction of these techniques would also radically alter the lucrative market for breeding animals: many fewer stallions would be required, for instance. Such an economic disruption would not be welcomed.

Nevertheless, scientific methods can help the breeding of Thoroughbreds. The industry would undoubtedly benefit from the more efficient procedures, already in use with other species for evaluating breeding animals. Raising mare fertility would also benefit everyone. That goal is the easiest one to achieve: all it requires is a change in the breeding regulations. New DNA techniques can simplify the challenges of verifying a Thoroughbred's parentage, and they may eventually improve the selection of animals for speed and other useful traits. All these beneficial changes are possible without destroying the basic biological uncertainties that go to make a horse race.

FURTHER READING

AN INTRODUCTION TO THE THOROUGHBRED. Peter Willett. London, Stanley Paul, 1977.

CARDIOVASCULAR RESPONSE TO EXERCISE IN THE HORSE: A REVIEW. G. F. Fregin and D. P. Thomas in *Equine Exercise Physiology.* Edited by D. H. Snow, S.G.B. Persson and R. J. Rose. Cambridge, Burlington Press, 1983.

A REVIEW OF THE INHERITANCE OF RACING PERFORMANCE IN HORSES. E. A. Tolley, D. R. Notter and T. J. Marlowe in *Animal Breeding Abstracts,* Vol. 53, No. 3, pages 163–185; March 1985.

ESTIMATION OF GENETIC TREND IN RACING PERFORMANCE OF THOROUGHBRED HORSES. B. Gaffney and E. P. Cunningham in *Nature,* Vol. 322, No. 6166, pages 722–724; April 21, 1988.

Mechanical Engineering in the Medieval Near East

Muslim engineers preserved earlier technology and added inventions as basic as the crank, the windmill and the suction pump. Some of their innovations anticipated much later developments in Europe

by Donald R. Hill

The West is accustomed to seeing its own intellectual development as having been shaped, in the main, by internal factors. This view of history traces our heritage back from the Industrial Revolution to the Enlightenment and Renaissance and, thence, via the monkish scribes of the Middle Ages, to the fountainhead: Greece, Rome and the ancient empires of the Fertile Crescent.

But the picture is incomplete because it ignores the intermediation of the civilizations of Greek Christendom (or Byzantium), Hindu India, Confucian China and Islam. Our subject here is the technology of medieval Islam—the knowledge it preserved, the new ideas it contributed to the medieval world and the inventions by which it anticipated later developments.

When the Prophet Muhammad died in A.D. 632, he left behind a new religion with its administrative center at Medina and its spiritual heart at Mecca. Within about a year of his death the rest of Arabia had joined the Muslim fold; by 750 the Arab Empire stretched from the Pyrenees to central Asia.

Although the advent of Islam brought immense political, religious and cultural changes, the technological traditions were largely unaffected. In mechanical engineering the Muslims adapted the techniques of earlier civilizations to satisfy the needs of the new society. These needs centered on a city life more extensive than any seen since Roman times.

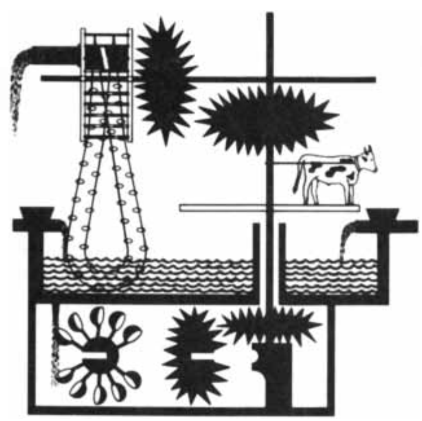
Baghdad's population is estimated to have reached about 1.5 million in the 10th century, and cities such as Cordoba, Cairo and Samarkand, although smaller, were still of considerable magnitude. Paris, by contrast, would not number 100,000 souls for another 400 years. Feeding and clothing the inhabitants of the Islamic world's vast urban centers placed great demands on agriculture and distribution. These, in turn, depended on technology for supplying irrigation water to the fields and for processing the crops into foodstuffs.

Water and waterpower, therefore, will constitute our first concern. Then we shall describe water mills and the windmills. Finally, we shall turn to descriptions, most of them in a handful of treatises that have come down to us, of water clocks, fountains and various automata, some of which might seem trivial to modern eyes. Yet they exploit concepts, components and techniques that did not enter the armamentarium of European engineering until the time of the Renaissance.

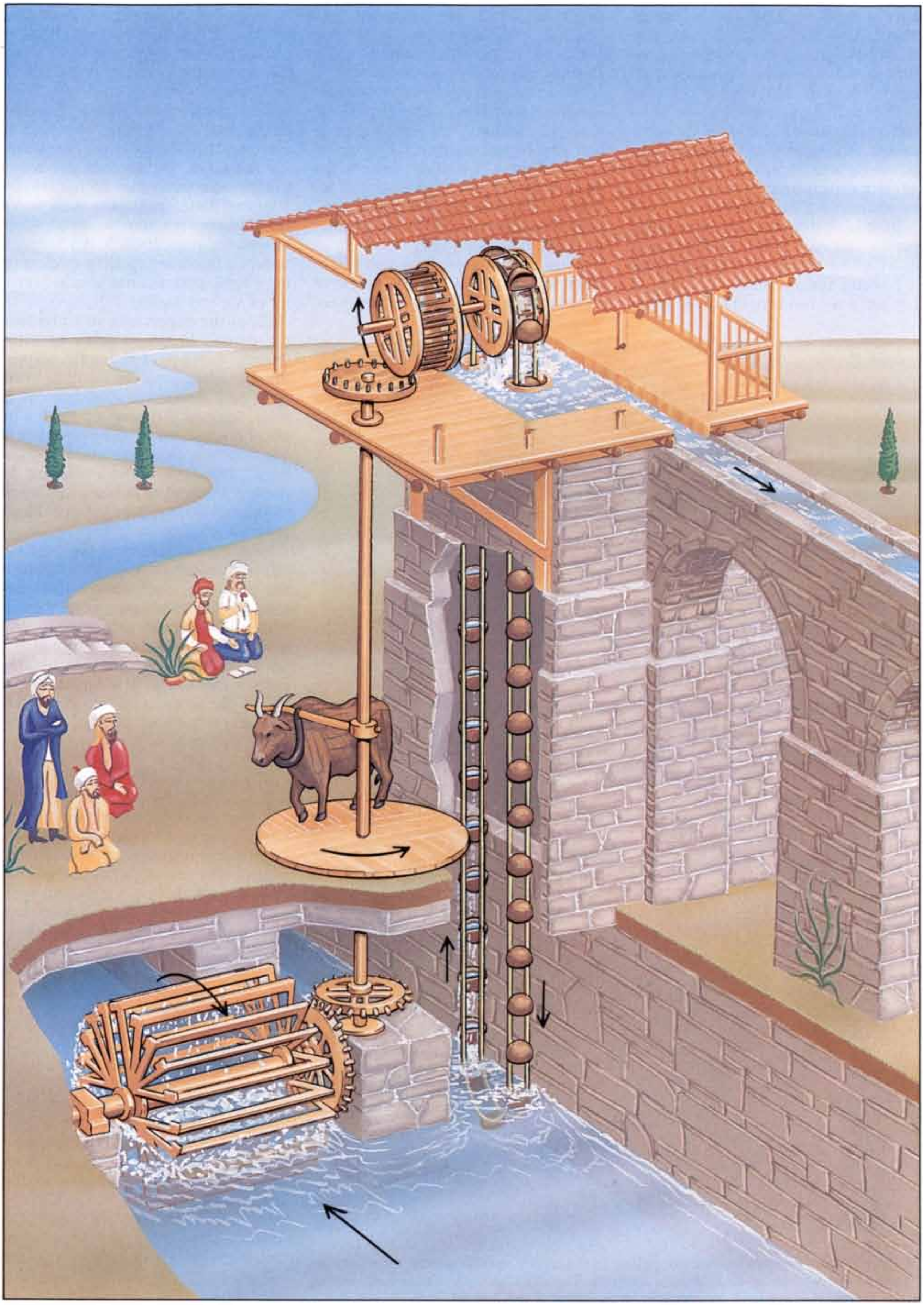
The most ancient water-raising machine is the shaduf, a counterweighted lever from which a bucket is suspended into a well or stream. It appears in illustrations from as early as 2500 B.C. in Akkadian reliefs and is still in use today in parts of the Middle East. Other traditional water-raising machines, introduced between the third and first centuries B.C., include the screw, or water snail, whose invention is attributed to the great mathematician Archimedes. It consists of a helical wooden blade rotating within a barrellike wooden cylinder, a design that could not push water up inclines greater than about 30 degrees, although 20 degrees was more common.

Higher lift was achieved by the noria, a large wheel driven by the velocity of the current. On the outer rim a series of compartments are fitted in between a series of paddles that dip into the water and provide the propulsive power. The water is scooped up by the compartments, or pots, and is discharged into a head tank or an aqueduct at the top of the wheel. Norias could be made quite large. The well-known wheels at Hama on the river Orontes in Syria

DONALD R. HILL, a retired engineer, became interested in Arabic while serving with Britain's Eighth Army in North Africa during World War II. After the war, he worked for the Iraq Petroleum Company, returning to England to join Imperial Chemical Industries. He later moved to senior positions in the subsidiaries of two U.S. petrochemical corporations, from which he retired in 1984. He now devotes his time to Arabic studies, in which he has earned a master's degree from Durham University and a Ph.D. from the University of London's School of Oriental and African Studies. His translation of al-Jazari's book of machines won for him a share of the 1974 Dexter Prize, awarded by the American Society for the History of Technology.



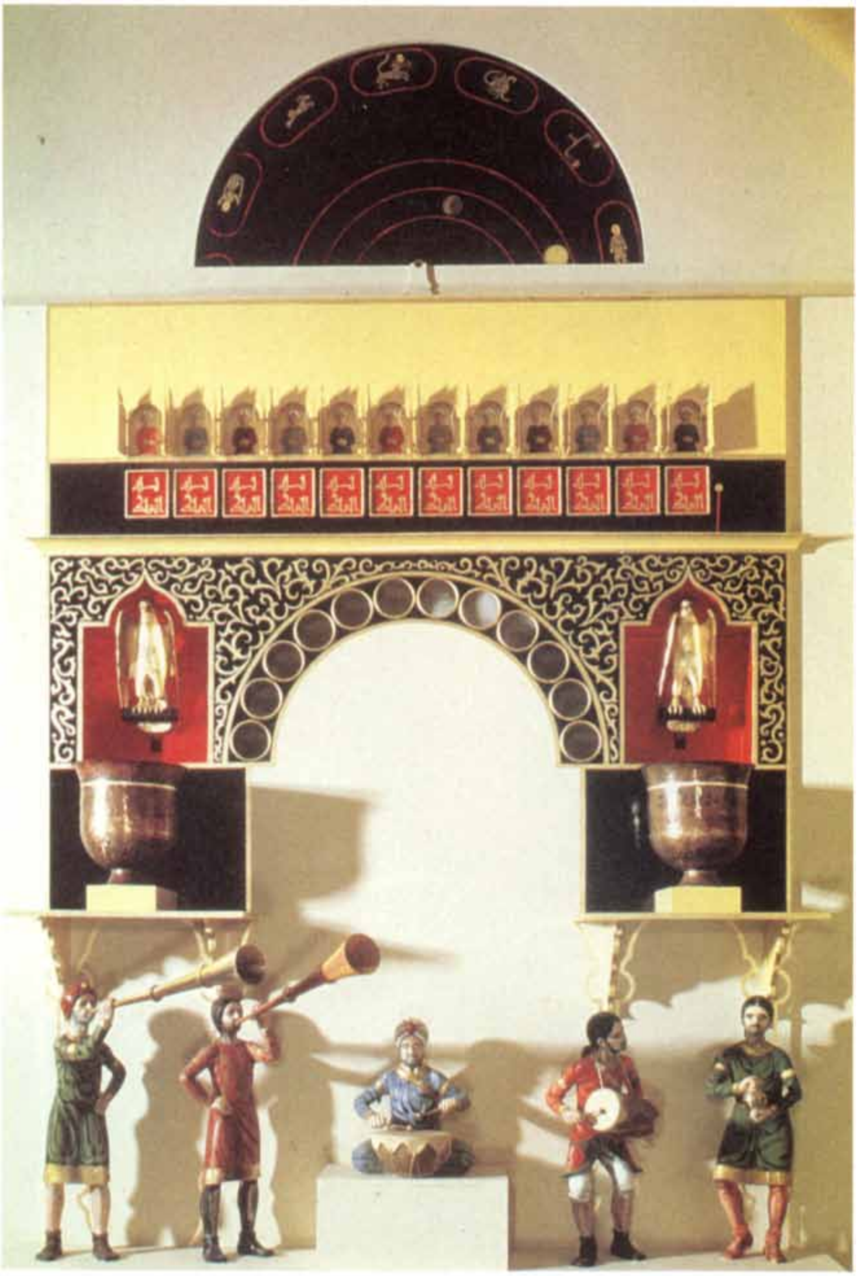
CHAIN OF POTS, or *saqiya*, raises water for cities and farms. A wooden ox seems to drive the *saqiya*, which is actually powered by a hidden waterwheel. The drawing (*right*) is based on one (*above*) given by al-Jazari, who flourished in Iraq at the end of the 12th century.



have a diameter of about 20 meters. The noria is self-acting, and its operation thus requires the presence of neither man nor beast. It is, however, expensive to build and maintain.

The *saqiya* is probably the most widespread and useful of all the water-raising machines that medieval Islam inherited and improved. It is a chain of pots driven by one or two animals by means of a pair of gears. The animals push a drawbar through a circle, turning an axle whose pinion meshes with a vertical gear. The gear carries a bearing for the chain of pots, or pot garland—two ropes between which earthenware pots are suspended. The chain of pots is optimal for raising comparatively small amounts of water from comparatively deep wells.

Other mechanisms, however, were required to raise large quantities of water relatively small distances. The problem can be solved by using a spiral scoop wheel, which raises water to the ground level with a high degree of efficiency [see "50 and 100 Years Ago," page 14]. The machine is very popular in Egypt nowadays, and engineers at a research laboratory near Cairo have been trying to improve the shape of the scoop in order to achieve the maximum output. Although it appears very modern in design, this is not the case: a 12th-century miniature from Baghdad shows a spiral scoop wheel driven by two oxen.



WATER CLOCK, reconstructed according to al-Jazari's specifications, incorporates "in-line" valves and other hydraulic controls. The clock measures time both by the hour and by the seasonal progression of the signs of the Zodiac.

These machines are still in use today in many oil-poor Middle Eastern countries, because for many purposes they are at least as efficient as diesel-driven pumps. Moreover, they do not require imported fuels, spare parts or labor. Vital time can therefore be saved, when the loss of even a single day's operation of a machine can kill a crop, making reliable performance literally a matter of life and death.

Given the importance of water-raising devices to the economy of many Islamic societies, it is hardly surprising that attempts were made to introduce new designs or modify existing ones. Some of the most interesting innovations are found in one section of Ibn al-Razzaz al-Jazari's great machine book, *The Book of Knowledge of Ingenious Mechanical Devices,* which was completed in Diyar Bakr in Upper Mesopotamia in 1206.

From our point of view, the most significant aspect of these machines is the ideas and components that they embody. For example, one of them is explicitly designed to eliminate out-of-balance loading and so produce a smoother operation. Another incorporates a crank, the first known example of the nonmanual use of this important component. Some of these devices functioned as curiosities.

The invention containing the most features of relevance for the development of mechanical design, however, was intended as a practical machine for high-lift duties: a twin-cylinder, water-driven pump [*see illustration on page 104*]. A stream turned a paddle wheel connected to a gear wheel meshing with a horizontal gear wheel, which was installed above a sump that drained into the stream. The horizontal wheel contained a slot into which a vertical pin fitted near the perimeter of the wheel.

The turning wheel moved two connecting rods back and forth, thus driving opposing pistons made of copper disks spaced about six centimeters apart, the gap being packed with hemp. The pistons entered copper cylinders, each one having a suction and delivery pipe. One piston began its suction stroke while the other began its delivery stroke. This machine is remarkable for three reasons: it incorporates an effective means of converting rotary into reciprocating motion, it makes use of the double-acting principle and it is the first pump known to have had true suction pipes.

Waterpower was clearly a prominent concern of medieval Islamic planners. Whenever they mentioned a stream or river, for example, they often included an estimate of how many mills it would operate. One might say that they assessed streams for "mill power."

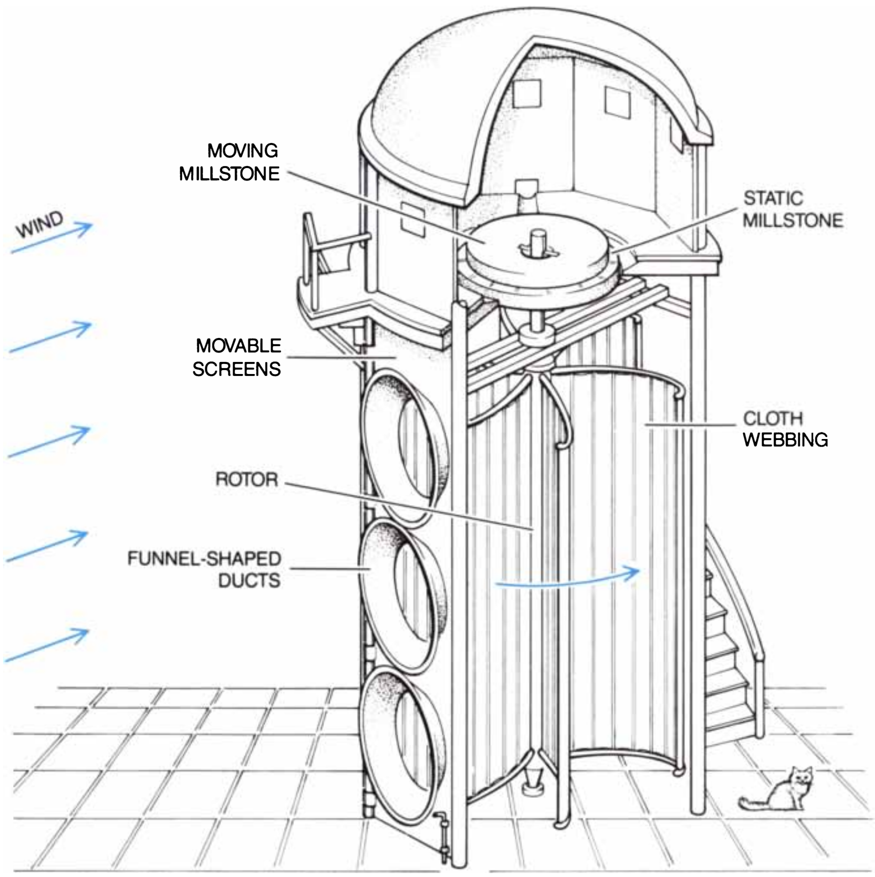
The three main types of waterwheel had all been in existence since Classical times—the horizontal wheel and two variations of the vertical wheel. The horizontal wheel has vanes protruding from a wooden rotor, onto which a jet of water is directed. In modern Europe the design was altered to use water moving axially, like air flowing through a pinwheel, creating the water turbine. Interestingly, wheels with curved blades onto which the flow was directed axially are described in an Arabic treatise of the ninth century.

The more powerful vertical wheels came in two designs: undershot and overshot. The former is a paddle wheel that turns under the impulse of the current. The overshot wheel receives water from above, often from specially constructed channels; it thus adds the impetus of gravity to that of the current.

When the levels of rivers fall in the dry season, and their flow diminishes, undershot wheels lose some of their power. Indeed, if they are fixed to the banks of rivers, their paddles may cease to be immersed. One way this problem was avoided was by mounting the waterwheels on the piers of bridges and taking advantage of the increased flow there. Another common solution was provided by the shipmill, powered by undershot wheels mounted on the sides of ships moored in midstream. On the rivers Tigris and Euphrates in the 10th century, in Upper Mesopotamia, which was the granary for Baghdad, enormous shipmills made of teak and iron could produce 10 tons of flour from corn in every 24-hour period.

Gristmilling—the grinding of corn and other seeds to produce meal—was always the most important function of mills. Mills were, however, put to many other industrial uses. Among these applications were the fulling of cloth, the crushing of metallic ores prior to the extraction process, rice husking, papermaking and the pulping of sugarcane. The usual method of adapting waterwheels for such purposes was to extend the axle and fit cams to it. The cams caused trip-hammers to be raised and then released to fall on the material.

Where waterpower was scarce, the Muslims had recourse to the wind. Indeed, it was in riverless Seistan, now in the western part of Afghanistan, that windmills were invented, probably early in the seventh century A.D. The mills were supported on substructures built for the purpose or on the towers of castles or the tops of hills. They consisted of an upper chamber for the millstones and a lower one for the rotor. A vertical axle carried either 12 or six rotor blades, each covered with a double skin of fabric. Funnel-shaped ducts pierced the walls of the lower chamber, their narrower ends facing toward the interior in order to increase the speed of the wind when it flowed against the sails.



FIRST WINDMILL was invented in the seventh century in Afghanistan, where waterpower was lacking. The rotor turned on a vertical axis, a design that spread throughout much of Asia. The Muslims never adopted the European windmill, with its horizontal axis, although the Crusaders erected such windmills in their castles.

This type of windmill spread throughout the Islamic world and thence to China and India. In medieval Egypt it was used in the sugarcane industry, but its main application was to gristmilling.

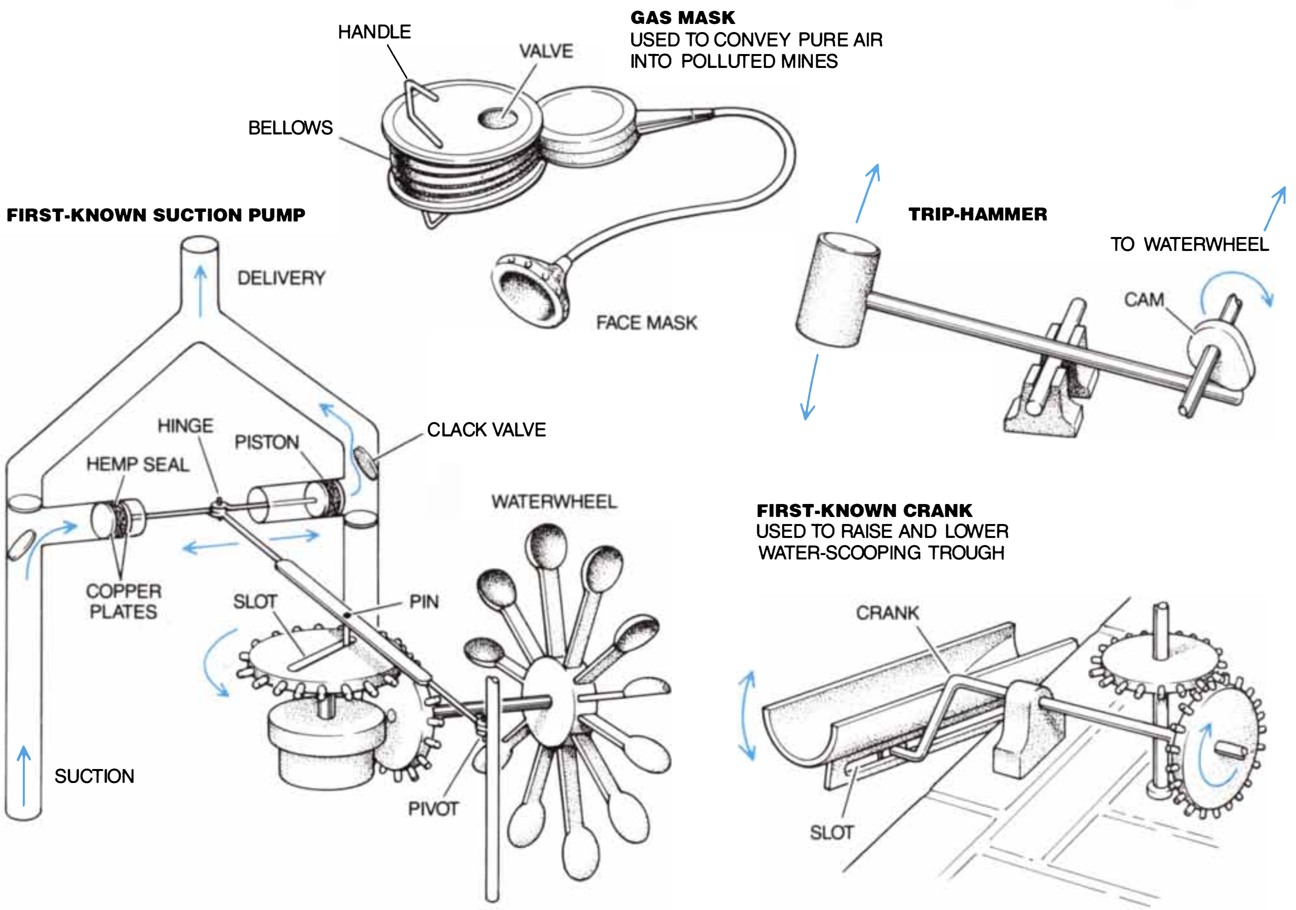
Now we turn to a type of engineering that is quite different from the utilitarian technology described so far. We may perhaps call it fine technology, since its distinguishing features derive from the use of delicate mechanisms and controls.

Some of these devices had obvious practical uses: water clocks were used in astronomical observations and were also erected in public places; astronomical instruments aided both observation and computation. Others gave amusement and aesthetic pleasure to the members of courtly circles. Still others undoubtedly had didactic purposes, for example, to demonstrate the principles of pneumatics as understood at the time. Apart from astronomical instruments and the remains of two large water clocks in Fez, Morocco, none of these machines has survived. Our knowledge of them comes almost entirely from two of the Arabic treatises that have come down to us.

The first is by the Banu (Arabic for sons of) Musa, three brothers who lived in Baghdad in the ninth century. They were patrons of scholars and translators as well as eminent scientists and engineers in their own right. They undertook public works and geodetic surveys and wrote a number of books on mathematical and scientific subjects, only three of which have survived.

The one that concerns us here is *The Book of Ingenious Devices*. It contains

Some Mechanical Devices of Medieval Islam



descriptions, each with an illustration, of 100 devices, some 80 of which are trick vessels of various kinds. There are also fountains that change shape at intervals, a "hurricane" lamp, self-trimming and self-feeding lamps, a gas mask for use in polluted wells and a grab for recovering objects from the beds of streams. This last is of exactly the same construction as a modern clamshell grab.

The trick vessels have a variety of different effects. For example, a single outlet pipe in a vessel might pour out first wine, then water and finally a mixture of the two. Although it cannot be claimed that the results are important, the means by which they were obtained are of great significance for the history of engineering. The Banu Musa were masters in the exploitation of small variations in aerostatic and hydrostatic pressures and in using conical valves as "in-line" components in flow systems, the first known use of conical valves as automatic controllers.

In several of these vessels, one can withdraw small quantities of liquid repeatedly, but if one withdraws a large quantity, no further extractions are possible. In modern terms one would call the method used to achieve this result a fail-safe system.

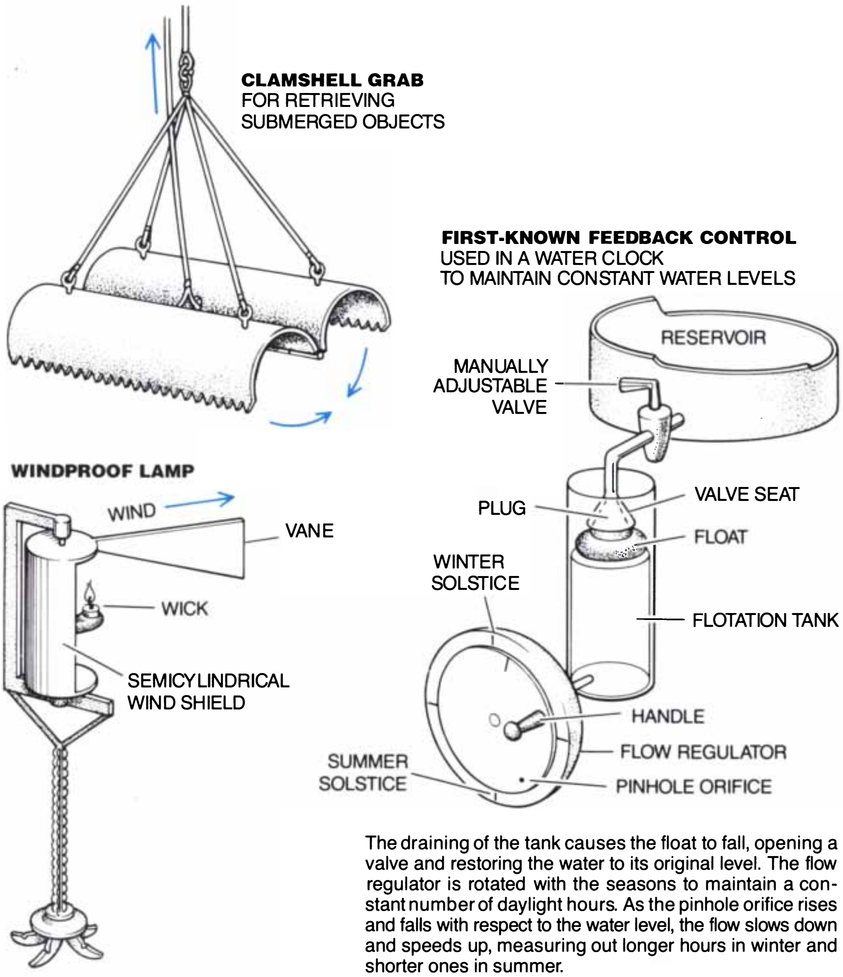
The second major treatise to have come down to modern times was written by al-Jazari at the close of the 12th century. He was a servant of the Artuqid princes, vassals of Saladin (who vanquished Richard the Lion Heart during the Third Crusade). His work places him in the front rank of mechanical engineers from any cultural region in pre-Renaissance times.

Several of al-Jazari's machines have been reconstructed by modern craftsmen working from his specifications, which provided far more detail than was customary in the days before patent law was invented [*see illustration on page 102*]. Such openness has rarely been encountered until recent times.

Al-Jazari's clocks all employed automata to mark the passage of the hours. These included birds that discharged pellets from their beaks onto cymbals, doors that opened to reveal the figures of humans, rotating Zodiac circles, the figures of musicians who struck drums or played trumpets and so on. Generally speaking, the prime movers transmitted power to these automata by means of pulley systems and tripping mechanisms. In the largest of the water clocks, which had a working face of about 11 feet high by 4.5 feet wide, the drive came from the steady descent of a heavy float in a circular reservoir.

Clearly, some means of maintaining a constant outflow from the reservoir was needed and was indeed achieved in a most remarkable way. A pipe made of cast bronze led out from the bottom of the reservoir. It was provided with a tap, and its end was bent down at right angles and formed into the seat of a conical valve. Directly below this outlet sat a small cylindrical vessel in which there bobbed a float with the valve plug on its upper surface.

When the tap was opened, water ran into the float chamber, the float rose and caused a plug to enter the valve's seat. Water was thus discharged from a pipe at the bottom of the float chamber, and the valve opened momentarily, whereupon water entered from the reservoir, the valve closed momentarily and so on. An almost constant head



The draining of the tank causes the float to fall, opening a valve and restoring the water to its original level. The flow regulator is rotated with the seasons to maintain a constant number of daylight hours. As the pinhole orifice rises and falls with respect to the water level, the flow slows down and speeds up, measuring out longer hours in winter and shorter ones in summer.

was therefore maintained in the float chamber by feedback control, and the large float in the reservoir descended at constant speed. Al-Jazari said he got the idea for his invention from a simpler version which he attributed to Archimedes.

This clock did not record equal hours of 60 minutes each, but temporal hours, that is to say, the hours of daylight or darkness were divided by 12 to give hours that varied with the seasons. This measurement required another piece of equipment: the pipe from the float chamber leading into a flow regulator, a device that allowed the orifice to be turned through a complete circle and thus to vary the static head below the surface of the water in the reservoir. Previous flow regulators had all been inaccurate, but al-Jazari describes how he calibrated the instrument accurately by painstaking trial-and-error methods. Another type of clock, which may have been al-Jazari's own invention, incorporates a closed-loop system: the clock worked as long as it was kept loaded with metal balls with which to strike a gong.

Al-Jazari also describes candle clocks, which all worked on a similar principle. Each design specified a large candle of uniform cross section and known weight (they even laid down the weight of the wick). The candle was installed inside a metal sheath, to which a cap was fitted. The cap was made absolutely flat by turning it on a lathe; it had a hole in its center, around which, on the upper side, was an indentation.

The candle, whose rate of burning was known, bore against the underside of the cap, and its wick passed through the hole. Wax collected in the indentation and could be removed periodically so that it did not interfere with steady burning. The bottom of the candle rested in a shallow dish that had a ring on its side connected through pulleys to a counterweight. As the candle burned away, the weight pushed it upward at a constant speed. The automata were operated from the dish at the bottom of the candle. No other candle clocks of this sophistication are known.

Other chapters of al-Jazari's work describe fountains and musical automata, which are of interest mainly because in them the flow of water alternated from one large tank to another at hourly or half-hourly intervals. Several ingenious devices for hydraulic switching were used to achieve this operation. Mechanical controls are also described in chapters dealing with a potpourri of devices, including a large metal door, a combination lock and a lock with four bolts.

We see for the first time in al-Jazari's work several concepts important for both design and construction: the lamination of timber to minimize warping, the static balancing of wheels, the use of wooden templates (a kind of pattern), the use of paper models to establish designs, the calibration of orifices, the grinding of the seats and plugs of valves together with emery powder to obtain a watertight fit, and the casting of metals in closed mold boxes with sand.

Precisely how Islamic mechanical technology entered Europe is unknown. Indeed, there may be instances of ideas being inherited directly from the Greco-Roman tradition into medieval Europe. Nor can we rule out cases of reinvention. When allowances have been made, however, it seems probable that some elements of the rich vein of Islamic mechanical engineering were transmitted to Europe.

Any such technological borrowing would probably have been mediated by contacts between craftsmen, by the inspection of existing machines working or in disrepair and by the reports of travelers. The most likely location for the transfer of information was Iberia during the long years in which Christians and Muslims coexisted.

The diffusion of the elements of machine technology from the lands of Islam to Europe may always remain partly conjectural. This should not in any way be allowed to devalue the achievements of the Muslim engineers, known and anonymous. Nor should we overemphasize the relevance of the Islamic inventions to the development of modern machinery. Of equal or greater importance is the contribution they made to the material wealth, and hence the cultural riches, of the medieval Near East.

FURTHER READING

ROMAN AND ISLAMIC WATER-LIFTING WHEELS. Thorkild Schiøler. Odense University Press, 1973.

A HISTORY OF ENGINEERING IN CLASSICAL AND MEDIEVAL TIMES. Donald R. Hill. Open Court Publishing Company, 1984.

ISLAMIC TECHNOLOGY: AN ILLUSTRATED HISTORY. Ahmad Y. al-Hassan and Donald R. Hill. UNESCO and Cambridge University Press, 1986.

CLEANING UP COAL

by Elizabeth Corcoran, *staff writer*

Pushed by Congress, electric utilities are on the verge of taking bold steps to cut emissions caused by burning America's most abundant fuel.

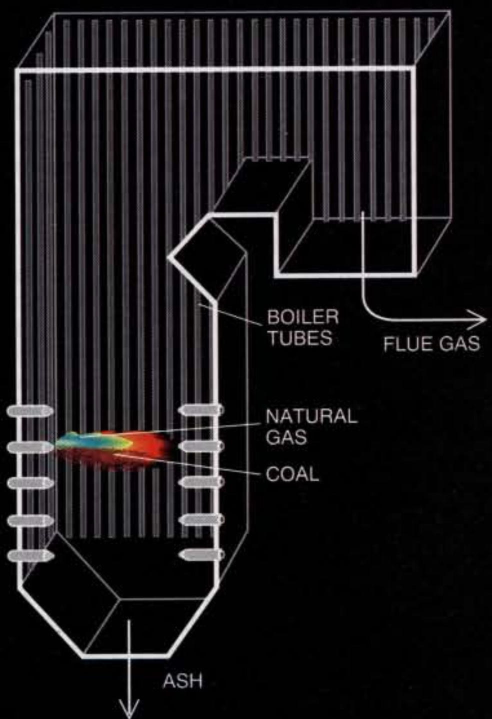
A seemingly endless chain of railroad cars brimming with shiny black chunks of coal snakes across the southwest Indiana plains. Its destination: the Gibson Generating Station, whose stacks rise 500 feet above the adjacent wildlife preserve and lake. Even by power plant standards, this facility is huge: it can churn out almost 3,200 megawatts of electricity—enough to light up more than a million and a half homes. To generate this energy, Gibson burns about 7.5 million tons of coal a year, or 200 railroad cars' worth every day. In the process, Gibson emits almost 300,000 tons of sulfur dioxide as well as other pollutants.

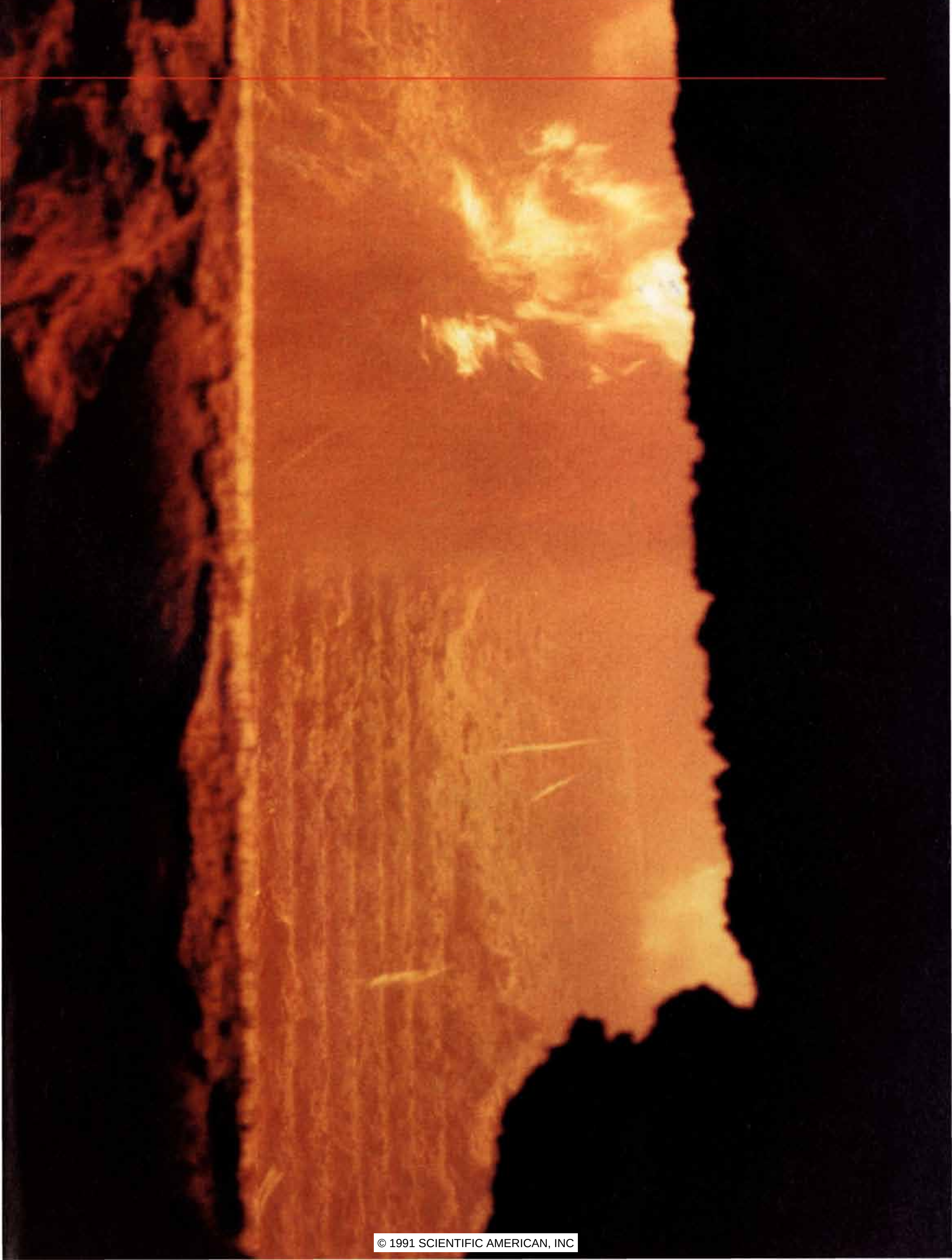
Up until 20 years ago, coal-fired utility plants were hailed as a triumph of modern technology for bringing light and heat into homes and businesses at affordable prices. Burning coal now generates about 56 percent of America's electricity.

More recently, environmental concerns have highlighted the hidden costs of such energy: coal-burning plants spew gases into the atmosphere that may fall back to the earth as acid rain or contribute to global warming. Among these gases are sulfur dioxide, oxides of nitrogen and carbon dioxide. Of the various sources of air pollution in the U.S., coal-fired power plants contribute about 70 percent of the SO_2, 30 percent of the NO_x and 35 percent of the CO_2.

Yet because coal accounts for more than 85 percent of potential U.S. fossil-fuel reserves, policymakers are not willing to abandon it. The Bush administration's National Energy Strategy calls for increased exploitation of all domestic fossil fuels. And even though burning natural gas creates negligible SO_2 and

VIEW INTO UTILITY BOILER shows coal dust and natural gas burning together. The heat from the combustion turns the water in boiler tubes (vertical bars) *into steam, which travels to a turbine-powered generator to produce electricity. Although most boilers burn only one type of fuel, workers from Energy Systems Associates have added natural gas igniters adjacent to the main coal burners on this boiler at Duquesne Light's Cheswick Generating Station near Pittsburgh. In the schematic below, the gas* (blue) *is used to ignite the coal stream* (brown). *Adding between 5 and 15 percent gas to the fuel mix also reduces emissions of nitrogen oxides by as much as 25 percent. This research project was funded in part by the Gas Research Institute.*





about half as much CO_2 as does burning coal, "we didn't promote natural gas at the expense of coal," says Linda G. Stuntz, deputy undersecretary for policy at the Department of Energy (DOE). (Natural gas accounts for 10 percent of potential U.S. fossil-fuel reserves.) To promote a large switch to natural gas might risk making the U.S. eventually dependent on foreign suppliers.

Congress, however, has ensured that the utility industry will not neglect its environmental duties. Last November lawmakers passed new amendments to the Clean Air Act, a rigorous collection of pollution-control policies aimed at tackling a range of air pollutants, including the two major precursors to acid rain. The law calls for utilities to reduce their annual SO_2 emissions from some 17.5 million tons to 8.9 million tons by the year 2000. It simultaneously calls for cuts in NO_x emissions of two million tons (to reach an annual emissions level of 11 million tons).

These pending cutbacks are forcing managers in the highly conservative power industry to take a hard look at the way they run their businesses and at the technologies they use. "The revision of the Clean Air Act has turned our industry upside down," says Gary R. Brandenberger, a vice president at Duquesne Light Company in Pittsburgh.

To meet the tough new standards, many utilities will still choose to wash, or "scrub," sulfur dioxide from flue gases, a costly solution that has practically been their only choice for nearly 15 years. But utilities are also scrutinizing a menu of lower-cost "clean-coal technologies" being demonstrated by the DOE. These include advanced methods of cleaning flue gases as well as techniques for removing pollutants from coal before and while it is burned.

Driving utilities to consider these options is the radical approach that Congress has unleashed to achieve the SO_2 reductions—an idea, long nurtured in universities, of granting tradable "allowances" for polluting. The policy is intended to minimize the total cost of controlling SO_2 by letting managers at every plant design their own antipollution strategy. Utilities can buy or sell allowances to help meet the emissions requirements that the Environmental Protection Agency will switch on in 1995. Those unable to balance their SO_2 emissions and allowance ledgers will be charged hefty fines. Congress has provided another incentive: their top officers will run the risk of jail.

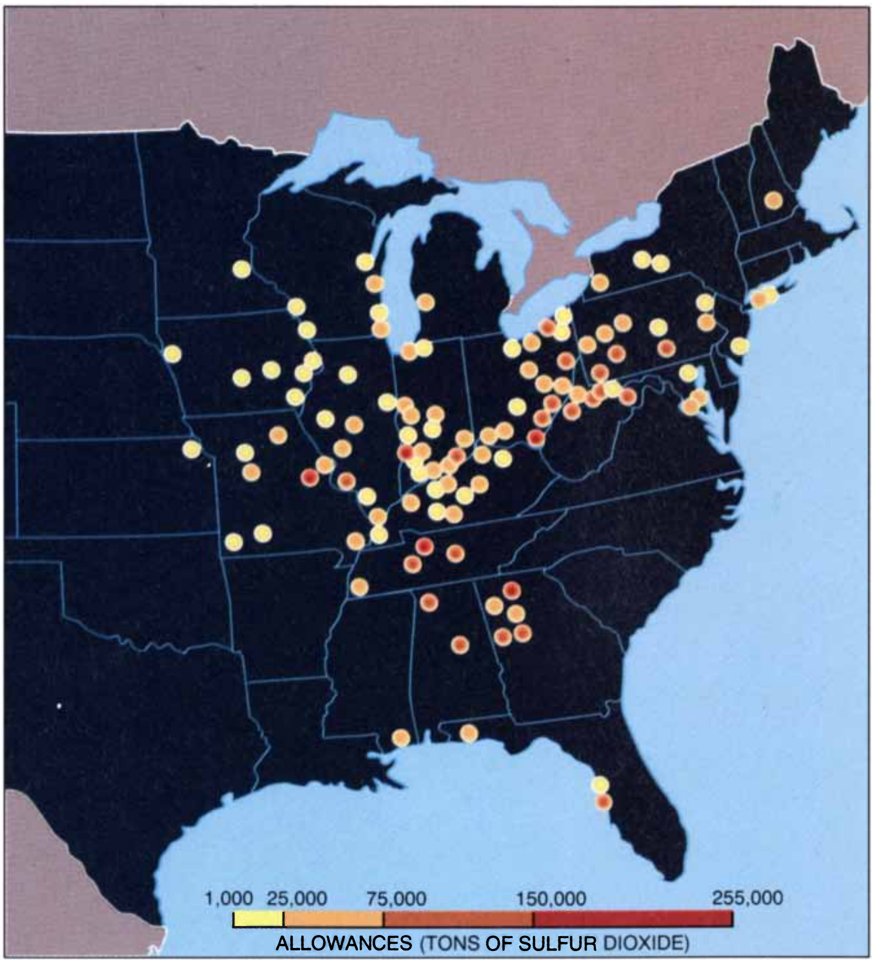
This time not only are the polluters on trial, so is the antipollution strategy. "The cornerstone of the Clean Air strategy is injecting competition into the utility industry," declares James E. Rogers, Jr., chairman of PSI Energy, the Indiana utility that owns Gibson. "Companies like ours have a great opportunity to reduce some of our costs if we show astuteness with respect to the allowance market." On the other hand, "it's an open question whether utilities have the right mind-set to play in this market," he says. But if emissions trading does reduce pollution at a lower cost than traditional "command and control" limits, it may well become the framework for future antipollution programs, such as limits on CO_2.

A God Named EPA

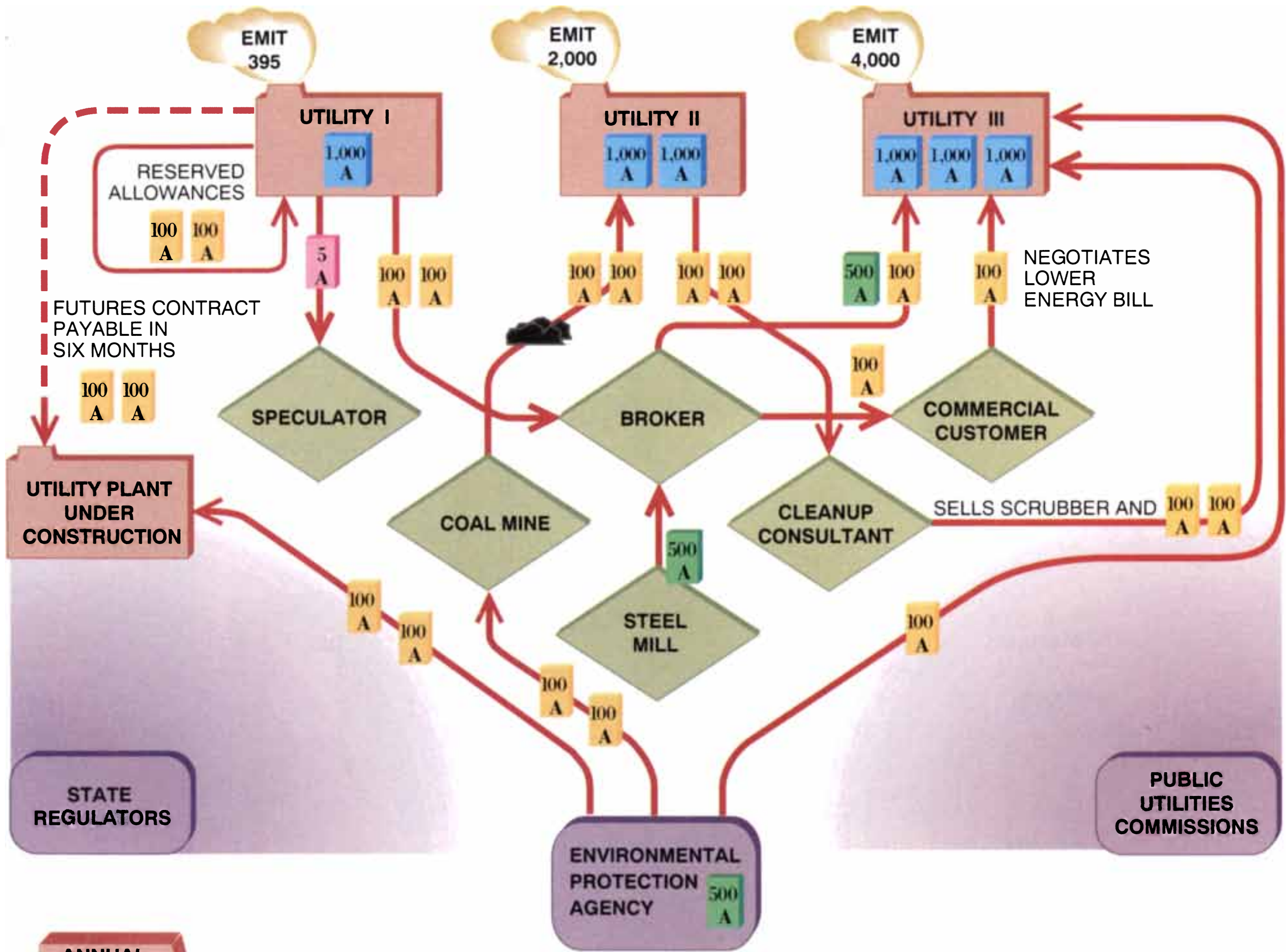
One of the incubators for the new antipollution program was an ornate 1912 building on Park Avenue South in New York City, the home of the Environmental Defense Fund (EDF). When Daniel J. Dudek, the fund's senior economist, is in town, he spends his time here in an office lined with economics texts, regulatory statements, computer equipment and a large chunk of shellacked coal. More often, Dudek is on the road, proclaiming the good tidings of emissions trading.

Dudek was one of a handful of people who coaxed first the EPA and then Congress into supporting emissions trading. He found allies in a variety of important corners, including Robert W. Hahn, then an economist with the Council of Economic Advisers. Yet none of them invented the idea. For more than 20 years, economists have suggested that rather than dictating how polluters should control emissions, the government should simply set targets, let the market figure out how to achieve them and reward the players who reach the goals the fastest. Only the "invisible hand" of the market will lead to the lowest-cost solutions, they argue.

But many environmentalists believed that controlling pollution called for the stronger arm of the law. In the late 1970s, for instance, Congress realized



POWER-GENERATING STATIONS TARGETED by the revised Clean Air Act have been granted annual allowances for emitting sulfur dioxide. (One allowance equals one ton of SO_2.) Beginning in 1995, these 111 plants will have to reduce their emissions to match their allowances or else buy additional allowances from one another.



ANNUAL ALLOWANCE ACCOUNTS

	UTILITY I	UTILITY II	UTILITY III	NEW PLANT
ALLOTTED	1,000	2,000	3,000	0
ACQUIRE	0	200 + COAL	1,000	400
SELL	405	200	0	0
RESERVE	200	0	0	400
EMIT	395	2,000	4,000	0

*SOURCE: John Palmisano, AER*X*

ALLOWANCE-TRADING SCENARIO portrays some of the players and strategies in the new market. Initially utilities hold most of the allowances (denoted by "A"). They either use their allowances (by emitting sulfur dioxide) or sell them. Brokers may arrange sales. Plants under construction must buy allowances to cover their future emissions. The EPA plans to hold annual auctions to give firms opportunities to buy allowances. Industrial sources of SO_2, such as steel mills, may "opt into" the program by complying voluntarily with the standards; the EPA then awards them allowances. Other companies—coal mines, cleanup consultants or commercial customers—may give or take allowances in lieu of cash payment for goods and services. The EPA also tracks utilities' balance of allowances (*left*). State regulators and public utilities commissions help the EPA oversee the program.

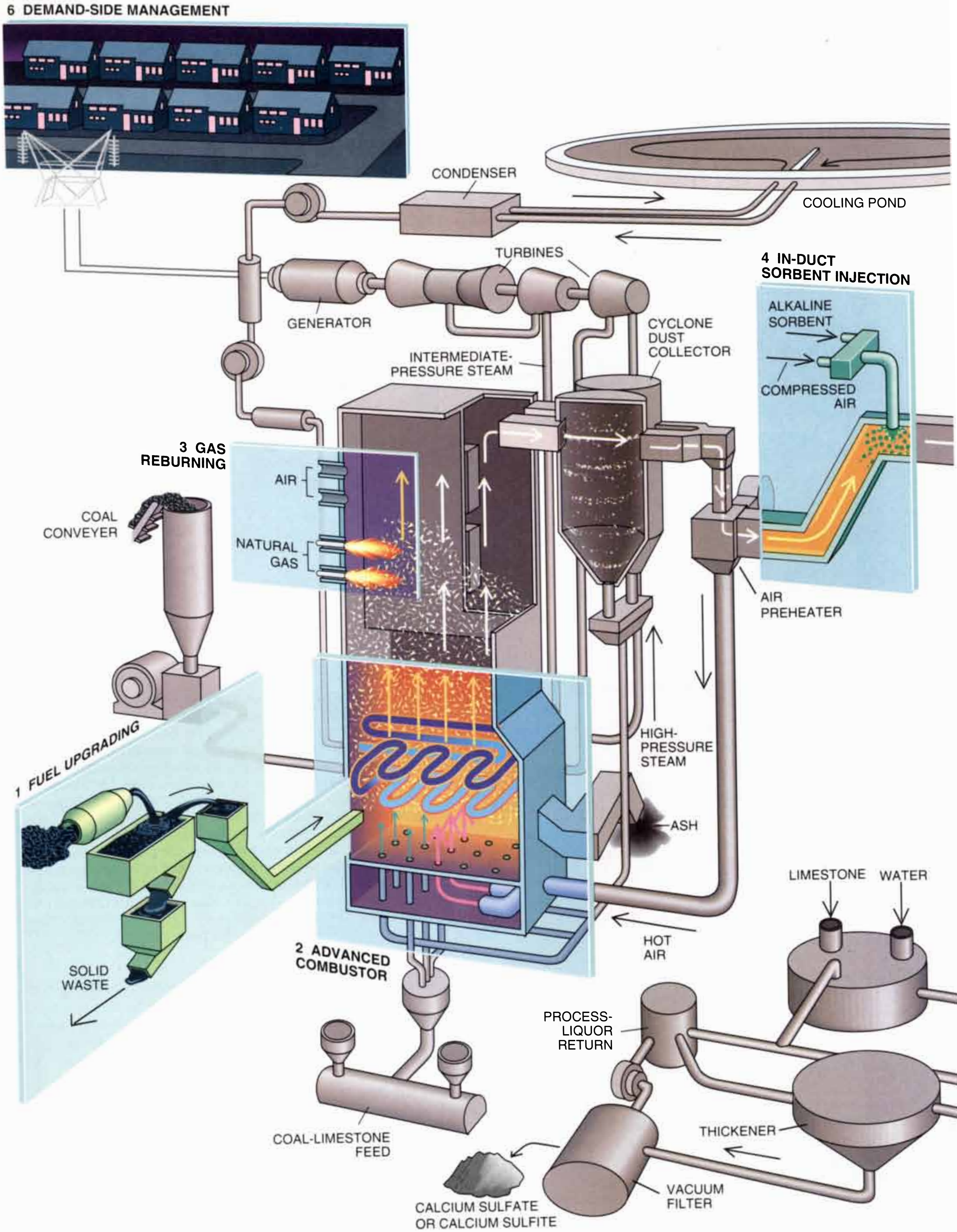
that its first Clean Air Act was ineffective. Regulators eventually required that newly built industrial sources of SO_2 be fitted with the "best available control technology," or BACT, to meet emissions standards. For utilities, it was, in effect, a velvet-gloved demand for a single solution: scrubbers.

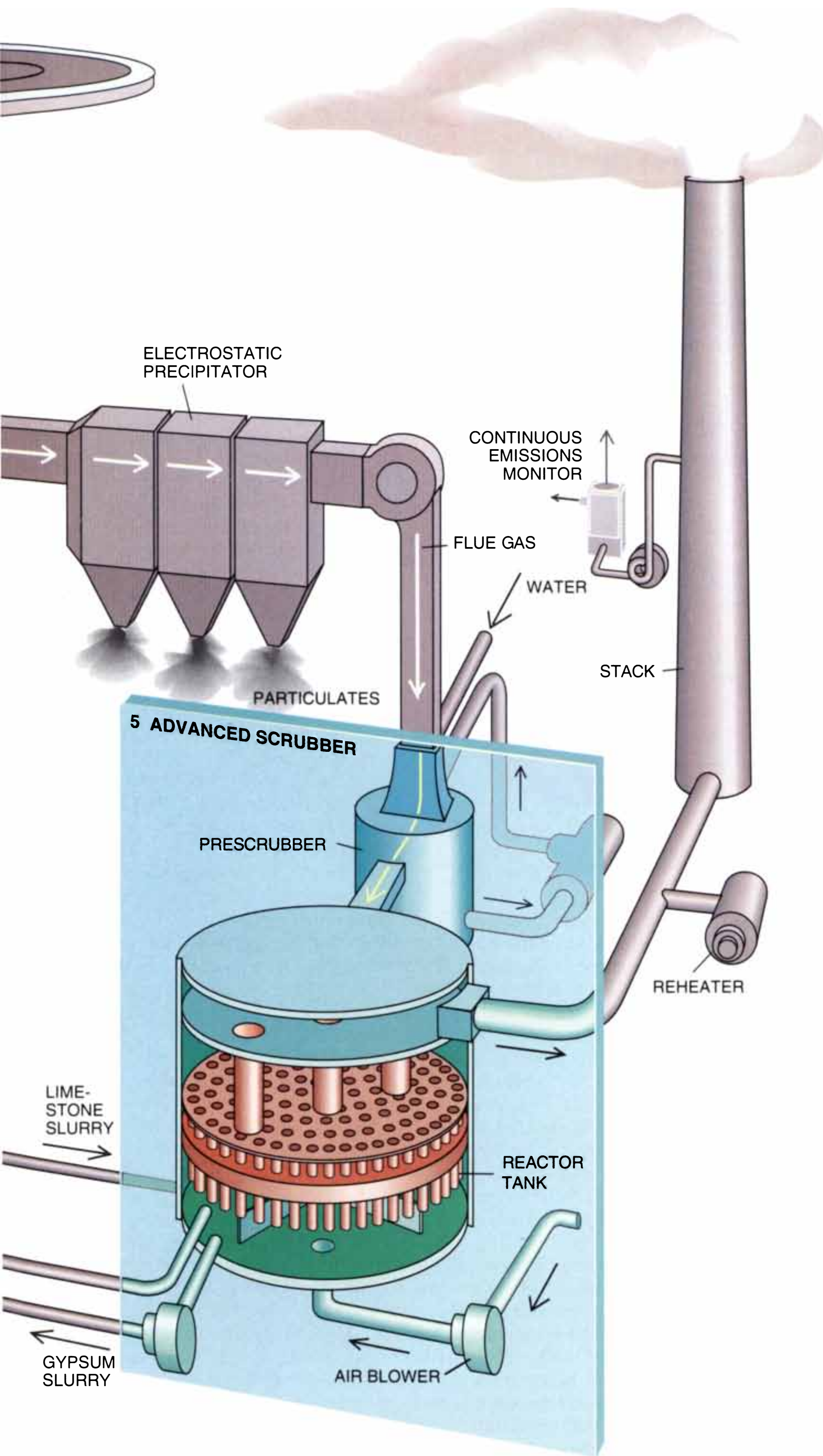
Scrubbers have been at best a mixed blessing. Formally named flue gas desulfurization units, scrubbers are mini-chemical plants that spray an alkaline mist—typically crushed limestone mixed with water—into the sulfur-laden flue gases produced by burning coal. The slurry reacts with the sulfur oxides in the flue gas and forms calcium sulfite or calcium sulfate (also called gypsum), which precipitates out as wet sludge.

More than 15 years and 150 scrubbers later, utilities have learned to deploy scrubbers in ways that slice SO_2 emissions by as much as 90 percent. But it has been a painful and costly learning experience. In the mid-1970s Duquesne became the first utility to scrub a plant. Those early days were memorable ones. "The acid vapor got the ducts wet and chewed them up to look like Swiss cheese," recalls Ralph L. Nelson, who manages fossil-fuel generation for the Pennsylvania company.

Scrubbers have also become a plant's best customer. The scrubber on one of Gibson's five boilers consumes between 10 and 12 megawatts of power, or about 2 percent of the boiler output. The unit's CO_2 emissions rise by roughly 4 percent, both because the utility must burn more coal to support the scrubber and because the scrubbing reaction itself generates some CO_2.

In addition, scrubbers churn out enormous waste. About a quarter mile from where the coal trains trundle onto


6 DEMAND-SIDE MANAGEMENT
CONDENSER
COOLING POND
TURBINES
4 IN-DUCT SORBENT INJECTION
GENERATOR
CYCLONE DUST COLLECTOR
ALKALINE SORBENT
COMPRESSED AIR
INTERMEDIATE-PRESSURE STEAM
3 GAS REBURNING
AIR
COAL CONVEYER
NATURAL GAS
AIR PREHEATER
1 FUEL UPGRADING
HIGH-PRESSURE STEAM
ASH
LIMESTONE
WATER
HOT AIR
SOLID WASTE
2 ADVANCED COMBUSTOR
PROCESS-LIQUOR RETURN
THICKENER
COAL-LIMESTONE FEED
VACUUM FILTER
CALCIUM SULFATE OR CALCIUM SULFITE



the Gibson site are the remains of the cleanup process: a growing pyramid of calcium sulfite that engineers expect will stand 80 feet high and cover about 80 acres when completed. "Some 1,000 years from now, people are going to dig here and decide, 'These people worshipped a god called EPA,'" quips Gregory L. Hauger, the operations superintendent at the Indiana plant.

More subtly, the BACT requirement froze out other emissions-control technologies, says Kurt E. Yeager, vice president of the Electric Power Research Institute (EPRI) in Palo Alto. Few other emerging technologies could attain the 70 to 90 percent reductions in SO_2 promised by scrubbers, even if some might have reduced emissions at lower costs. Moreover, utilities have had limited incentives to chase lower-cost solutions because their profit margins are set by state public utilities commissions, Yeager points out. Consumers not only are paying for preserving the air—they may be paying more than is necessary.

In the past, only a few regulators have experimented with using market-based programs to tempt utilities to seek lower-cost pollution controls. In the late 1970s the EPA encouraged local air-pollution regulators to create programs in which firms could offset emissions from new sources by reducing emissions from existing ones. California launched the first such program in 1976, aimed at new and modified industrial plants that emit smog-pro-

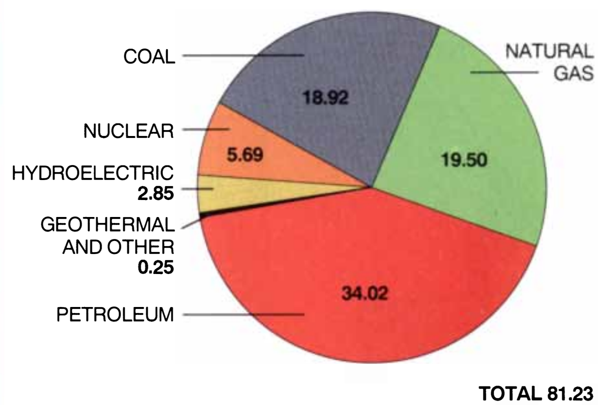
MENU OF TECHNOLOGIES for reducing emissions of sulfur dioxide (SO_2) and nitrogen oxides (NO_x) shows cleanup options. A facility might use some—but not all—of these choices. Vigorous physical and chemical cleaning before combustion (*1*) can reduce the sulfur content of some coals significantly. Replacing the boiler with a fluidized-bed combustor (*2*), in which coal and limestone are mixed together, can cut SO_2 emissions by as much as 90 percent. This chamber also burns coal more efficiently—and so at a lower temperature—than a conventional boiler, thereby reducing the formation of NO_x. In a conventional boiler, NO_x emissions can be trimmed by raising the ratio of fuel to oxygen by adding jets of natural gas and air (*3*). Another, low-cost strategy for reducing SO_2 emissions relies on injecting an alkaline sorbent into the duct between the boiler and the stack (*4*). The most effective and expensive way to cut SO_2 emissions is to scrub; advanced scrubbers, such as a jet-bubbling reactor (*5*), promise improved efficiency. All emissions can be reduced by convincing consumers to conserve electricity (*6*) and to use more energy-efficient appliances.

ducing volatile organic compounds. Yet largely because these programs were isolated test runs, many became entangled in red tape.

Then, about five years ago, a series of overlapping events conspired to push emissions trading into the acid-rain limelight. Although the voluminous \$500-million, 10-year National Acid Rain Precipitation Assessment Program would not wrap up until 1990, Congress began readying itself to revise laws limiting emissions of sulfur dioxide and nitrogen oxides.

With the administration being goaded by Canada to address acid rain, new laws seemed inevitable. The utility industry and the DOE initiated a joint program to develop alternatives to conventional scrubbers. The DOE promised to spend about \$2.5 billion between 1986 and 1994 to help build prototype and full-scale demonstrations of emissions-control technologies, many of which had been gathering dust on shelves or had been used abroad for years. Industry would match the federal dollars and provide the demonstration sites. The program "was a good-faith effort to show the industry's desire and willingness to explore new technology," Yeager says.

Sources of U.S. Energy, 1989
(QUADRILLION BRITISH THERMAL UNITS)



SOURCE: Energy Information Administration

At about the same time, Dudek was attending United Nations workshops on phasing out chlorofluorocarbons (CFCs), which destroy the earth's protective ozone layer. An international market for trading emissions of CFCs and other greenhouse gases "seemed to make eminent sense to me," Dudek says. So "I was quite unprepared for the response." The policymakers told him: prove such a scheme will work.

Many papers and much politicking later, Congress passed the Clean Air Act Amendments of 1990, featuring an emissions-trading program for precursors to acid rain as its most novel component. (There are several other sections to the amendments, including restrictions on CFCs, on airborne toxic chemicals and on emissions from mobile sources such as cars.)

To reach the goal of reducing utilities' annual SO_2 emissions to 8.9 million tons, Congress divided the requirements into two phases. By 1995 some 111 "large" generating stations (those producing 100 megawatts or more) will have to reduce their SO_2 emissions to 2.5 pounds per million British thermal units (mm-BTU) per year. By the year 2000 almost all coal-fired generating facilities will have to meet an SO_2 emissions benchmark of 1.2 lbs/mmBTU.

Trading comes into play because those targeted plants have been apportioned specific allowances for polluting. (The allowance awards are equivalent to the average fuel consumed by each plant between 1985 and 1987 multiplied by an emissions rate of 2.5 lbs/mmBTU.) These plants can buy or sell allowances to anyone willing to pay the going price: other utilities, emissions brokers, small, independent power producers, coal mines—even environmentalists who retire a ton of SO_2 to strike a public relations coup. (Planting trees would likely bring more environmental benefits for the dollar.) The price of a ton of SO_2 will vary depending on supply and demand. Some observers estimate the prices will range between \$400 and \$800 a ton. Because utilities must also conform to state emissions limits, no region is likely to wind up as an SO_2 dumping ground.

Other players can join the program. For instance, industrial SO_2 sources, such as steel mills, are not obliged to conform to the new standards. By limiting their emissions voluntarily and so "opting into" the program, they can create and then sell allowances. Moreover, any new power facility, be it a utility plant or small, independent power producer, must buy allowances from the market before generating emissions.

The EPA's role ought to be analogous to that of the scorekeeper at a baseball game. The challenge will be to track both the allowances held and the emissions reported from every affected unit at a utility plant. (Utilities typically have several units, or boilers, at a single site.) At the end of the year the accounts will be tallied: plants that exceed their allowance limits will be charged fines, have the emissions deducted from the next year's balance and risk incurring a criminal record.

Unused allowances will be carried over to the next year's account. After the year 2000, the annual allotment of allowances will remain at 8.9 million tons. This "cap," environmentalists say, will ensure that U.S. emissions of sulfur dioxide do not rise.

On Whose Shoulders?

How efficiently this market will work depends, observers say, on several factors, beginning with the rules created by the EPA. "Based on what people tell us, I think we're going to try to make the rules fairly simple and keep them that way," says Brian McLean, deputy director of the EPA's acid-rain division.

Yet some holes have been punched in the legislation. Extra allowances, for instance, will be available for utilities that comply early or build scrubbers (and so preserve some coal miners' jobs). The EPA will also hold an annual auction of allowances to give all a chance to buy shares.

Even so, "it's beyond controversy that the reductions will happen faster" under the new scheme, says John Palmisano, a former EPA official who helped to start the early emissions programs and had a key role in developing the revised laws. Palmisano now runs AER*X, an emissions brokerage in Washington, D.C., which arranges trades in the ongoing programs. He adds one caveat: "The EPA must have good enforcement and detection systems and be vigorous with its protection."

The real burden of making the amendments work will fall on those who had little to do with writing the new law: state environmental regulators and refrigerator-size monitoring instruments called continuous emissions monitors, or CEMs. "The states are on the front line of investigating and reporting on emissions," points out James K. Hambright, director of the bureau of air quality in Pennsylvania. "A conservative estimate is that all of the state programs would logically have to double or more" in staff and funding to carry out the new law, he notes.

Even the tools that regulators have will be hard-pressed to cope with the tougher standards. CEMs record the levels of specific contaminants in flue gases at fixed time intervals (such as once every 15 minutes). By 1995 all smokestacks of coal-fired plants will have these devices; although roughly three quarters of the largest utilities are already equipped with CEMs, few of the hundreds of units affected by the second phase of the new program do.

But CEMs, which can cost more than $100,000 to install, are often fussy. The devices are sensitive to temperature swings and easily become clogged with contaminants if they are not rigorously maintained. Even under good conditions, CEM measurements may vary by as much as 40 percentage points, according to Henry E. Beal, a vice president at Research-Cottrell Companies, a cleanup consulting firm in Somerville, N.J.

"There's no way we will find that kind of variability acceptable," counters Joseph Goffman, a principal architect of the amendments who is now helping the EPA devise regulations. Instead, he says, the EPA will probably write rules that add a "tax" proportional to the variability on top of any measurements.

Beal hopes that such regulations are settled quickly enough to give device makers time to build and test new instruments. Manufacturers can devise more accurate CEMs, "but we won't until someone wants to have one," he says.

Also complicating the work of state regulators will be the sheer volume of documents they must handle. This task is something Hambright, who began his career as an air pollution-control engineer almost 30 years ago, appreciates daily. Pennsylvania is one of the few states that already requires CEMs on utility smokestacks. The state's SO_2 limits are relatively tough: 1.2 lbs/ mmBTU in air basin regions, such as the Pittsburgh area; 4 lbs/mmBTU elsewhere.

Companies must report emissions quarterly. The Harrisburg headquarters receives about 170 thick reports every quarter to track 85 utility boilers and other industrial sources. Over the next few years, as additional plants install CEMs, the number of quarterly reports is likely to leap beyond 700, says Joseph C. Nazzaro, chief of Pennsylvania's continuous emissions-monitoring unit.

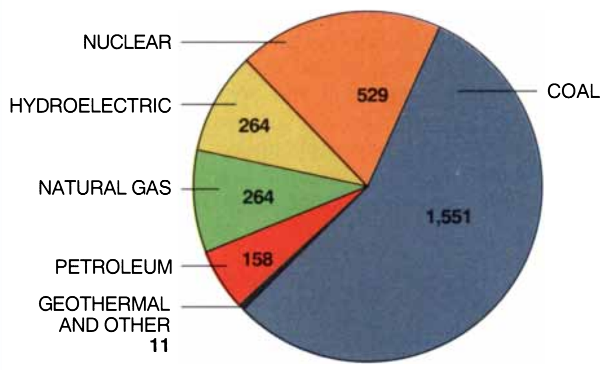
Crammed into a corner, his desk overflowing with papers, Nazzaro worries about keeping up with the work. To handle the upcoming regulations, additional staff "should be in their seats in this office right now," Hambright says. Because of a state budget squeeze, however, Hambright does not yet have the funds to hire any.

Public Watchdogs

Still, both the utilities and state environmental regulators agree that the wild cards in the evolving emissions market are the state public utilities commissions, or PUCs. The PUCs specialize in worrying. Among their duties, the PUCs aim to protect consumers from possible price gouging by the utilities and to ensure there will be sufficient energy to meet future demand. The commissions traditionally approve rate hikes proposed by the utilities and decree the maximum profit a utility can make on its sales.

Sources of U.S. Electricity, 1989

(BILLION KILOWATT-HOURS)



SOURCE: Energy Information Administration

Over the next year the PUCs will also scrutinize the mix of technology, strategy and trading options that utilities hope to use to meet their emissions limits. When they review a compliance plan, commissioners will weigh whether the utility has picked the least-cost options and so is justified in asking for any proposed rate hike. A PUC could decide that saving allowances for future growth amounts to wasting a valuable asset. Or a PUC projecting large future demand could compel a utility to hang onto excess allowances.

Utility managers fear the worst. At PSI Energy, Rogers seems convinced that complying early with the Clean Air standards will—over the long run—prove the cheapest strategy. "We've got a huge reservoir of SO_2 allowances," he says. Building scrubbers early would free PSI Energy to sell those allowances. But because he must make investment decisions first and appeal afterward for rate increases, Rogers worries about "Monday morning quarterbacking." As a result, Rogers is petitioning for full approval from the Indiana PUC before embarking on his compliance campaign. "I'd like all of us holding hands on our view of the future," he says.

In response, some PUCs are trying to clear the regulatory air. More than a year ago the Ohio public utilities commission promised neither to encourage local utilities to hoard allowances nor to penalize them automatically for overcomplying with the pollution regulations if they think they can make money selling excess allowances. "We said that we'd compare them with other compliance plans," explains Ashley C. Brown, an Ohio commissioner whom many consider to be among the most open to a flexible trading market.

How to satisfy future demand for power in a world with a tight emissions cap is nonetheless a vexing question for PUCs. One option is to postpone the need for additional capacity through demand-side management, or conservation. Designing adequate incentives that will encourage utilities to launch conservation programs is tricky, Brown acknowledges. Currently profits are tied directly to sales of electricity. Uncoupling these could inadvertently reward utilities for every dip in demand—even if the "conservation" was motivated by warm weather in December.

Many PUCs are likely to move slowly, reluctant to loosen their grip on the power producers—a plausible consequence of a truly fluid trading market. "The law says one ton cut in California can reduce one ton in Ohio. Coping with this is going to be a challenge for the PUCs," predicts Charles H. Goodman, a vice president at the Southern Company in Atlanta. Already officials in Florida and New York have hinted they may frown on local utilities engaging in out-of-state trades. Brown remains hopeful. "It's kind of an experiment," he adds. "There's no harm in trying."

There is, however, irony in the idea of experimenting with the U.S. electric utilities. Even those in the industry acknowledge they represent one of the slowest-moving businesses in the country. The government "couldn't have picked a more paranoid group of people to try this experiment on," says Kris A. McKinney, an administrator of emissions allowances at Wisconsin Power & Light in Madison. The utility is one of the few that has advertised that it will likely be selling credits.

Two of Wisconsin Power's plants stumbled over the federal trip wire of 2.5 lbs/mmBTU of sulfur dioxide. But the company will have to overcomply with the 1995 regulations; its state emissions limits are tougher. By 1993 Wisconsin utility systems must average no more than 1.2 lbs/mmBTU. "We'll probably be at that standard at both of the plants," McKinney says, and so they will wind up in the sellers' market. McKinney consequently spends his time analyzing which companies may be interested in buying allowances "whether they know it now or not."

Lacking the spurs of oncoming state regulations, other utilities are traveling similar routes, albeit more slowly. "A year ago I surveyed the electric utilities," says Palmisano of AER*X. "They said, 'Hell no, we're not going to do anything about trading.' Last spring 5 to 10 percent were looking into trading; in September, 18 percent. Now I see that people have got religion on this."

Regardless of the number of allowances traded, most utilities will still need technological means of controlling SO_2—before, during or after combustion. All will have to limit nitrogen oxides as well. In the absence of government BACT requirements, balancing the mix of options "turns into sheer economic analysis," says Jim Stevenson, a manager at Southern.

Technology Strategy

Many of the technologies being explored were pioneered in the U.S. but have been commercialized by companies in Europe and in Japan. "Germany views SO_2 and NO_x reductions as a business opportunity," says Brown of the Ohio PUC. Anticipating that other governments would eventually limit these pollutants, Germany created early, tough emissions restrictions that gave local companies a head start in developing cleanup technologies. In contrast, the Bush administration's recent budget proposals recommend shutting down the DOE's clean-coal program in 1992.

At the largest and dirtiest coal-fired plants, scrubbing and switching to a lower-sulfur variety of coal will still top the list of compliance options. Even these methods take on a shine under the allowance-trading program.

A handful of utilities are cutting their teeth on the next generation of scrubbers. One promising approach is a jet-bubbling reactor developed by Chiyoda Chemical Engineering & Construction Company in Japan. In this technique, flue gases are forced to bubble through a large vessel filled with an alkaline slurry. After absorbing the SO_2, the slurry is run through a centrifuge and transformed into dry gypsum. The reactor seems to present fewer maintenance problems than conventional scrubbers do, in part because the process is confined to a single reactor rather than three or four scrubber modules. Although effective, scrubbers may prove a short-sighted strategy if the government imposes limits on CO_2.

A relatively low-cost—and low-pain—alternative that will enable many utilities to meet their 1995 requirements relies on burning coals that naturally have a lower sulfur content. (Such fuel switching requires some boiler modifications because lower-sulfur fuels tend to produce more ash and less heat than their higher-sulfur cousins.) The increased interest in so-called compliance coals, however, is forcing high-sulfur coal companies to calculate ways of protecting their market share and miners' jobs. One strategy is to buy allowances and then sell them along with

GIBSON GENERATING STATION in southwest Indiana is bordered by a 3,000-acre artificial lake and a 160-acre wildlife preserve, both of which were built by the utility. Flue gas emissions from one of Gibson's five turbine generators are "scrubbed," or washed; the white plumes from one of the stacks is water vapor produced by scrubbing.


SCRUBBING SULFUR DIOXIDE begins when the flue gases produced by burning coal react with an alkaline mist in a scrubber module, producing sludge. The precipitates are transferred to "thickeners," or settling ponds (*left*), where the heavy materials sink and water is drained off. After further processing to remove additional water, the sludge is moved by conveyer belts to a staging area (*right*). Later, the material is transported by trucks to a landfill near Gibson.

the high-sulfur coal; another is to remove the sulfur from the coal before shipping the fuel to a utility.

Sulfur in coal is either chemically bound to the carbon atoms (called organic sulfur) or combined with finely dispersed iron particles that are distinct from the coal (pyritic sulfur). The proportions of pyritic and organic sulfur vary depending on the coal.

Much pyritic sulfur can be physically separated from the rest of the coal. One such technique, "froth" flotation, was first tried in the 1960s. A recent version of froth flotation relies on the fact that pyritic sulfur has a significantly higher specific gravity than does the organic material. By mixing pulverized coal into a fluid that has a lower specific gravity than the pyrite, the impurities sink and the organic matter floats.

J. Kelly Kindig, a principal researcher at Custom Coals International in Pittsburgh, is working with Duquesne on a more thorough coal-cleaning method that involves grinding the coal into particles measuring only a few microns in diameter, then separating the pyritic sulfur with a centrifugal cyclone. To attack the organic sulfur, Kindig adds a brew of limestone, soda ash and catalysts to the coal dust—a combination similar to the reactants in a scrubber. The cleaned coal is later reshaped into pellets so that it can be moved through a plant by existing conveyers.

"We've got a 575-megawatt plant [called Cheswick] that's not scrubbed, and we're trying to find a way not to scrub it," explains Brandenberger of Duquesne. Cleaning the coal beforehand looks particularly attractive because Duquesne owns the local mine. "Some coals, including that from Duquesne's mine, can be cleaned to within 1.2 lbs/mmBTU," Kindig asserts. Cleaned coal also produces less ash—and so more heat—per pound, he adds.

One aggressive chemical treatment is being tested by TRW in Los Angeles with DOE funding. In this molten caustic leaching process, coal is subjected to heated sodium hydroxide for about two hours, then washed and filtered. Both the sulfur and ash contents of the coal are significantly reduced. The process is still costly and experimental.

Farther down the road, researchers hope biological agents will come into play, says David J. Boron, a manager at the DOE's Pittsburgh Energy Technology Center. "We're trying to find out what makes bugs metabolize sulfur, then improve and regulate that sulfur metabolization," Boron says. "We want to be able to turn them on and not have them go on a prolonged coffee break."

Cleaning coal should bring another benefit: it should reduce the toxic minerals in the coal that could be emitted as airborne particles during combustion. The amendments call for the EPA to launch a three-year investigation of toxic emissions from utilities. Air toxics, says Yeager of EPRI, are "like a sword of Damocles hanging there."

Chlorides, for example, are ubiquitous in coal. "Even a small plant might need scrubbers to comply with the 10-ton limit" on toxics recently established for industrial sources, Yeager says. Some toxics may be sifted out by vigorous physical cleaning. Coals that carry significant doses of other elements, such as mercury and selenium, which are volatile at average stack temperatures, may require additional control techniques, Yeager warns, and such chemical treatments will not be cheap.

Baking Soda Fix

At marginally dirty plants and old facilities slated to be retired, other clean-coal strategies may rise to the fore. These methods largely fall into two categories: injecting a dry sorbent into either the boiler or duct or co-firing a coal-burning boiler with gas.

Dry sorbent-injection technologies are much like scrubbing minus the special equipment. The sorbent, often a hydrated lime, is injected into gases while they are in the boiler or as they are traveling between the boiler and the stack. The sorbent reacts with the gases, creating a dry waste product that can be hauled out with the fly ash. In principle, the technique could trim SO_2 emissions by 50 to 70 percent.

A small Houston company, NaTec Resources, is pushing its own sorbent twist. It sprays sodium bicarbonate into the duct. The highly reactive dry sodium sorbent produces a dry sodium sulfate by-product. "The wet-scrubbing market will still take 50 to 60 percent" of the business of utilities, concedes Glenn Hobratschk, an executive vice president at NaTec. But he believes dry sorbent injection can make a play for the balance of the market.

Burning natural gas in a coal-fired plant may seem like blasphemy to a generation of power plant managers who grew up thinking of natural gas as a limited, precious commodity. But some are beginning to extol the benefits of using gas to lower NO_x emissions. It can also replace oil to initiate firing and to supplement coal as a fuel.

Although combustion frees the nitrogen molecules in coal and so creates some NO_x emissions, most of the NO_x from power plants is a result of the burning process itself. At typical combustion temperatures of 3,000 degrees Fahrenheit, nitrogen in the air combines with oxygen to form nitrogen oxides. Lowering the boiler temperature or increasing the ratio of fuel to oxygen consequently trims NO_x emissions.

According to the consulting firm Energy Systems Associates, based in Pittsburgh, changing the mix of boiler fuel to include 5 to 15 percent natural gas raises the fuel-to-oxygen ratio in the boiler, cutting NO_x formation by as much as 25 percent. In tests conducted at Duquesne's Cheswick plant, operators also realized that gas could retard the buildup of slag, or molten ash, within the boiler. "We knew the benefits of it in about the first six months," says Nelson of Duquesne. Engineers at Energy Systems report that they can get as much as a 10 percent reduction in SO_2 by adding 5 percent gas.

NO_x can also be treated chemically. One costly technique, more widespread in Japan and Germany than in the U.S., is selective catalytic reduction. In this scheme, ammonia is mixed with flue gases in a chamber separate from the scrubber. The gases react, forming benign products, namely, water and molecular nitrogen.

The most extensive power plant surgery calls for replacing the boiler with a fluidized-bed combustor. It is not a cheap alternative even for new plants, such as the Texas–New Mexico Power Company's 150-megawatt unit that began pumping out electricity last September. Fluidized-bed combustors may also find a place in so-called repowered facilities, aging plants whose power-generating capacity has been boosted by an extensive overhaul.

Fluidized-bed combustion dates back to chemical processing work in Germany in the 1920s. In this technique, crushed coal and limestone are suspended in a boiler on jets of air. The churning of the particles ensures efficient combustion. Moreover, because boiler tubes are in direct contact with the burning particles, far more heat is transferred than in the conventional steam boiler. The boiler can therefore operate at relatively lower temperatures, minimizing NO_x formation. The limestone captures about 90 percent of the sulfur dioxide emissions. Some fluidized-bed combustors operate at atmospheric pressure. Others are pressurized and can drive a combined cycle, in which hot gases in the combustion chamber first drive gas turbines, then produce steam for a conventional turbine.

Even as utilities consider these and a battalion of other ways to meet the fast-approaching deadlines for SO_2 and NO_x emissions, they are keeping one eye on the coming front of environmental control, namely, limits on carbon dioxide. "I'm clearly a little cynical," Brandenberger says. "But I think CO_2 limits are not far down the road. That pushes me harder to find an alternative to scrubbers."

Many utilities scent a political deal in the wind. If Congress moves to limit carbon dioxide emissions, it will also have to relax the regulations for permitting nuclear power plants, they say. "You can't go to serious CO_2 controls without changing nuclear plant regulations," says Richard A. Abdoo, chairman of the Wisconsin Electric Power Company (WEPCO) in Milwaukee.

There are few technological fixes for reducing CO_2 caused by burning coal. The simplest solution is to burn less fuel—or from the vantage of utilities, to convince customers to use less power through demand-side management programs. In those states where the PUCs support conservation, utilities are reporting early successes.

Since 1987, for instance, WEPCO has committed about $100 million to conservation efforts and trimmed the growth in demand by 250 megawatts. Among its programs, WEPCO offers customers U.S. savings bonds for turning in old, leaky appliances such as refrigerators and investing in more energy-efficient ones.

Saving Power

Yet finding ways to reduce demand significantly is not always either obvious or easy. Workers at Battelle Pacific Northwest Laboratory had calculated that the Bonneville Power Administration could reduce electricity consumption by about one third if it encouraged homeowners to take such low-pain energy-saving measures as turning down thermostats. When the consultants surveyed homes, however, they found that most consumers were already sparing in their use of power. "We found that you get about half the savings that the energy models predicted," says W. Michael Warwick, a project manager at Battelle in Portland, Ore.

Longer term, one way to continue to make use of coal and still reduce CO_2 emissions relies on an updated version of a 19th-century concept: coal gasification. In this technique, coal is broken into methane by high-temperature steam and oxygen (or air). The purified gas can be burned and so run a gas turbine or a combined-cycle system. (In a combined-cycle system the heat from the burning gases is used to power a steam turbine in addition to the gas turbine.) A number of combined-cycle units are in use in the U.S.; these rely on natural gas, however. Integrated coal-gasification combined-cycle systems are still under development. PSI Energy, for instance, plans to transform part of an old plant into such a power generator—provided it wins DOE support for the $250-million project.

Even so, there is no way to burn either natural or coal gas without creating CO_2 emissions. Efforts focus, as a result, on how to capture those gases. Planting more trees creates one sink for absorbing CO_2; pumping the gases deep into the ocean may be another way to sequester them. Along with the Solar Energy Research Institute in Golden, Colo., DOE researchers are also trying to use microalgae to convert the CO_2 into biomass, liquefy it, then use the product as transportation fuel.

No utility manager believes that the U.S. can afford to abandon its abundant coal resources altogether. "The cornerstone of energy policy is coal," Goodman declares. And if the world does move toward global restrictions on CO_2 emissions, trading may well be a part of that calculus. "We've got to balance the three E's"—energy, the environment and economic development, Rogers says. "And sulfur dioxide is a great experiment to see if market solutions can work in resolving environmental problems."

FURTHER READING

COAL-FIRED POWER PLANTS FOR THE FUTURE. Richard E. Balzhiser and Kurt E. Yeager in *Scientific American*, Vol. 257, No. 3, pages 100–107; September 1987.

LAW AND ECONOMICS SYMPOSIUM: NEW DIRECTIONS IN ENVIRONMENTAL POLICY. *Columbia Journal of Environmental Law*, Vol. 13, No. 2, 1988.

THE THEORY OF ENVIRONMENTAL POLICY. William J. Baumol and Wallace E. Oates. Cambridge University Press, 1988.

CLEAN AIR ACT AMENDMENTS OF 1990. Report 101-952: Conference Report to accompany S. 1630. U.S. Government Printing Office, 1990.

ENERGY FOR PLANET EARTH. *Scientific American* special issue, Vol. 263, No. 3; September 1990.

CLEAN COAL TECHNOLOGY DEMONSTRATION PROGRAM: PROGRAM UPDATE, 1990. U.S. Department of Energy, February 1991. Available from National Technical Information Service.

SCIENCE AND BUSINESS

Relaxation receptors, fiber-optic aircraft, remote sensing for petri dishes, real estate bubbles

Handful of Pain

Pressure mounts to alleviate repetitive-motion injuries

In his *Treatise on the Diseases of Workers* in 1700, Bernardino Ramazzini, an Italian physician and philosopher, made reference to the "harvest of diseases" that workers experience from "certain violent and irregular motions and unnatural postures of the body." The 1893 edition of *Gray's Anatomy* described a "sausage-shaped swelling" of the hand that it characterized as washerwoman's sprain.

Today the names may be different, but the pain is the same. The clinical term "carpal tunnel syndrome" or "tendinitis" may be dubbed pricer's palsy by a store clerk, pickle-pusher's thumb by a worker at a food-processing plant (the last pickle may go in the jar manually) and Nintendonitis by an overzealous video-game player.

The plethora of names points to a growing awareness of problems that can result from the forceful repetition of hand, wrist and upper-body movements while using a computer keyboard, a factory tool or an electronic checkout scanner. "People have gone beyond the denial phase," says Thomas J. Armstrong, a researcher at the Center for Ergonomics at the University of Michigan at Ann Arbor.

CUMULATIVE-TRAUMA disorders are endemic in the meat-packing and poultry industries. Source: United Food and Commercial Workers Union.

Indeed, the number of such cumulative-trauma cases reported by workers increased nearly fourfold from 1985 to 1989. Responsibility may lie with the rapid-fire pace of factory work and office assembly lines, where work is parceled into highly specialized tasks. "Jobs are more repetitive than they used to be," observes Franklin E. Mirer, director of the United Auto Workers health and safety department. "The light-duty and rotational jobs have been cut out of the system, and people are pushed all the time."

Under pressure from unions and Congress, the Occupational Safety and Health Administration (OSHA) is now taking a harder stance on cumulative-trauma injuries. The agency is assessing penalties for "egregious" conditions, fining a company for each employee exposed to a violation instead of levying a single fine. That policy has increased the amount of fines from thousands of dollars to sometimes more than a million. General Motors, Ford and Chrysler have agreed to pay such penalties during the past 18 months. Earlier, OSHA singled out some of the largest meat-packing firms—IBP and John Morrell & Co.

The meat-packing industry has been one of OSHA's major targets because its workers suffer from cumulative-trauma injuries about 12 times more than do workers who produce nondurable goods such as clothing. Last August then Secretary of Labor Elizabeth Dole released OSHA's Ergonomics Program Management Guidelines for Meatpacking Plants. "In some plants we went into, nearly 80 percent of the people in some jobs had cumulative trauma," says Raymond Donnelly, director of OSHA's enforcement division.

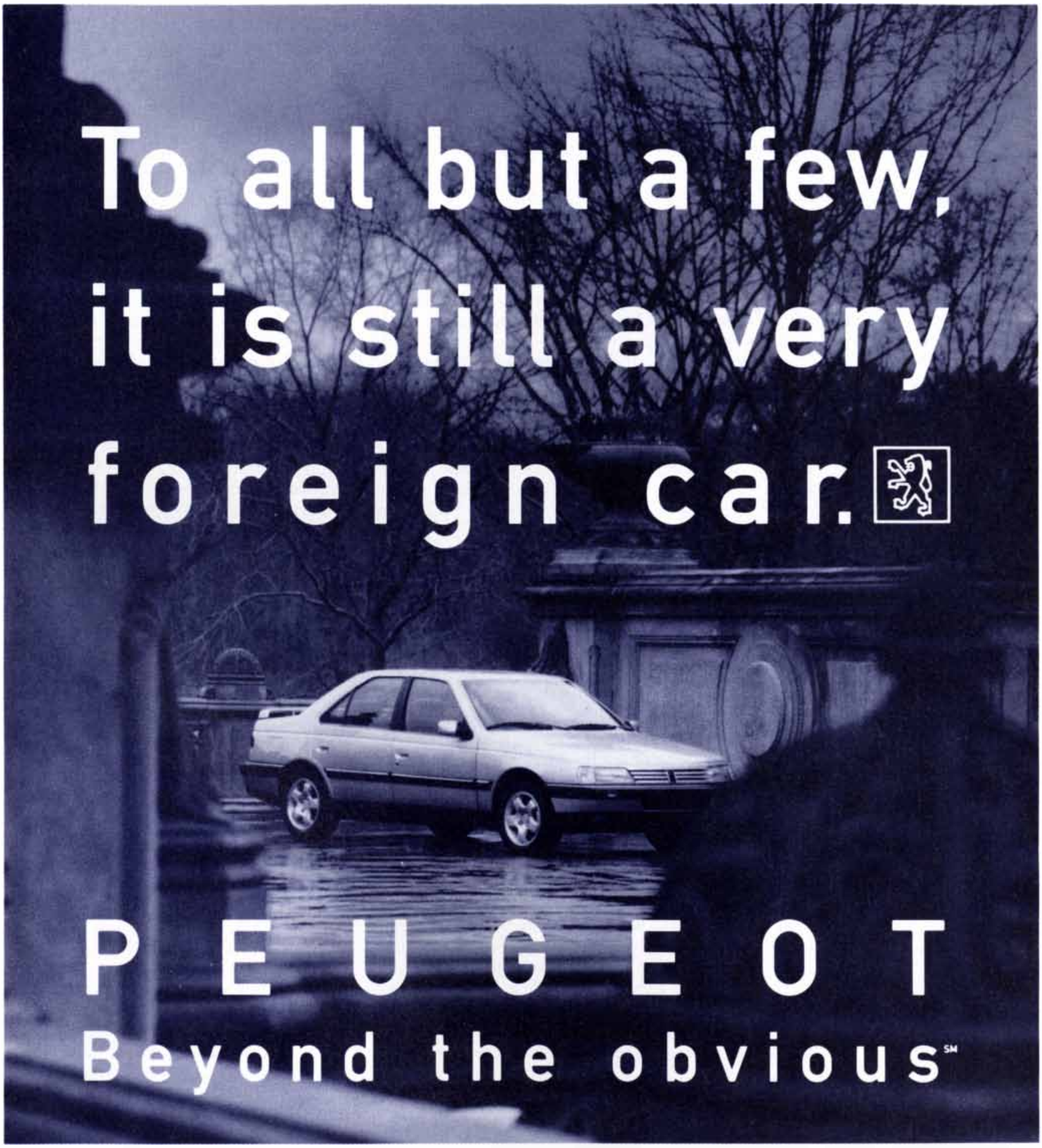
To fight the growing wave of workmen's compensation claims, some employers have relied on medically dubious treatments, such as dispensing vitamin B_6 tablets, applying hot compresses to already inflamed areas and requiring that employees keep an injured joint immobilized while continuing to work, a practice that may expose them to further harm.

Health and safety trade magazines, moreover, advertise services that can supposedly measure nerve activity in the hand and wrist to determine susceptibility to carpal tunnel syndrome. But no accurate technique exists to identify likely candidates for the disorder, researchers say.

More serious efforts to prevent injury rely on developing tools that require less force while allowing the hand and wrist to assume more natural positions. According to the United Food and Commercial Workers Union (UFCW), one instrument, a circular-bladed power knife called the Whizard Knife, has been a contributor to repetitive-motion injury. Its manufacturer, Bettcher Industries, has designed a new model that damps vibration and allows for a better grip. Separately, workers at IBP in Dakota City, Neb., have concocted their own designs for boning knife handles and blade angles, but finding a domestic manufacturer for the knives has proved difficult.

In the office, designers also are trying to develop keyboards that reduce persistent pounding and wrist bending. One model, the TONY!, splits a keyboard in half to allow the sides to be rotated, or even angled upward into an A shape, to accommodate the typist's hands. Another new model, AccuKey, has four keys for either hand, each of which can be manipulated into one of three positions. Using all eight keys, a typist can form "chords" in a total of 5,561 combinations, although only about 500 combinations are actually used by typists.

For now, however, designers of work implements are crippled by the meager literature documenting the complex interplay of force, posture and repetition involved in manual work in the office and factory. Managing work—the meaning of the word "ergonomics"—has oc-



1991 Peugeot 405 models are intelligently priced between $15,490 and $21,990. MSRP. Excludes tax, title, options, registration and destination charges. In Canada call 1-416-566-1900.

Europeans have long appreciated the remarkable engineering and world-class styling of a Peugeot.

Yet, here in America, a Peugeot 405 is a rare pleasure. Perhaps because a full appreciation of one requires the kind of thorough scrutiny few car buyers exercise.

The 405's patented 8-valve shock absorbers, for example, are far from obvious, yet they contribute to the renowned Peugeot road feel that is immediately apparent. Two densities of foam are a subtle but effective way to eliminate seat springs and the road vibrations they transmit. And less obvious still is the sophisticated composite barrier beneath the roof that absorbs road noise.

But you'll quickly understand the value of every aspect of the Peugeot 405 with closer scrutiny. For the dealer nearest you call 1-800-447-4700.

cupied only a few university and industry researchers trying to answer the question of how much toil is too much. "We don't really have any hard evidence of how many repetitions can create an injury in a certain joint," says Manny Halpern, a biomechanical ergonomist at the Hospital for Joint Diseases in New York City.

Measurement of the amount of wrist and hand stress is an inexact science. Typically researchers videotape people working and then watch the tapes in slow motion, noting posture and number of repeated movements. These observations are supplemented by measurements of vibration and force exertion. For example, electromyographs can monitor activity at the surface of muscles to estimate force.

The Center for Ergonomics at Ann Arbor has studied tool and workstation designs that minimize the force and frequency of manual tasks, and it has come up with software models of mannequins that can be used in computer-aided design to test whether a tool position results in awkward hand posture. An American National Standards Institute committee is trying to assemble existing research in order to develop guidelines for tool and workplace design. OSHA may eventually use the committee's work as a partial basis for formulating mandatory regulations.

Because of the growing number of workmen's compensation claims, insurers also are interested in studying ergonomics. The Liberty Mutual Insurance Group, the nation's largest underwriter of workmen's compensation insurance, has set up a laboratory to study what happens when someone performs the same manual task over and over. It has recruited about 30 women to spend a month's time doing nothing but flexing their hands every four seconds for seven hours a day on a handgrip. The research is directed toward developing workplace guidelines for employers, similar to ones the company researched for lower-back injuries. "We tell people, we want you to work as hard as you can without going home at night with a sore hand," says Stover H. Snook, the project director for ergonomics.

Epidemiological studies and new technologies may help reduce trauma from force and awkward posture. But it has been difficult for management to resist the inexorable pressure to speed the assembly line.

In *Working for the Japanese*, published last year, Joseph and Suzy Fucini document a higher than normal incidence of cumulative-trauma injuries at a Mazda plant in Flat Rock, Mich., which began operation in 1987. They quote union dissidents who criticize the plant's just-in-time production system, which keeps workers busy 57 seconds out of every minute, in contrast to 45 seconds at plants owned by the Big Three automobile makers.

Automakers are also now under increasing pressure to put on the brakes. Making the job fit a worker's abilities is the goal of about 75 management-labor committees that are the foundation of Ford's company-wide ergonomics program. The UAW-Ford Ergonomics Process was set up to meet the terms of the OSHA settlement and an earlier agreement with the United Auto Workers. (General Motors and Chrysler have adopted similar initiatives.)

Some unions fear, nonetheless, that OSHA's limited ability to monitor industry may enable dangerous work practices to persist. Deborah E. Berkowitz, director of the office of occupational safety and health at the UFCW, which represents the meat-packing industry, says understaffing at the agency has led to a controversial proposal to allow meat-packing companies to enter into voluntary agreements with OSHA. A congressional hearing on the OSHA plan was held in mid-March.

The UFCW is concerned that these agreements may allow for self-policing that will enable the industry to avoid making fundamental changes in working conditions. But Roger Stephens, OSHA's chief ergonomist, is adamant that such adjustments are needed. "If somebody spends six hours putting cakes into wrappers, the company should find something else for that person to do for two hours," he insists. —*Gary Stix*

Who Gets Cumulative-Trauma Disorders?

INDUSTRY	INCIDENCE*
Meat packing	799
Motor vehicles	453
Shipbuilding and repair	242
Frozen bakery products (except bread)	219
Pens and mechanical pencils	206
Metal office furniture	195
Vacuum cleaners	177
Average for private sector	**19**

*Per 10,000 full-time workers, 1989

SOURCE: U.S. Bureau of Labor Statistics

Light Flight

Optical fibers may be the nerves of new aircraft

The wiring in some modern commercial aircraft would stretch in a straight line from London to the Strait of Dover. But while shuttling vital control signals from cockpit to flight-control computers and back, the copper conductors sometimes take on a less desirable role. "Wires are becoming the dominant antennae in these aircraft," says Robert J. Baumbick, a senior electronics research engineer at the National Aeronautics and Space Administration's Lewis Research Center in Cleveland.

The danger posed by lightning bolts, radar or a radio transmitter means that aircraft designers must carefully shield copper cabling, adding a substantial burden of weight to the aircraft. New airframes have also begun to incorporate composite materials that eliminate the natural barrier to potentially dangerous electromagnetic interference (EMI) furnished by the metallic airframe.

So designers of aviation electronics systems at such government laboratories as Lewis and at major airframe manufacturers are now developing fly-by-light control systems. In such systems, optical fibers would become the medium for controlling electronic devices dispersed throughout the aircraft, reducing the amount and weight of cabling by as much as 50 percent.

Military reality provides an important part of the incentive. The potential for an electromagnetic blizzard above a war zone has been recognized by the Department of Defense since the 1970s, when it commissioned the YC-14, a Boeing short-takeoff and landing prototype that never went into production. The transport airplane connected flight computers with fiber-optic cables.

Susceptibility to EMI has already embroiled an army helicopter program in controversy. In 1988 the army ordered that additional shielding be added to the Sikorsky Aircraft Black Hawk helicopter. It made the announcement after the tail rotor pedal froze when the aircraft passed by a radio transmitter in Germany. A news report by the Knight-Ridder Newspapers in 1987 quoted Pentagon sources and documents as saying that EMI may have been responsible for a number of Black Hawk crashes, a contention denied by the army.

The term "fly by light" is reserved for the network of sensors and actuators that receive optical signals from an onboard flight-control computer to adjust the position of an aileron or rudder. (The electronic predecessor to fiber-optic systems was known as fly-by-wire.)

In fly-by-light, optical sensors on the wing or tail of an aircraft note the position of a flap or rudder, and on an engine, they check the temperature, throttle position or other parameters. As a flight-control surface, such as a rudder, moves, so too does a sensor, which changes the amplitude or another characteristic of an optical signal generated by a transmitter in the fuselage and relayed over a fiber. This information may then be relayed to cockpit instruments, or it may be processed by the flight-control computer to make automatic maneuvering adjustments to the wing or tail surfaces.

Despite its promise, however, fly-by-light technology has not moved beyond the prototype stage. The most extensive test of optically controlled flight was a modified Black Hawk helicopter, dubbed the Light Hawk. Boeing engineered the system—called Advanced Digital Optical Control System (ADOCS)—for the army in the mid-1980s.

Although the Light Hawk prototype successfully logged more than 500 flight hours, the program also underlined reliability and manufacturing problems that systems designers are still wrestling with: the high cost of components, signal loss from bending of the fiber, failure of laser diodes used in the transmitting units and contamination of the sensors. "Most people have seen that the technology will work," says Walter L. Glomb, Jr., a manager of product development for the United Technologies Research Center. "Now the question is, Will it perform well enough over a 20-year lifetime?"

Integrating readily producible sensors in a fighter jet is the current endeavor of NASA's Lewis Research Center in Hampton, Va., and the navy. Called Fiber Optic Control System Integration, or FOCSI, the program is scheduled to test sensors for position, speed, temperature, propulsion and pressure in the navy's F/A-18 fighter by mid-1993.

Still needed are better light sources for transmitters used to send optical signals to sensors. Most lasers cannot tolerate the intense heat from aircraft engines and supersonic flight without external cooling, so less powerful, light-emitting diodes have been used. Boeing's High Technology Center in Seattle has begun tests on a laser made of gallium indium arsenide that can withstand elevated flight temperatures and could be used to monitor many sensors at one time.

Once sensor technology is perfected, development is likely to shift toward research on optical components that would replace the electrical control of hydraulic actuators used in moving flaps, rudders and other flight-control surfaces. The United Technologies Research Center has developed an actuator that works by using an optical signal to heat a small amount of hydraulic fluid. The tiny pressure generated by the heated fluid is amplified by the actuator's hydraulic system and used to move a wing flap or tail surface.

Fiber-optics may eventually supply a self-diagnosing system for detecting structural impairments to the aircraft while in flight. These "smart skins"—built from optical sensors and fibers embedded in a composite airframe—could detect the presence of cracks and other stresses from wear or battle damage. Composites would enable a radar or infrared detector to be built as an integral section of the aircraft body, with signaling to avionics computers carried out through fiber-optic links.

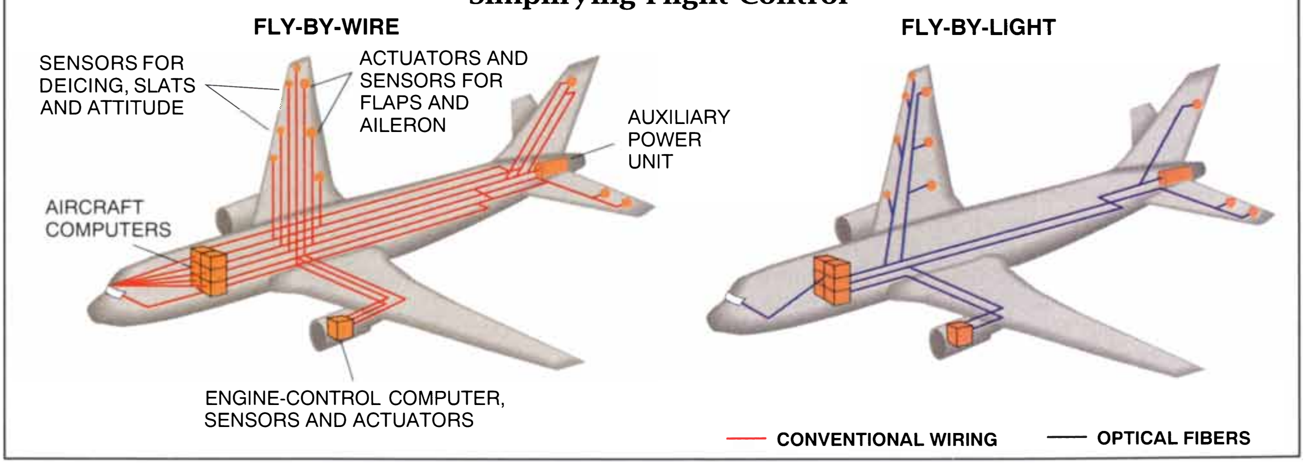
The lack of a unified effort has prompted NASA to propose the introduction of an approximately $70-million five-year program coordinating industrial, university and federal laboratory work on fly-by-light for commercial aircraft. "We want to achieve a nationwide foundation in fly-by-light technology," says Cary R. Spitzer, manager of program planning and contracts for NASA's Langley Research Center. "As it is now, people are working on bits and pieces of it."

For now, the military is not ready to embrace the concept fully. The development program for the army's new Light Helicopter, a replacement for a fleet of Vietnam-era armed reconnaissance helicopters, evaluated and rejected optical sensors as immature technology. The U.S. Air Force's Advanced Tactical Fighter (ATF) development program, a successor for the 16-year-old F-15, also looked at and discarded such a system. The ATF and the Light Helicopter, however, may incorporate a high-speed optical data bus, a fiber network connecting various flight computers.

Enabling computers to talk to one another over high-capacity fiber offers a distinct benefit in the cockpit, where fighter pilots are barraged with instrument data. The tens of millions of bits streaming each second over a fiber network provides enough communications capacity to "fuse" infrared, radar and other information readily into a single image for display to the pilot. This image can be superimposed on the position coordinates of a digital map.

Integration of aircraft systems using lightwaves will likely continue. One day, during a stopover, a maintenance technician may hook up a fuel line. A few feet away another worker will plug a fiber into an optical interface on the fuselage. Route and weather information, along with *Rocky,* part XXV, will be loaded into the big digital flying machine at hundreds of millions of bits per second. —*Gary Stix*

Simplifying Flight Control



SOURCE: United Technologies Research Center

Cadmium Charges

The environmental costs of batteries are stacking up

"Get the lead out" was the first message to battery makers from environmentalists intent on keeping toxic metals out of landfills. So manufacturers agreed to take back spent car batteries and recycle the lead. Next came mercury. In the mid-1980s the poisonous metal that coats electrodes in alkaline batteries accounted for 1 percent of a cell's weight. It will contribute no more than 0.025 percent by early 1992. Some batteries already meet this requirement.

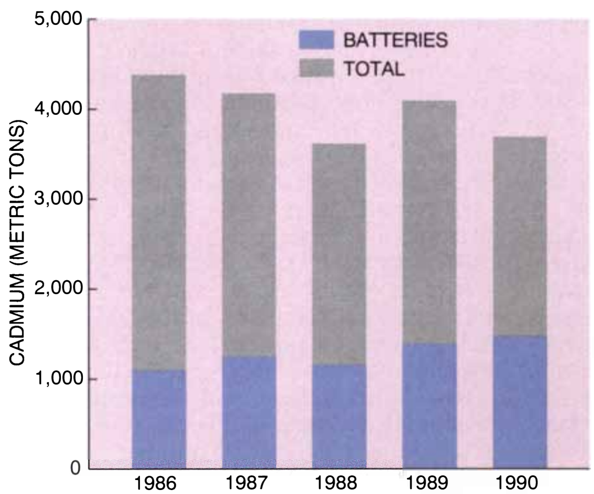
In line to leave now is cadmium. The carcinogenic metal powers rechargeable batteries by way of a reversible chemical reaction with nickel. Most of the 280 million Ni-Cd cells purchased in the U.S. last year went home sealed deep within cordless appliances such as power tools, miniature vacuum cleaners and toothbrushes. Manufacturers now are beginning to promote removable Ni-Cds for portable radios and other electronic products. Sales of Ni-Cds are climbing steadily as consumers endeavor to limit what they throw away.

Unfortunately, plenty of rechargeable batteries are being disposed of. More than half of the estimated 1,775 tons of cadmium in the municipal waste stream came from batteries in 1987, according to the most recent study conducted for the Environmental Protection Agency. In landfills the batteries can release toxic cadmium. When they are burned with other garbage, the resulting ash becomes so contaminated with cadmium that it must be considered hazardous waste. Dumping fees at specially lined landfills run up to $1,000 a ton.

Connecticut and Minnesota have already passed legislation aimed at keeping cadmium out of landfills and incinerators. More states, including New Jersey, Vermont, Michigan, California and Oregon, are considering similar actions. "The bottom line is that Ni-Cds cannot go into the waste stream," declares Jean Wagenius, a Minnesota state representative from Hennepin County. "Manufacturers must be ultimately responsible."

Minnesota's 1990 legislation—which other states are regarding as a model—was co-authored by Wagenius and Senator Gregory L. Dahl. It gives makers a few years to label consumer batteries by their electrometallic content, which should simplify sorting and ultimately reprocessing. By then, consumers must also be able to remove batteries from products, to prevent incognito trips to the dump. Eventually Minnesota wants to give consumers and retailers incentives to recycle Ni-Cds. The state's pending 1991 legislation would tack a redeemable surcharge onto the product, similar to a five-cent can deposit. The amount, however, would be hefty enough to encourage compliance.

Rechargeable Batteries' Share of U.S. Cadmium Consumption



SOURCE: U.S. Bureau of Mines

The prospect of a wave of restrictive laws has raised concern among battery manufacturers. "We want to be responsible, but we don't want to make portable rechargeable tools things of the past," insists K. Fred Wehmeyer, chairman of Battery Products Alliance (BPA), an international trade association composed of battery and tool and appliance manufacturers. The BPA has written model legislation of its own, which it will take first to states already considering their own lawmaking and then to federal bodies such as the EPA and Congress. Wehmeyer says that most companies "want consistency of legislation. This state-by-state approach is a marketer's nightmare."

Battery makers favor separating Ni-Cds from waste as part of a community waste-collection program. "We think the municipality has some obligation to collect and maybe sort," states Terry Telzrow, manager of product safety and standards at Eveready Technology Center in Westlake, Ohio. Wagenius retorts: "They want to shift the cost for collection to the user rather than the generator of the product. They want government to absorb the cost, and I say no fair."

Perhaps the biggest obstacle to recycling consumer batteries is cadmium's designation as a hazardous waste. Ni-Cds do not pass the EPA's materials toxicity test, which came into effect in September 1990. The designation means extra, expensive paperwork for government and industrial Ni-Cd users, who are accustomed to recycling batteries from backup energy units in hospitals, marine buoys and other sources. The law makes it more difficult for municipalities who have no such systems to dispose of or recycle spent batteries. "We're reluctant to seek consumer cells [for recycling], simply because of all the permitting and cost. Until it's made easier, it's barely worth bothering," says Derek J. Benham, president of F. W. Hempel & Co. in New York City.

Nor do recyclers want to become marketers of what they reclaim. The Swedish conglomerate NIFE, among the world's largest industrial manufacturers of Ni-Cds, would like to recycle for consumers but has been unable to get batches large enough to be practical, explains Arne O. Nilsson, executive vice president of the NIFE office in Greenville, N.C.

"This is something that the EPA is going to have to get into," Telzrow asserts. The agency could extend the exemption for household hazardous waste, he suggests, so that batteries could be transported for recycling. Industrial batteries might be interpreted as scrap metals. Ni-Cds might also be granted a specific exemption from hazardous-waste designation, as was done for lead-acid batteries on their way to regeneration.

Still another option is to find replacements for cadmium. Matsushida, the parent company of Panasonic, has already begun to test market nickel–nickel hydride batteries in Japan. (It has no immediate plans to introduce these prototypes to the U.S., however.) Metal hydrides—porous compounds that store hydrogen—are promising substitutes for low-power devices such as toys and photoflash units. The new metals will probably not suffice for power tools, because they cannot maintain a constant rate of discharge. The prospect of detoxifying at least some rechargeable batteries is a bright spot. It is one that many would argue simply illuminates the need for recycling. But on whose nickel? —*Deborah Erickson*

Protein Probe

Remote-sensing technique screens bacterial cultures

Spy satellites use spectral imaging to look for camouflage. Other remote-sensing craft use it to track pollutants and to locate mineral deposits. The technique works because of the ability to differentiate between wavelengths of visible and infrared radiation associated with, say, a sand dune in the desert or a tank covered with khaki camouflage.

In the early 1980s, however, Douglas C. Youvan, now an associate professor of chemistry at the Massachusetts Institute of Technology, confronted the absence of just such an imaging technology for identifying the various inhabitants of the small world encompassed by a laboratory culture plate. His eventual goal is to amass data on proteins in a million different mutant strains of photosynthetic bacteria, each of which could potentially be identified by a distinct spectral signature.

Yet at that time, the only instrument available to capture the big picture in a petri dish was a double-beam spectrophotometer. It allowed the spectral characteristics of a single culture of photosynthetic bacteria to be analyzed, a process that took five days. But Youvan needed to examine hundreds of bacterial colonies spread across a dish at one time or else find another line of research. So his need became the mother of an invention called a digital-imaging spectrophotometer, a device he hopes may eventually be used widely to screen not only bacterial proteins but pharmaceuticals—maybe even salmon food as well.

The instrument Youvan devised, like the human eye, combines the spectral resolution of the spectrophotometer with the spatial resolution of a camera. But it is not hindered by the eye's limited spectral resolving power. Youvan and two associates, Adam Arkin and Mary M. Yang, use a video camera to record images of up to 500 colonies on a petri dish. A slide carousel changes Fabry-Perot filters in quick succession to capture as many as 50 distinct spectral images during a 10-minute span. Each 10-nanometer slice recorded by the filtered camera is digitally combined to produce a display of absorbance spectra at wavelengths from the visible to the near infrared.

Simply gathering the data is not enough. Without further processing, the information generated from a single scan would take up 40 million bytes of computer storage. Youvan's group used image-processing algorithms to segregate an area corresponding to each bacterial colony. Spectral values for a colony are calculated by averaging the picture elements for each area, reducing the amount of data to be stored by a factor of 1,000.

A scan is then displayed on a computer screen broken into several windows. In one window, an arrow can be placed on a row of a color contour map that depicts the amount of light absorbed by each colony. The 50 or so segments along the row, each of which corresponds to a 10-nanometer band of spectrum, are colored to represent how much light has been absorbed. Other windows show the location of colonies in a dish or a graph of the composite absorption spectrum for the entire row.

The photosynthetic bacterium Youvan is studying—*Rhodobacter capsulatus*—contains bacteriochlorophyll and carotenoids, pigments that absorb light in the visible and near infrared range. Different absorbance characteristics represent varying protein structures, a spectral signature for every mutant bacteria. "We know that if we could image millions of different events, it would be comparable to an evolutionary time scale," Youvan says. "It would take nature a vast amount of time to search out and test this number of mutants. We will have effectively speeded up the mutational clock."

Youvan believes the technology may be good for looking at more than colored bacteria. By extending the spectral range of the device, it could help to screen the effectiveness of microorganisms in bioremediation for neutralizing pollutants, to identify bioengineered bacteria that produce compounds used as pharmaceuticals or even to judge the amount of a pigment in yeast fed to domesticated salmon to give the fish their characteristic orange hue.

It might also allow the imaging of a microtiter tray with up to 25,000 wells. The conventional tray used for biological assays holds 96 wells, Youvan says. "By increasing the plate density almost 250 times, you will be able to screen the samples hundreds of times more rapidly," he notes.

Youvan has talked to E. I. du Pont de Nemours, Beckman Instruments and Eli Lilly about the technology, although none of these companies have as yet announced firm plans to commercialize it. The interest is there, however. "What you can do is an order of magnitude bigger than what you can do by classical measurements," observes Jeff Quint, a research biochemist for Beckman, a manufacturer of instrumentation for the life sciences. —*Gary Stix*

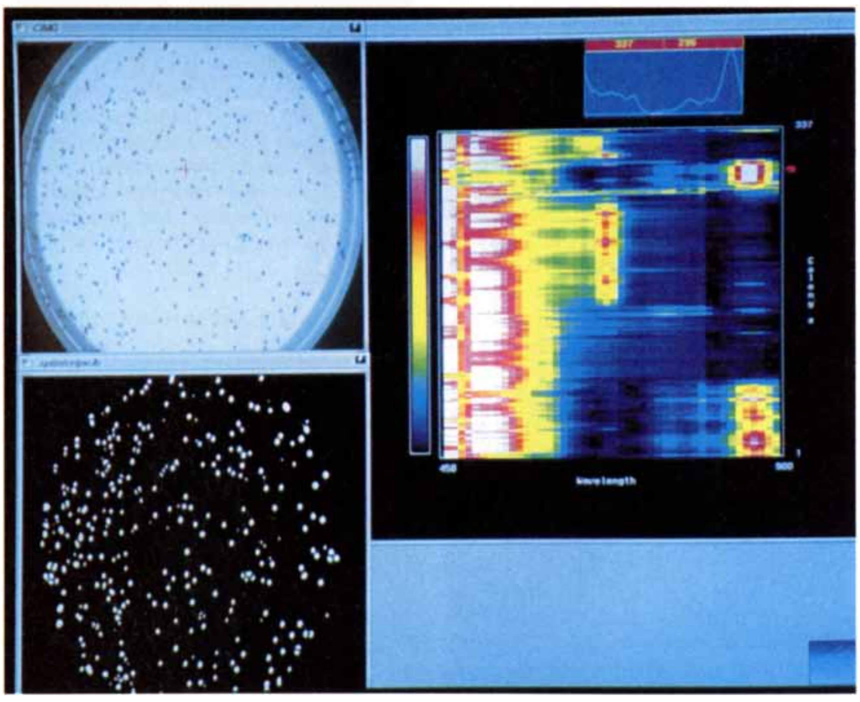


IMAGE PROCESSING of several hundred colonies of mutant photosynthetic bacteria on a petri dish shows the location of the bacterial colonies* (upper and lower left). *The spectral data* (right) *for each colony is represented as a row* (indicated by the red arrow) *within a color contour map. A graph of the absorption spectrum for the row is shown above the map. Source: Douglas C. Youvan.

Open Channels

Hormone derivatives may combat PMS and epilepsy

When is a sex hormone not a sex hormone? In the 1950s researchers observed that epileptic women had fewer seizures during menstruation, when progesterone levels are on the wane. No one knew why until the 1980s. When the hormone is no longer needed by the body, it breaks down into metabolites that do something completely different. For 10 minutes or so, the former sex hormones become steroids that reduce neural excitation.

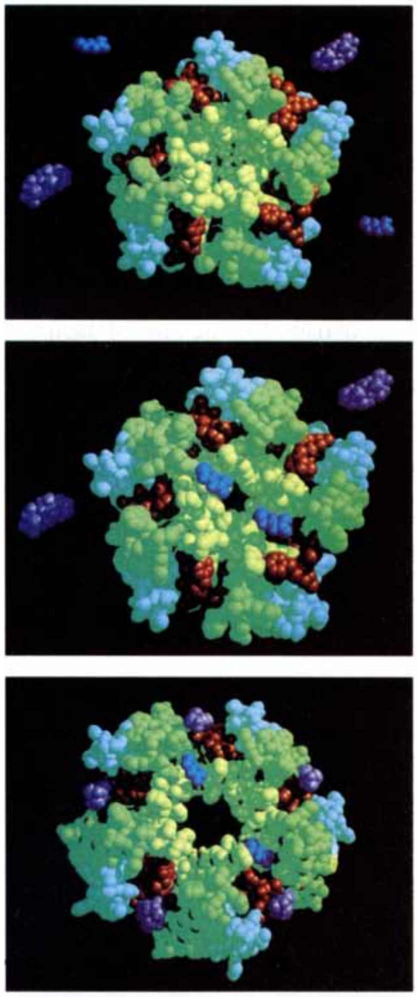
Researchers did not realize they had discovered a class of natural allosteric modulators until 1987. These "other site" substances enhance the function of a nerve cell receptor by binding to it at a place distinct from that taken by a neurotransmitter. The new drugs are called epalon compounds by their developer, CoCensys, Inc. The start-up company in Menlo Park, Calif., believes that by tailoring the compounds to fit specific subunits of the receptor, it will be able to tease out quite different neurological responses.

One modification might produce a "soft" drug to relieve the tension and irritability of premenstrual syndrome (PMS) without causing drowsiness. Another version might induce sleep, and still another could stop the convulsions of epilepsy. CoCensys increases the hormonal metabolite's half-life with a chemical extension to the part of the molecule the body degrades first.

Epalon compounds bind to the receptor complex for a neurotransmitter called GABA. GABA is but one of many such chemical messengers, each with its own receptors throughout the central nervous system. But unlike most neurotransmitters, such as serotonin and the catecholamines, which are primarily excitatory, GABA is mainly an inhibitor.

Under normal physiological states, GABA binds to its receptor, thereby briefly opening chloride channels in the neuron's outer membrane. Negatively charged chloride ions flow in, rendering the cell less prone to excitation typically caused by movement of positive ions. This instant of relaxation—measured in milliseconds—resets nerve cells so they may fire again. When chloride channels clench shut, this produces feelings of anxiety. Seizures result when many channels close in enough brain regions.

Since the discovery of GABA, drugmakers have searched for chemicals that can mimic its function but do so for a prolonged period. Anticonvulsants such as the benzodiazepines help GABA maintain open channels by binding the receptor complex, as do barbiturates. Epalons also work this way but appear to bind different places.



MOLECULAR MODEL of a partial receptor complex for GABA, an inhibitory neurotransmitter, shows closed chloride channel **(top).** *Channel opens when GABA* **(blue)** *binds this receptor in a nerve cell membrane* **(middle).** *Progesterone derivatives* **(purple)** *called epalon compounds also bind the receptor* **(bottom),** *enhancing channel opening.*

CoCensys was not the first to identify progesterone metabolites. In 1986 National Institute of Mental Health (NIMH) researchers Steven M. Paul and Maria D. Majewska reported that the derivatives would bind—albeit not tightly—to GABA receptor complexes. The team suspected the drugs would behave like barbiturates.

This announcement did not go unnoticed by Kelvin Gee, an associate professor of pharmacology at the University of Southern California School of Pharmacy, who was also working on GABA receptors. Gee decided to see what the progesterone metabolites did when exposed to the receptor along with physiological concentrations of the neurotransmitter that acts on it. He confirmed that the derivatives bound to the GABA receptor, but now much tighter, and to a site other than the one recognized by barbiturates. A quick test on mice bore out Gee's hunch—the compounds prevented convulsions. He filed for patents on the derivatives in 1987.

That was almost the end of the story. "Most pharmaceutical companies were scared away because they were steroids," Gee says. "People have this preconceived notion that anything that's a steroid is bad." But the derivatives have no affinity for hormonal receptors, just as hormonal steroids have no effect on the GABA complex. "It's like taking a wheel off a tricycle," says Robert G. McNeil, the venture capitalist who proved willing to take the risk. (McNeil functions as CoCensys's president from his offices at Sanderling Ventures. The research is contracted out to founding scientists at U.S.C.)

There is much more research to do before epalons can be commercialized. Scientists at CoCensys and elsewhere are trying to determine what the receptor complex looks like in fine detail. So far they have identified five subunits, dubbed alpha through epsilon. CoCensys co-founder Michael Bolger observes, "It is likely that in different brain regions, different forms of the GABA receptor complex will predominate, each made up of different subunits."

This promise of diversity sets drugmakers to dreaming. A compound that bound a subunit combination most prevalent in the cerebellum, for example, which controls voluntary muscular movement, could bring about much different effects from a drug tailored to fit a subunit cluster in the brain stem, which regulates functions such as heart rate and breathing. The specificity might well reduce side effects.

Instead of trying to isolate receptor subunits from brain tissue, Nancy C. Lan, another CoCensys founder, is making them by genetic engineering. The company will test modified versions of its epalon compounds against combinations of subunits to see which fit best.

Still, safety will have to be proved with a great deal of rigor before the Food and Drug Administration will consent to clinical tests for a PMS drug to be taken by otherwise healthy women. So CoCensys will focus initially on West syndrome, a type of childhood epilepsy. If epalons work as expected, they are likely to find wide-open markets. —*Deborah Erickson*

Bursting bubbles

I have long been struck by the fact, and puzzled by it too, that in all the arsenal of economic theory we have absolutely no way of predicting how long such a [bubble] will last. To say that prices will fall back to earth after they reach ridiculous heights represents a safe but empty prediction.

—PAUL A. SAMUELSON, 1957

Pick up almost any Sunday newspaper in the U.S. and a special section devoted to real estate advertisements practically falls out. Scan the indices of standard economics textbooks, on the other hand, and real estate seldom surfaces. An economic fact of life for 64 percent of America, the buying and selling of homes has received scant attention from economists.

Why the indifference? In the past, neither macroeconomists nor microeconomists could divine a specific reason for watching real estate sales. From the vantage of macroeconomists, other factors—namely capital, labor and technology—drive the economy. Real estate has largely provoked yawns. House sales do not show up in any indicators of economic performance. (The value of homes is also not included in measures of national savings.)

Microeconomists have been uninterested in real estate sales as well. They study consumers' decisions to buy goods in the context of a market. Since one efficient market ought to function much like any other, the same economic principles should describe decisions to buy a house or a car.

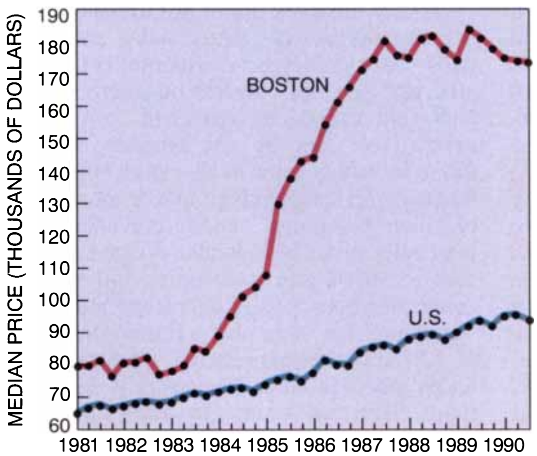
These observations, however, are proving to be wrong. Economists who do scrutinize real estate sales are concluding that such markets are far from efficient. One glaring sign of the market's inefficiencies may be "bubbles," or rapid escalations in prices, followed by equally rapid declines. Moreover, these economists believe that real estate bubbles puffed up by consumers' wishful thinking can cause gross misallocations of resources, ultimately crippling a local or regional economy.

"This has been controversial stuff, but it's getting less controversial by the minute," asserts Karl E. Case, an economist at Wellesley College. More than five years ago Case began analyzing inflation in home prices. Data showed that in 1985 the median sale prices of single-family homes in the Boston area jumped 38 percent. No fundamental factors—such as changes in population, employment or income—that Case considered could account for the enormous increase. So he speculated that the rise in prices was a bubble driven by home buyers' rosy expectations.

To test this hypothesis, Case and a colleague, Robert J. Shiller of Yale University, plunged into a more detailed study of real estate prices. In spite of the Sunday papers, there is a dearth of reliable data on home sales. The U.S. Bureau of Labor Statistics even gave up calculating an index of changes in home prices because the data were plagued by inconsistencies.

Case and Shiller skirted these problems by constructing a new index of house prices. Their "weighted, repeat sales" index measures the changes in the prices of homes sold at least twice during a given time. This avoids the erratic shifts in mean house prices created by lumping together sales of uptown penthouses and downtown walk-ups.

The Boston Home Price Bubble



SOURCE: National Association of Realtors

Applying the index to data on home sales in four U.S. cities between 1970 and 1986, Case and Shiller found significant excess returns from buying and selling a house—profits above the return earned on other investments such as Treasury bills. Furthermore, these excess returns were predictable, leading to a simple rule for making money in these markets: if the returns predicted by the index were higher than the current mean return, buy immediately. If not, wait a year. The rule lent more weight to the argument that real estate booms are driven by people's expectations in addition to fundamental values.

To further unravel what fuels expectations, Case and Shiller surveyed about 1,000 home buyers in four other cities. Most saw buying a home as an investment. But their reasons for expecting prices to rise were largely anecdotal and not rooted in serious assessments of why prices change. "Among the most popular clichés were 'The region is a good place to live,' and 'There is not enough land,'" Case reports. "Neither of these is news, and neither could explain a sudden boom."

Their conclusion—that in a real estate boom, prices race ahead of what sound judgment dictates—is more than academically interesting. "It means there's been a terrible misallocation of resources," Case says. Banks suffer because they lend more money than the assets are worth. Homeowners profit, but those who do not already own watch their chances of buying a home wither. When the bubble bursts, the economy loses "a ton of money," Case declares.

Case has further argued that the Massachusetts economy was pushed into its current recession by the real estate boom that lasted from 1984 to 1987. Over that time, some $100 billion in real estate equity was "created" in Boston by the rising prices of single-family residences. In response to the demand for housing, construction went ahead at a feverish pace. When the boom ended, tens of thousands lost their jobs. In February unemployment in Massachusetts stood at about 9.3 percent. Case believes that in the absence of the boom, the economy would have slowed but not reached the "potentially catastrophic recession" it is now experiencing.

Real estate may not be the only inefficient market plagued by destructive bubbles, Shiller points out. "People talk about inefficient markets, but they don't do research on them," he says. Shiller claims to have found the spoor of inefficiency in financial markets as well. Nevertheless, he adds, bubbles have been an unfashionable subject of study because of economists' deeply held reliance on efficient markets.

As a result, economists have yet to sort out what initiates real estate or other market bubbles and why they burst. But, on the other hand, airing the prevailing theories might provide a salve: at least in principle, the more people who know the market is inefficient, the fewer should jump on the speculative bandwagon. Pass it along.

—Elizabeth Corcoran and Paul Wallich

MATHEMATICAL RECREATIONS

The theory of rigidity, or how to brace yourself against unlikely accidents

by A. K. Dewdney

It helps to be flexible when you think about rigidity. I learned this lesson in the summer of 1978 as my father, my son, Jonathan, and I fixed up our cabin in the Canadian North. To patch the leaky roof, my father had built a scaffold from freshly cut spruce poles. When my father and Jonathan climbed to the top of the scaffold, the rustic framework groaned and swayed. I mentioned that the scaffold looked a little shaky, but my father scoffed, "Why this thing will hold 10 men—and I used the absolute minimum number of poles."

Who was I to argue with my father, an expert woodsman and an amateur mathematician to boot? I returned to my chores inside the cabin. Less than a minute later I heard a whoosh, a thump and two startled cries. Racing outside, I found Jonathan and my father sprawled on the moss. The scaffolding had scaf-*folded,* so to speak. The two stood up, and my father grinned sheepishly, exclaiming, "Isn't that the damnedest thing!"

I can hardly blame my father for building an unstable scaffold. Mathematicians and engineers have been struggling for centuries with the theory and practice of constructing rigid frameworks. Mathematicians call the subject rigidity theory. I recently investigated this topic, hoping that a few insights might save my family and others from further injury. My research has also uncovered a host of amusing puzzles to flex the mind.

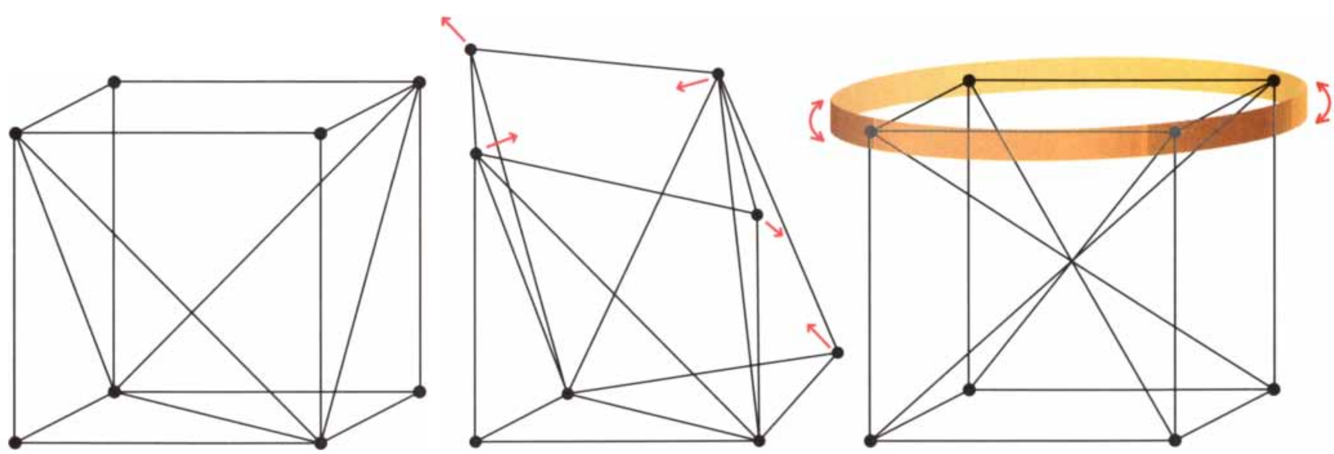
Rigidity theorists prefer not to make frameworks out of spruce poles and nails. Instead they have a mental construction set that consists of abstract bars that cannot be stretched, compressed or bent by any amount of force. Such bars come in all conceivable lengths, and if the ends of two or more of them touch, an instant universal joint is formed. The joint allows the two bars to swivel and twist unless other connecting bars constrain their motion.

Imagine, for example, a framework of 12 bars of equal lengths arranged into a cube. The cubic framework is not rigid. Placed on a table, it would flop over in an instant. Indeed, if such a framework were rigid, bridges and towers would not need diagonal girders.

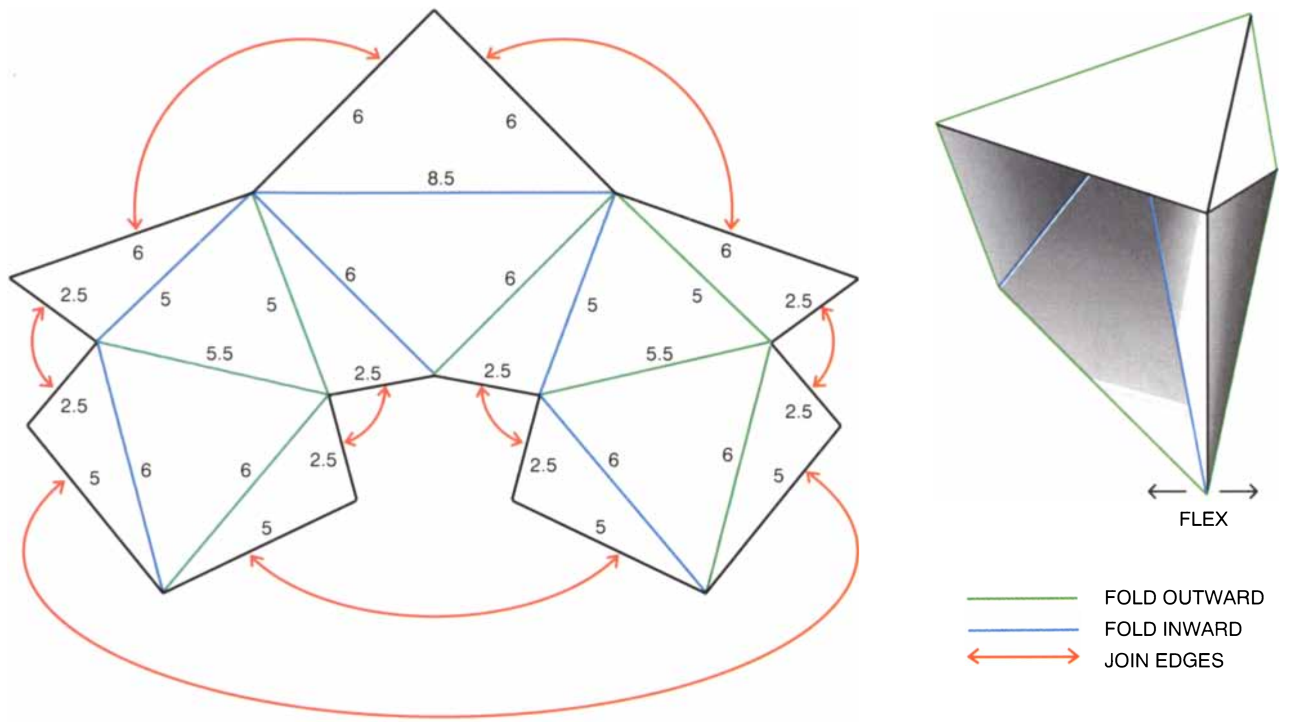
Recently I attempted to brace a cube by adding diagonal bars to some of its six square faces. How many bars would it take to make the cube rigid? Four bars, judiciously placed, seemed enough until I found a way to make the cube flex. Even a cube that had diagonal bars added to five of its faces turned out not to be rigid [*see illustration below*]. The five bars, along with six of those already present in the cube, form two tetrahedrons that are hinged along a common bar. If the two tetrahedrons are folded against each other, two of the four joints that define the unbraced square face move toward each other as in the illustration. The other two joints move outward. No matter how five diagonal bars are added to the faces of a cube, there will always be a way to flex it. No fewer than six bars are needed.

Instead of bracing a cube on its faces, what if it were braced by diagonal bars that run from one joint right through the center of the cube to the opposite joint? (With the careless élan of theorists, readers may ignore the intersections of the diagonals.) A cube braced by four interior diagonals has a strange kind of flexibility that theorists call an infinitesimal flex. In some sense, an infinitesimal flex is a motion of one part of a framework relative to another. The motion is so small, however, that it does not even exist.

Let me explain. The diagonally braced cube shown in the illustration below has arrows that indicate a tiny rotation of the top face in relation to the bottom face. Because all bars making up the cube are made of ideal materials that will not suffer the slightest change in their length, the top face cannot be truly rotated, even by a tiny amount. Yet one may *start* to rotate the top face and the bottom face in opposite directions. During this vanishingly tiny moment, there is no resistance from any other part of the cube because all bars that connect the upper face to the low-



The braced cube at the left has an ordinary flex* (center)*, whereas the one at the right has an infinitesimal flex



How to construct a Connelly-Steffen surface

er one make right angles with the direction of rotation.

If this diagonally braced cube were made of real materials, it would be distinctly vulnerable to small but measurable rotations. The structure would wobble. (My father avoided this particular style of bracing.) Frameworks that have only infinitesimal flexes are considered rigid, but those that have no flexibility whatsoever are called infinitesimally rigid.

Besides their mental construction sets, rigidity theorists also have a mental tool kit containing a great many theorems and techniques that can be applied, among other things, to bracing a cube. One of the simplest and most effective tools was discovered by 19th-century engineers. A framework that has J joints must have at least $3J-6$ bars to be infinitesimally rigid. This theorem can be applied to the cube; its eight joints mean that $J = 8$. The corresponding magic number computed by the formula is $(3 \times 8) - 6 = 18$.

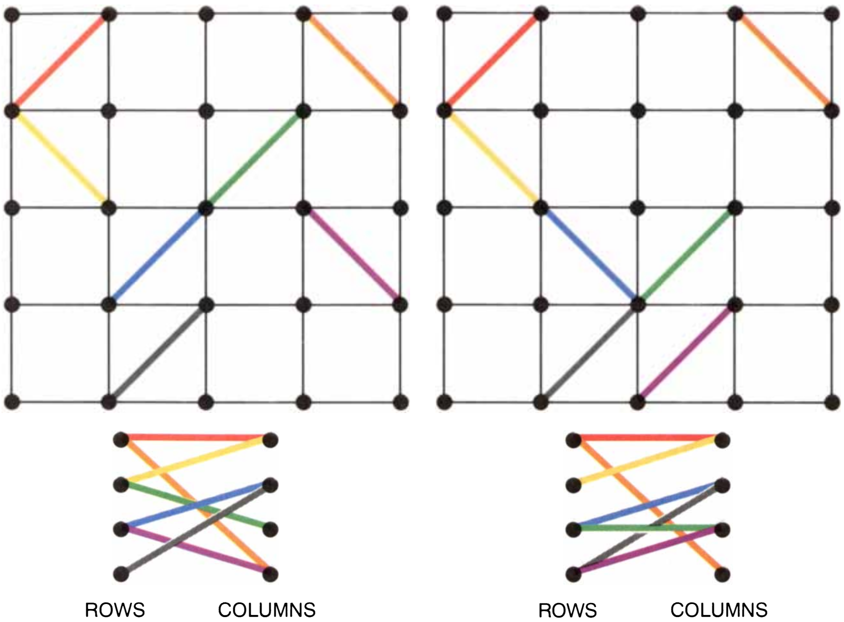
To show that a cube composed of 18 bars (12 edges and six braces) is actually infinitesimally rigid, one might appeal to a theorem invented by the Russian geometer A. D. Alexandrov in the 1940s. Alexandrov studied rigidity in frameworks based on a convex polyhedron. These faceted surfaces include everything from cubes to cut gems to the geodesic domes of R. Buckminster Fuller. Alexandrov proved that any framework based on one of these shapes can be made infinitesimally rigid by adding bars to the framework so that every face is composed of triangles. As far as Alexandrov's theorem is concerned, then, a triangular bracing of each face of the cube (one bar each) will make it strong.

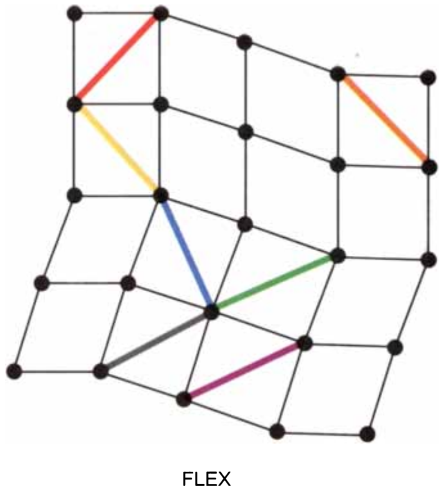
I sympathize with any readers who have problems visualizing the cubic bracings. Even the diagrams on the opposite page are a bit complicated. Perhaps it is time to descend from the three-dimensional space that gave birth to the theory down to the plane, a two-dimensional space inhabited by a vast panoply of various flat frameworks. Although readers can easily figure out that a square can be made rigid with a single diagonal, they will find it rather challenging to figure out how to brace a grid of squares. For example, how many diagonals must be added to make a four-by-four grid of squares immune to flexes? The illustration on the next page shows two ways to brace such a grid with only seven diagonals. But one of the braced grids is not rigid. Can readers tell which one?

The answer can be deduced in the following manner. Make up a diagram composed of two sets of dots. The first set represents the four rows of the grid, one dot per row. Likewise, the second set corresponds to the four columns. For each of the seven diagonal bars in the grid, connect the appropriate dots in the diagram. For example, if there is a diagonal bar in the square situated in the third row and the fourth column, then draw a line from the third dot in the first set to the fourth dot in the second set.

Whether the grid is now rigid can be answered by asking the following question: Is the dot diagram connected? In other words, is there a continuous path from any dot in the diagram to any other? If (and only if) so, the grid is rigid. This elegant theorem—first proved by Henry Crapo of the INRIA near Paris and Ethan D. Bolker of the University of Massachusetts at Boston—can help readers quickly determine whether a grid will flex. The diagram for the grid on the left is connected, but the other is not. As an exercise in rigid thinking, I will leave readers with the problem of using the Crapo-Bolker theorem to decide why seven is the minimum number of bars necessary to brace the grid. As far as I know, there is no corresponding theory to advise readers, or my father, about how to brace scaffolds or other cubic grids.

Sometimes the search for rigidity requires flexibility in the literal sense. No story better illustrates the point than the history of the famed rigidity conjecture. In the 17th century the French mathematician Augustin-Louis Cauchy wondered whether all convex polyhedral surfaces were rigid. Such surfaces include the triangulated polyhedrons of





***Dot diagrams reveal whether a grid is rigid* (left) *or not* (center *and* right)**

Alexandrov's theorem and many more. Their facets, or faces, are bounded by plane polygons with any number of sides. Being convex, they have no indentations or hollows of any kind. In 1813 Cauchy proved that a convex, polyhedral surface is rigid if all its faces are triangles. The theorem meant that any convex surface one could construct from triangles, each triangle sharing each of its bars with one other triangle, would be rigid.

Despite the restriction of Cauchy's theorem—that the surface be convex—mathematicians were beginning to wonder whether all surfaces composed of triangles were rigid—even those surfaces that were not convex. Such surfaces may appear to be folded, twisted or contorted in quite crazy ways. The only requirement was that they be simple in the topological sense. If suddenly converted to rubber and inflated, they must be (more or less) spherical. Additionally, a simple surface required that no part of it touch another part of the same surface. Mathematicians conjectured that if a surface had all these properties, then no matter how deformed it happened to be, a version composed of triangles would suffer no flexes.

For more than a 100 years, no one was able to prove this so-called rigidity conjecture, nor could anyone disprove the conjecture by finding a flexible, nonconvex surface made of triangles. The strongest supporting evidence for the conjecture came in 1974, when Herman R. Gluck of the University of Pennsylvania showed that "almost all" such surfaces were rigid. In other words, examples counter to the conjecture, if they existed, would be rare indeed. Even a contrarian would find this much evidence in favor of a conjecture discouraging.

But Robert Connelly of Cornell University was convinced in some corner of his being that the rigidity conjecture was false. After visualizing surface after surface that looked as though it should flex, Connelly realized one day that he was working against Gluck's theorem. His office was full of models sent to him by amateur mathematicians who claimed flexibility for them. Gluck's theorem said, in effect, "Not likely!" Faced with the same difficulty, Connelly decided to examine mechanisms, namely, frameworks that he knew would flex.

Starting with a very simple flexible framework, he employed his knowledge of topology, spanning parts of the framework with simple triangles. Then one day he felt close. Before him was a nonconvex surface that flexed. But it was not quite what topologists call a sphere. Two edges within the surface touched each other, like a deflated basketball in which one side is pressed against the other. The thing was distinctly annoying. So near and yet so far.

It was then that the idea of a crinkle came to him. He suddenly thought of a way to introduce a subdivision of the annoying edges and surrounding triangles that amounted to a fold—enough to take the two lines out of contact. The model he built flexed!

The counterexample to the rigidity conjecture appeared in the literature in 1978. Shortly after, the German mathematician Klaus Steffen found an even simpler surface, based on Connelly's idea, that flexed. Readers who would like to flex their own version of the Connelly-Steffen surface will find it laid out in the illustration on the preceding page. To obtain a size that is easy to work with, readers should interpret the edge numbers as centimeter lengths. Arrows that connect the edges in pairs indicate attachments to be completed by armchair rigidity theorists.

When the Connelly-Steffen surface is completed, the two central triangles make a fold by which one hand may grasp the surface from above. With the other hand, it will be possible to reach up under the model and then (delicately!) to flex the bottom vertex from side to side, but only by a small amount, roughly 10 degrees.

When this tiny flex is performed, the surface bounds the same volume. These days Connelly ponders whether the constant volume property holds true for all flexible, nonconvex surfaces made of triangles. If he conjectures that they do, he himself may have to be flexible. Some young upstart may find a counterexample.

As something of an upstart myself, I gave my father some trouble over the collapse of the scaffolding. But within a few hours of the accident, the scaffolding was up again. It was identical to the previous structure, except for one extra spruce pole. My father climbed the scaffold confidently. I am sure the tiny wobbles I detected were merely infinitesimal flexes.

FURTHER READING

CONNECTIONS: THE GEOMETRIC BRIDGE BETWEEN ART AND SCIENCE. Jay Kappraff. McGraw-Hill, 1991.

BOOKS

Reading the earth's strata, something new under the sun, a possible rival for chocolate

by Philip Morrison

A GEOLOGIC TIME SCALE 1989, by W. Brian Harland, Richard L. Armstrong, Allan V. Cox, Lorraine E. Craig, Alan G. Smith and David G. Smith. Cambridge University Press, 1990 ($49.50; paperbound, $19.95). **A GEOLOGIC TIME SCALE 1989: WALLCHART,** by W. B. Harland, R. L. Armstrong, L. E. Craig, A. G. Smith and D. G. Smith. In color, 22 by 40 inches. Cambridge University Press, 1990 (folded or rolled, $12.95).

A finger's breadth of clay parts two thick rock strata in the Umbrian hills; below it lie the rocks called Maastrichtian, uppermost formation of the Cretaceous, above it the Danian, lowest of the Tertiary. By now every scientific reader knows of that abrupt end to a long chapter in the Book of Sediments, an episode judged by many to record celestial blows that ended the reign of the reptiles. That KT boundary is here dated 66 million years ago, plus or minus about two million years (adjusted for overall consistency, the adopted age is 65 million years). But just how do we know that chronology?

This volume, a lucid summary and a technical work of reference as well, documents a modern time scale for the whole of geologic history. The task has a dual nature that flavors the whole. What is dated—the dependent variable, one might say, and a formidably varied one—is some 140 or 150 stages (once called formations), found and distinguished in the rocks worldwide mostly by the field geologists of the 19th century. This generates a wealth of taxonomic and historical appraisal. One geochronological stage is simply about as small a column of rock as experience has shown can be treated as a single unit. That varies with the case; for crude guidance, imagine a typical stage as a pile of rock layers the height of a high cliff.

Few boundaries are marked by crisp bookmarks like that clay at Gubbio, "golden spikes." Most stage boundaries have been related by 20th-century work to the specific organic forms found in the rock, macrofossil and microfossil species and their population counts. This is like ordering the scattered and incomplete pages of an old manuscript by making sure that the joined text fragments make sense. Then consensus is sought to partition that shifting evolutionary continuity into discrete stages and to label those by an international nomenclature. New fossils are being found; forms once characteristic may no longer seem unique; discordant assignments lead to boundary shifts. Many stage boundaries are settled by the judgments of expert committees. (The authors show a little pique about that famous discontinuity in clay. "Along with much publicity for catastrophic events," they report, the official committee has yet to recommend formally its reference point for that key boundary.)

If we describe the KT boundary in the way used for most stages, we would stipulate only that it overlies the rock of one specific quarry in the chalk of South Holland but stays below the chalky limestone present at a second classical locality in Denmark. The abundant fossil plankton seen in those two limy beds are quite distinct.

Stages once arranged in sequence can be dated, but only through some token of the uniform flow of physical time, the grandest of independent variables. Today the time label is radioactive decay within the rock, independent of geologic history and of living forms, indifferent to terrestrial environment. It is something of a surprise to see how many dates here depend on one single instance of the alchemical change, the slow beta decay of the rare potassium isotope of mass 40 into stable argon 40. This change has been measured in the right samples for a generation by direct mass-spectrographic analysis for those two isotopes. The right samples are mostly unweathered crystals of black mica, found in the lavas and tuffs and other volcanic inclusions that lie now and again amid the layered sediments of land or seafloor. That mica has plenty of potassium; once formed, its crystals hold argon safe against diffusive leak at "closure temperatures," about that of an extremely hot oven. The volcanologists and mineralogists have certified how little difference that will make under the cooling conditions of volcanic flows.

For 20 pages, the isotopic data base reports a few hundred samples, every sample tied to a date. The authors have drawn the numerical data mainly from careful compilations published during the past 25 years. They critically examine enough measurements of other mineral types and of other isotopes to dispel any sense of fragility in depending heavily on a single superior method.

Now the stages stand in agreed sequence, with plenty to be said of the history and reliability of most of the choices. For each date, there is a statistical estimate of counting error. If there were one well-known date fixed at the midpoint of each stage, the matter would be straightforward. Utopian! Dates with different errors are scattered in the rocks, sometimes to date different places within one stage, sometimes skipping several stages. Dates for lower stages are sometimes younger than those for stages that lie higher in the rocks.

Out of this pile of troubles the authors have devised a compelling visual means of displaying errors quantitatively. Plotted against a trial date, they compute a measure of error, on the model of least squares. It would ideally generate a symmetric curve, a valley whose lowest point would touch the time axis at the most likely date, the error of fit fixed by the valley width. You can examine here some 120 of these curves, boundary by boundary; the curves are twisted, asymmetric, sometimes missing the axis, sometimes cut off by it. The final case made is strong. Stage after stage can hardly be in error by more than a few million years each way. Considering the entire sequence allows even better judgments, amply presented and documented. The uncertainty in the fossil sequences is not forgotten, and the linear progress of seafloor spreading with time aids the interpolation at awkward places in the long list.

With increased depth in time, rocks and their data points grow scarcer. The "stages" become longer, less secure than the one to five million years that span the stage durations for the upper three fourths of the stages. Below 570 million years, in the Precambrian, the grand clues of visibly intricate organic form fail almost completely. There, the relics of life are rarer, microscopic, their complexity biochemical, well hidden from the tweezers and microscopes of the paleontologists. Before that time, atomic dates alone rule, and the classical rock stages must give way to broader divisions bounded by round numbers in atomic time.

Back, back we peer, to the start of the scale in the "Cryptic division." That apt name arises out of varied clues from the earliest history of the forming moon, 4.5 billion years ago, older than the oldest-known rocks on the earth by 500 million years, and we are at the lowest of rock-based horizons.

A companion wall chart reproduces the concluding tables of the rich text, with color added. One panel sums up the results. Along the three-and-a-half-foot length of the chart, the whole of the earth's geologic time is plotted, not at the detail of the stage but only by period. Two other panels fill out the space with zoom views. One extends stage by stage from the first metazoans to *Homo erectus;* the other expands the past two million years, *Australopithecus robustus* to Cro-Magnon. The elegant austerity of the time scale itself is at the end ornamented to everyone's pleasure by the entry of a couple of hundred varied geologic events, authors' choice, without references or justification: early grasses, Antarctic glaciation, primates (before the rodents), India-Eurasia collision. This team from Canada, the U.K. and the U.S. is fully as scrupulous, learned and precise as calibrators should be; they are uncommonly good-humored besides. It is sad to report that co-author Allan V. Cox of Stanford University died before the book went to press.

THE ELEMENTS BEYOND URANIUM, by Glenn T. Seaborg and Walter D. Loveland. John Wiley & Sons, Inc., 1990 ($49.95).

When in 1869 Dimitri Mendeleev published his first periodic table of the elements, it ended with uranium, the heaviest atom in nature, known somewhat vaguely as a constituent of Bohemian pitchblende since the 1780s. Not until 1940 was the first transuranic element found by two Berkeley physicists, later named neatly enough after the first planet beyond Uranus. The next higher element was already implied by the instability of that first find, but plutonium was not securely identified until the first months of 1941. That identification remained secret until August 1945. The unprecedented delay signals the fateful coupling between the ends of the periodic table and the chronicle of our warlike century. Scientific curiosity began the study of the transuranics, but war and the fear of war enabled and enlarged it. The first visible speck of any artificial element was a microgram or so of plutonium isolated in August 1942; plutonium is now stored by the hundred ton. With that change in scale by 14 orders of magnitude runs a deep fault in history where a slip might shatter the nations.

This concise 50th-anniversary summary is a book with mixed aims; its personal accounts—the senior writer was co-author of the plutonium discovery paper and a Berkeley participant in more than half of the discoveries of transuranics since then—and its wide scope can well serve the general reader. Much of the text is a kind of compact handbook, presenting quantitatively if briefly the rich nuclear data and detailed theory of our day. For instance, the properties of some 160 isotopes of the 17 known transuranic elements, now up to element 109, are tabulated (not all elements bear official names because the act of discovery is contested). Although the plutonium in the bombs is hidden and remote, one transuranic atom, americium, available by the kilogram, is now in wide public use. In tiny sealed samples it is the source of ion current within millions of smoke detectors.

Virtuosity is commonplace among these investigators: they carry out multiple chemical separations from moving streams of hot gas within seconds; they can separate and detect in vacuum new radioactive nuclei by their mass and charge, provided their lifetime is at least something like a millisecond; they have confidently analyzed samples with only tens of atoms, often all of them on the surface of a single tiny bead of absorbent resin.

Time is the arbiter of elements. Uranium is simply the heaviest atom that could endure for the billions of years it took to evolve chemists. Only one of all the transuranic species known has a mean life as long as a geologic era, about 100 million years; that one, ^{244}Pu, cannot be extinct, "although the total amount... on earth is less than 10 g." We know it was part of what our planet inherited from interstellar gases, for isotopes and tracks characteristic of its fission are found in overabundance in certain mineral grains of meteorites. In those yeasty times, solar matter was more radioactive.

The transuranics are genuinely artifacts, something new under the sun. They were by no means new to the ancestral stars; even today, somewhere else in the Milky Way, in the debris of some mass-shedding star, they are synthesized anew. Several intricate graphs in the book make dramatic this branch of applied astrophysics. In the first thermonuclear explosion at Eniwetok in 1952, when one islet of the atoll necklace became vapor, the internal flood of neutrons reached unheard of heights. In nanoseconds, uranium nuclei captured neutron upon neutron to form isotopes in measurable amounts all the way from ^{239}U up to mass number 255. Those quickly decayed, to produce a swath of transuranic species from uranium up to element 100, first isolated from that bomb debris and named fermium.

Other graphs clarify the expected run of successive neutron captures expected somewhere in dying stars. Their mark is seen in the distribution of heavy element abundances on the earth, when star fallout, long decayed, coagulated into the solar nebula. The calculations fit what we find; the star-bomb debris held highly transient elements up to 112, never seen yet under the sun. The physics of this field centers on the energy of nuclear droplets. They are held together by short-range nuclear attractions and pushed apart by long-range electrostatic repulsions that act among protons alone. There is no hope for an element with more than about 125 protons: too much mutual repulsion, its center cannot hold. Increasing the neutron number, whether in bomb, star or high-powered reactor, is the wholesale path to transuranics. Accelerator bombardment of a well-chosen heavy target nucleus with a fast neutron-rich heavy ion, say, ^{39}Ca, is a more controlled method and has claimed most recent successes.

It has turned out that the model of a charged nuclear fluid at fixed density is not accurate enough to predict the processes of fission and alpha decay that usually end the heaviest transuranics. The energy of a nucleus is described by such a fluid to within about 0.5 percent. But prediction of the path of a reaction depends on small differences, and favorable arrangements of small subsets of particles, shells, must in fact be allowed for. It was always evident that the decay of heavy nuclei much favored losing four nucleons all at once, the two pairs so well bound into an alpha particle.

The modeling is finicky, but it seems now that the ridge of relative stability (we no longer expect any superheavies with lifetimes of geologic duration) runs unbroken from where we are now out toward the element 120, the nuclei steadily decreasing in lifetime. In all, some 500 species of transuranics are expected with lifetimes greater than a microsecond. We might hope to form a superheavy nucleus gently, never so overexciting it that it flies apart. Meanwhile we can admire the neutron stars, superheavy balls of nuclear matter, stable only because they are bound at long range by the fourth force, gravity.

Advances in New Crops. Edited by Jules Janick and James E. Simon. Portland, Ore., Timber Press, 1990 ($65).

The black seeds, set like jewels in the green slices of kiwifruit, came out of China to many countries before 1905. Only in New Zealand did nurserymen plant and select for new crops. By 1940 they had bred several commercial varieties; about 1950 a clever marketer in Los Angeles encountered the fruit and became its champion. The distant growers learned how to wrap and store their furry delicacy. Once called Chinese gooseberry, the kiwi was given its market name. The New Zealand acreage in kiwi vines increased by a few hundred times; during the 1960s their fruits sold in the U.S., Europe and Japan for 79 cents apiece.

Some Californian growers near the plant introduction station in Chico, where long before good kiwi vines had come, began to pick fruit about 1967. A Growers Organization was formed to improve standards and to nurture the U.S. market. By 1985 California kiwi "were in surplus and the boom was over," with production at almost 30,000 tons a year. New Zealand, the first to finish the course, saturated their market in 40 years; Chile, the latest to enter, will make it in about 12 years. France, Japan and Italy came between. "Any country can take a new crop and overproduce it in less than 20 years."

Decorative kiwi is hardly a staple, even on the trendiest of tables. But the soybean has been a cornerstone of human nutrition in East Asia for 3,000 years. That bland, white proteinaceous curd puzzled naive European travelers, who saw no dairy farms at all. In 1665 observant Friar Navarrete understood: "They drew the milk out of the Kidney-Beans and...make great Cakes of it like Cheeses."

Soybean has been grown on and off as a forage crop in the U.S. ever since its introduction near Savannah, Ga., in 1765. The new agricultural field stations began to promote it during the 1880s, with mixed response, until in

1888 it was found that all legumes fixed nitrogen in the soil through the microflora of their root nodules. Soon W. P. Brooks showed how to encourage soybean by planting it along with a pinchful of soil dust from previous roots. The new crop was truly rooted in America: by the 1920s the crushed seed was a source of high-protein feed.

The World War II shortage of edible oils and feed induced a great expansion. Between 1924 and 1984 the U.S. crop of soybean, largely exported, had increased 400-fold. Crop restrictions on corn and wheat favored soybean, which now earns more for U.S. farmers than any other (legal) crop save those two cereals. The whole Western world still eats soybean mainly at second hand, via its poultry, pigs and cattle; in Asia they frugally avoid middle creatures and enjoy their tofu at considerable saving.

This thick volume reports the First National Symposium on New Crops, held in Indianapolis in the fall of 1988, with many sponsors. Some 30 papers treat systematic topics, such as plant breeding, biotechnology and crop evaluation. About 120 papers discuss specific crops, grouped as cereals, oilseeds, fiber and industrial crops, floral crops, vegetables, fruits and nuts, spices, medicinals, even space crops. (Cowpeas and garden peas appeal to space gardeners, because they are self-pollinating and offer a variety of uses of leaf and seed.)

Starchy Andean tubers like the potato, but in many forms and colors, protein-rich grains amaranth and quinoa, the rye-wheat hybrid triticale, a success on marginal lands in Eastern Europe, and one perennial relative of wheat—the best among 100 grass species for a new approach to the husbandry of cereals—are all surveyed here, in place and in promise. None are yet wonder crops. A long list of tropical fruits and vegetables that may enter our markets includes a possible rival to the divine chocolate: it is another such rain-forest fruit, cupuassu, whose pulp now fetches top cruzeiro around Manaus for its intense perfumed aroma, "a little overpowering at first."

One enthusiastic expert offers concise prophesy for a new crop, not as food or fiber but to guard the environment. World Bank staff in India report a certain deep-rooted grass able to grow in snow, shade or desert sands, there able to halt soil erosion and hold moisture but never to spread weedlike. "I think we'll see hundreds of thousands of miles of vetiver grass strips banding hillsides in every continent," and perhaps we will.

ESSAY

The sting of truth

by John Kenneth Galbraith

There is a troubling question: Does mainstream economic theory and instruction serve to conceal economic reality? And, a closely related matter, do accepted political attitudes and instruction likewise serve concealment? The answer is, alas, that both do. It is with that misguidance that this essay is concerned.

In the great tradition of economics, nothing is more central than the sovereign role of the consumer. It is to this authority, manifested in the market, that the producing firm responds. This it does with basic efficiency; such response is the natural product of profit maximization in the competitive market. Monopoly or imperfect competition or defective knowledge may invade and frustrate this process. But only what so offends as aberration stands against the basic power of effectively served consumer choice. The textbooks, with appropriate bows to market imperfection and producer error, so affirm.

The facts of economic life are manifestly different. That the producing firm reaches forward to shape and thus to subsume the consumer sovereignty it is assumed to serve is patently clear. Even the most austere of economic scholars watch television.

Not less evident, and especially in these last years, has been the commitment of those who manage the modern corporate enterprise to the maximization of compensation and also of the enjoyments of power not for the firm but, and frequently with grave damage to the firm and the public, for themselves. There has always been a question as to why corporate managers should be powerfully committed to profit maximization but in a selfless way for distant, dispersed, unknown and effectively powerless stockholders. Why shouldn't they, if so motivated, maximize for themselves? We now see that, at least within limits, they do. The mergers and acquisitions mania and the protective leveraged buy-outs have not been for profit maximization by the firm. They have been struggles for power, position and reward by management or those who seek to control and manage. Of this, given the media attention, economists can also hardly have been unaware.

The self-service of the management of the enterprise, not the service of the public, is the reality that the established doctrine conceals. Economic education then extends that concealment in colleges and universities.

In civilian industry this broad and visible tendency to bureaucratic and corporate self-service is only a smaller shadow of its much larger manifestation in the field of military organization and production. Here there is a truly massive escape from public purpose as the business enterprise becomes part of a vastly larger bureaucratic and political complex. The compelling case of bureaucratic and economic power in pursuit of its own purposes and controlling the means by which these purposes are achieved is the modern military establishment, including the intimately associated weapon-producing firms and the more than adequately subservient political acolytes.

This coalition defines the military goals to be pursued, notably the weaponry held to be needed, which is thus developed and produced. These goals and the serving instruments are those not of the public but of the organization. Serving the self-constituted aims of organization is control over effective demand—over the provision of the public funds that pay for the pursuit of organization goals. This control is by a decisive influence on the legislative process. Overall sanction, the equivalent of presumptively beneficent market response, is then held to come from the democratic process. The people, through the Congress and the president, have made the decision and given their assent and the money. Democracy is held to justify what serves the great bureaucracy as it serves its own interest. The result is our highest form of self-sanctioned, self-serving economic and bureaucratic power.

This is evident with rewarding clarity at the present moment. The cold war was long cited as the sustaining case for weaponry that was ever more sophisticated, ever more expensive, ever more excessive in destructive power and ever technically more implausible, going on to the B-2 bomber, the Strategic Defense Initiative and beyond.

With the somewhat reluctant agreement of the president, the cold war now has come to an end. This, at most, has been only mildly troublesome to the military establishment. Its self-constituted authority continues. The weapon production and development and the supporting budgets go on. Accordingly, we see with even greater clarity the independent or autonomous power of the military-industrial complex. It does not really need a plausible enemy.

In the months just past there was indeed much talk of a peace dividend. Those attracted to the idea were regrettably innocent of the deeper situation. The cold war was serving the military-industrial power only as an available justification. When the cold war ended, all that was changed was that the self-sustaining military power became slightly more revealed.

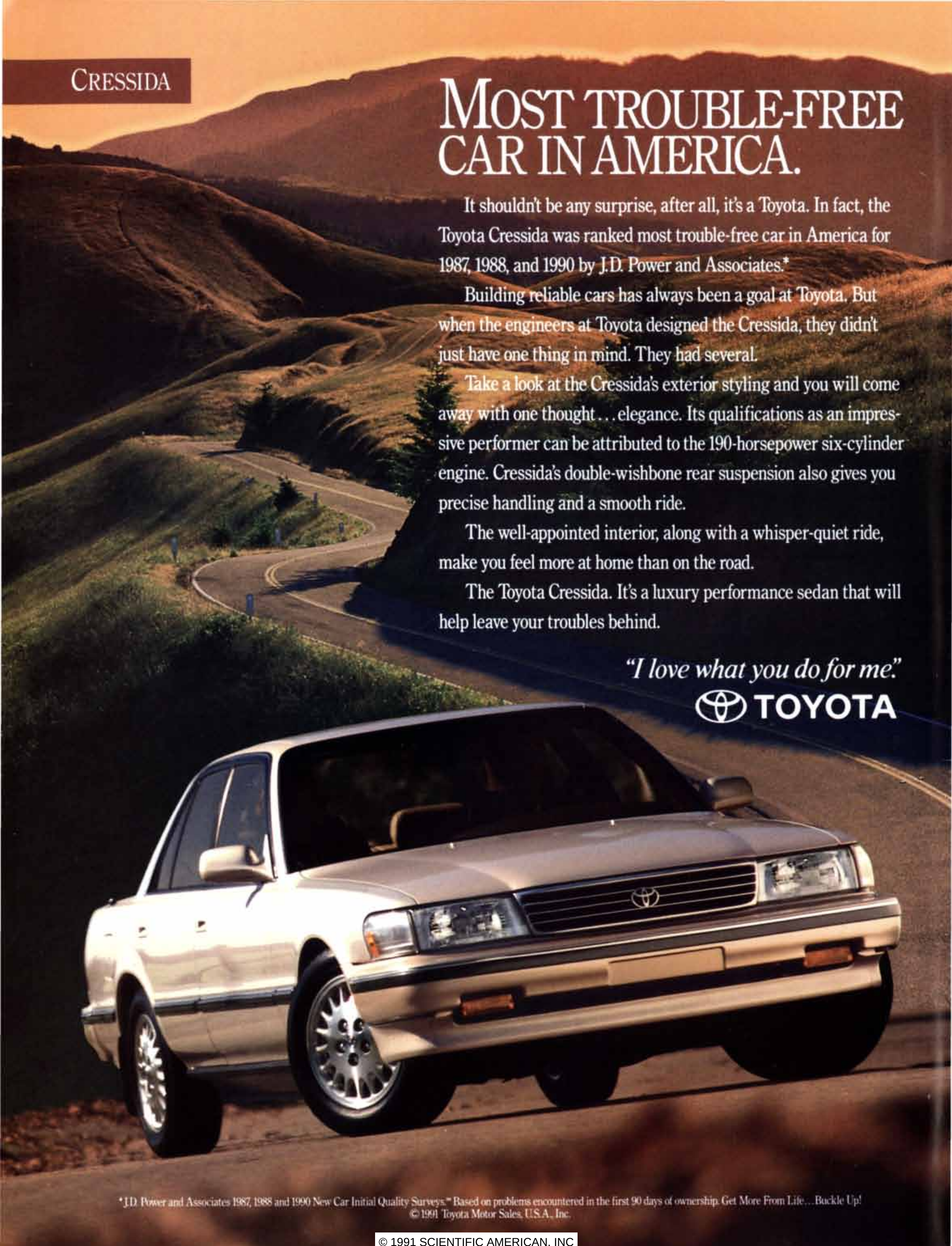
In yet more recent times, as this is written, Saddam Hussein has indeed emerged as a justification for the military-industrial power. There is a sobering possibility that the undue eagerness for conflict on and over the desert sands reflected in some measure military needs and purposes. And, in any case, the larger point remains. Advanced mass-destruction weaponry and other ultrasophisticated weapons had, finally, to give way to ground warfare.

The lesson here for economists and not less for political scientists and for scholars in general is plain. Great organization is a commonplace in our time. The myth that it serves social purpose is deep in our faith and in our scholarly instruction. This must not continue.

The military-industrial bureaucracy, the extreme case, serves extensively its own ends. From its extended arm comes the revenue that buys its products and otherwise finances its existence. Democracy, like the market, is then the covering facade behind which it pursues its own interest. Only as we realize and urge this will we be in the service of countering and therapeutic truth.

And only then will we have an acceptable view of a very substantial part of all economic activity—of between 5 and 6 percent of GNP in calendar 1990 and a full one quarter of all federal expenditure in a time of grave competing need. In dismal contrast with Japan and Germany, our vast and commanding military organization absorbs a very substantial share of our capital and highly qualified manpower. Those who lost in military conflict, in contrast, have used their capital and qualified manpower to defeat us in the larger arena of modern economic competition. Can we in our economic instruction gloss over the motivation in so important a sector of the national economy?

The pursuit by great organization of its own interest should now be central in our research, writing and instruction. If it is not, we are cooperating in a major and economically, socially and politically damaging exercise in concealment. Of this as scholars we must not be guilty. Rather we should welcome, not without pleasure, the sometimes stinging inconvenience of truth.


CRESSIDA
MOST TROUBLE-FREE CAR IN AMERICA.
It shouldn't be any surprise, after all, it's a Toyota. In fact, the Toyota Cressida was ranked most trouble-free car in America for 1987, 1988, and 1990 by J.D. Power and Associates.*
Building reliable cars has always been a goal at Toyota. But when the engineers at Toyota designed the Cressida, they didn't just have one thing in mind. They had several.
Take a look at the Cressida's exterior styling and you will come away with one thought . . . elegance. Its qualifications as an impressive performer can be attributed to the 190-horsepower six-cylinder engine. Cressida's double-wishbone rear suspension also gives you precise handling and a smooth ride.
The well-appointed interior, along with a whisper-quiet ride, make you feel more at home than on the road.
The Toyota Cressida. It's a luxury performance sedan that will help leave your troubles behind.
"I love what you do for me."
TOYOTA
*J.D. Power and Associates 1987, 1988 and 1990 New Car Initial Quality Surveys.℠ Based on problems encountered in the first 90 days of ownership. Get More From Life…Buckle Up!
© 1991 Toyota Motor Sales, U.S.A., Inc.